

Кузнецов Михаил Павлович

ПОСТРОЕНИЕ МОДЕЛЕЙ ОБУЧЕНИЯ ПО ПРЕДПОЧТЕНИЯМ
С ИСПОЛЬЗОВАНИЕМ ПОРЯДКОВЫХ ЭКСПЕРТНЫХ ОЦЕНОК

05.13.18 — Математическое моделирование, численные методы и комплексы программ

Диссертация на соискание ученой степени
кандидата физико-математических наук

Москва — 2016

Оглавление

Введение	4
1 Основные определения и обозначения	11
1.1 Обзор предметной области и базовые методы	11
1.2 Основные понятия, используемые для описания предпочтений	20
1.2.1 Частично упорядоченное множество	20
1.2.2 Свойства матрицы предпочтений	22
1.2.3 Конусное представление предпочтений	25
2 Построение модели предпочтений на наборе объектов, заданных порядковым описанием	28
2.1 Постановка задачи восстановления предпочтений	29
2.2 Восстановление предпочтений с использованием матрицы попарного доминирования объектов	30
2.3 Восстановление предпочтений с использованием конусов	34
2.3.1 Определение функции потерь с использованием конусного представления предпочтений	34
2.3.2 Линейная полиэдральная модель восстановления предпочтений	37
2.3.3 Прямая оценка параметров полиэдральной модели	38
2.3.4 Оценка параметров с использованием центральной точки конуса	39
2.3.5 Построение модели с использованием полиэдрального представления конусов	42
2.4 Восстановление предпочтений методом криволинейной регрессии	45
2.5 Восстановление предпочтений с использованием функции копулы	48
3 Построение модели предпочтений с использованием экспертных оценок	57
3.1 Постановка задачи и базовые методы	59
3.1.1 Ранжирование объектов без учителя	60
3.1.2 Ранжирование объектов с учителем	62

3.2	Согласование экспертных оценок в задаче восстановления предпочтений	63
3.2.1	Согласование линейных экспертных оценок	64
3.2.2	Согласование порядковых экспертных оценок	66
4	Анализ прикладных задач	75
4.1	Категоризация объектов Красной книги РФ	75
4.2	Эксперименты на данных UCI	87
4.3	Ранжирование заповедников для оценки эффективности их управления	89
	Заключение	91
	Список иллюстраций	93
	Список таблиц	94
	Список основных обозначений	95
	Литература	96

Введение

Диссертационная работа посвящена построению математических моделей обучения по предпочтениям. Разработанные методы опираются на анализ частично упорядоченных множеств экспертных оценок объектов и признаков.

Актуальность темы. Область обучения по предпочтениям [41] объединяет задачи машинного обучения с теорией группового выбора [54, 4, 53] и агрегирования предпочтений для восстановления порядковой зависимости на множестве объектов с использованием линейных и порядковых экспертных оценок. В последние годы область обучения по предпочтениям приобрела особую актуальность в связи с возрастающей необходимостью решения задач информационного поиска [61], ранжирования, порядковой классификации и регрессии [64, 44], построения монотонных композиций алгоритмов [3, 43].

Центральной задачей работы является задача построения модели предпочтений на множестве объектов, заданных линейным и порядковым описанием. В работе рассматриваются несколько типов задач, различающихся природой описания признаков и целевой переменной, заданной на множестве объектов.

Первым типом задач является построение модели предпочтений на множестве объектов, заданных линейным признаковым описанием. Впервые задача в подобной постановке была рассмотрена для решения задачи информационного поиска [61]. Требуется построить отображение множества объектов во множество действительных чисел, наиболее точно приближающее целевое отношение предпочтения. В задаче информационного поиска целевое отношение задается оценками ассессоров на наборе документов, соответствующих запросу. Рассматривается класс линейных моделей и функция ошибки особого вида, штрафующая пары объектов, не согласованные по экспертно заданному отношению предпочтения. Моделями подобного типа являются модификация метода опорных векторов (RankSVM) [44, 17] и модификация бустинга (RankBoost) [40]. В случае, когда целевое отношение предпочтения задается конечным набором меток, решается задача порядковой классификации и регрессии [56].

Вторым типом задач является построение модели предпочтений на множестве объектов, заданных порядковым описанием. Другими словами, на множестве объектов задан набор отношений предпочтения, соответствующих порядковым признакам. Требуется корректно определить отображение, приближающее целевое отношение предпочтения. Впервые задача в подобной постановке была рассмотрена в работе В. Коэна и Р. Шапиро [20]. Для восстанов-

ления линейного порядка был предложен метод оценки матрицы попарного доминирования объектов. Основным требованием к отображению является монотонность по набору порядковых признаков [27, 18]. Для решения этой задачи разработаны методы построения монотонных композиций алгоритмов, рассмотренные, например, в работах К. В. Воронцова [3], а также в [22]. Для соблюдения условий монотонности В. В. Стрижовым был предложен подход на основе конусного представления предпочтений [67], развиваемый в данной работе.

Третьим типом задач является моделирование и согласование линейных и порядковых экспертных оценок. В данной работе рассматриваются экспертные оценки объектов, заданных линейным признаковым описанием. Численные методы, разрабатываемые в рамках диссертационной работы, развивают подход В. В. Стрижова [69, 80] и используют элементы теории экспертного оценивания, исследуемой Б. Г. Литваком [28] и А. И. Орловым [48]. В качестве частного случая задачи экспертного оценивания рассматривается задача учета экспертной информации о попарном доминировании признаков. Для ее решения используются элементы теории важности критериев, разрабатываемой В. В. Подиновским [47].

Цели работы.

1. Разработка новых подходов к моделированию объектов, заданных порядковым описанием.
2. Разработка и обоснование эффективных вычислительных методов распознавания объектов, заданных линейным и порядковым описанием.
3. Реализация алгоритмов распознавания и проведение вычислительных экспериментов для проверки адекватности разработанных моделей.

Задачи работы.

1. Разработать подход к моделированию порядковых признаков с использованием полиэдрального представления порядковых данных.
2. Разработать алгоритм построения модели предпочтений на множестве объектов, заданных порядковым признаковым описанием.
3. Разработать метод решения задачи порядковой классификации объектов, заданных линейным и порядковым описанием.
4. Разработать методы моделирования и согласования линейных и порядковых экспертных оценок.

5. Реализовать алгоритмы распознавания порядковых объектов и провести вычислительные эксперименты для установления границ применимости методов.

Основные положения, выносимые на защиту.

1. Метод моделирования объектов, заданных порядковым описанием, с использованием полиэдрального представления предпочтений на множестве объектов.
2. Метод построения модели предпочтений объектов на основе суммы конусов предпочтений.
3. Численный метод решения задачи порядковой классификации объектов на основе оценивания матрицы попарного доминирования объектов.
4. Методы моделирования и согласования порядковых экспертных оценок объектов, заданных линейным признаковым описанием.
5. Программный комплекс, включающий в себя методы распознавания объектов, заданных линейным и порядковым описанием. Проведены вычислительные эксперименты, подтверждающие адекватность методов.

Методы исследования. Для достижения поставленных целей используется теория обучения по предпочтениям [41, 51]. Для исследования суперпозиции порядковых признаков используются механизмы теории голосования и общественного выбора, исследуемые в середине XX века К. Эрроу [4] и Дж. Кемени [53], а также в более современных работах В. Данилова [24] и Д. Саари [71]. Для учета целевой переменной при восстановлении отношения предпочтения используются наработки в области обучения ранжированию [61], описанные в работах В. Коэна и Р. Шапиро [20]. Исследуется обобщение задач порядковой регрессии [22] и классификации [56], а также методы построения монотонных композиций, исследованные, в частности, в работах К. В. Воронцова [3, 77]. Для введения алгебраических структур, описывающих порядковый признак и частично упорядоченное множество, используется подход на основе конусов, впервые разработанный В. В. Стрижовым [67], а также теория анализа экспертной информации, исследуемая Б. Г. Литваком [28] и А. И. Орловым [48]. Кроме того, используются элементы теории выпуклой оптимизации [9], теории вероятностей, теории графов и теории важности критериев, разрабатываемой В. В. Подиновским [47].

Научная новизна. Разработан новый подход к построению математической модели предпочтений на основе полиэдрального представления порядковых данных. Разработаны новые вычислительные методы агрегирования порядковых признаков, основывающиеся на построении суперпозиции конусов предпочтений. Показано, что предлагаемая модель является обобщением модели порядковой классификации с монотонными ограничениями. Установлено теоретическое соответствие между конусом и матрицей инцидентности графа предпочтений. Проведено обобщение известных порядковых метрик на случай частичных порядков с использованием матрицы предпочтений.

Теоретическая значимость. Диссертационная работа устанавливает теоретические обобщения ранее полученных результатов в области обучения по предпочтениям путем введения понятия конуса предпочтений. Введенные понятия помогают провести обобщения для известных ранее понятий ранговой корреляции и монотонных ограничений, а также провести теоретический анализ свойств алгоритмов восстановления предпочтений и порядковой классификации.

Практическая значимость. Разработанные вычислительные методы показывают существенное улучшение качества решения реальных прикладных задач ранжирования и порядковой классификации. Реализованный программный комплекс позволяет проводить самостоятельные исследования в области обучения по предпочтениям и сравнивать результаты с известными методами. В рамках диссертационной работы решена практическая задача категоризации таксонов Красной книги по экспертным оценкам, предоставленным Министерством природных ресурсов РФ.

Степень достоверности и апробация работы. Достоверность результатов подтверждена математическими доказательствами, экспериментальной проверкой полученных методов на реальных задачах ранжирования и порядковой классификации, публикациями результатов исследования в рецензируемых научных изданиях, в том числе рекомендованных ВАК. Результаты работы докладывались и обсуждались на следующих научных конференциях:

- Международная конференция «International Conference on Operational Research», 2011.
- Всероссийская конференция «Математические методы распознавания образов» ММРО-15, 2011, [35].

- Международная конференция «25th European Conference on Operational Research», 2012.
- Международная конференция «Интеллектуализация обработки информации» ИОИ-9, 2012, [36].
- Международная конференция «26th European Conference on Operational Research», 2013.
- Международная конференция «20th Conference of the International Federation of Operational Research Societies», 2014.
- Всероссийская конференция «Математические методы распознавания образов» ММРО-17, 2015, [33].

Публикации по теме диссертации. Основные результаты по теме диссертации изложены в шести изданиях из списка ВАК [30, 46, 68, 58, 79, 57], из которых три индексируемы Scopus, трех сборниках докладов конференций [33, 36, 35] и других печатных изданиях [52, 78, 34, 32, 31].

Личный вклад. Все приведенные результаты, кроме отдельно оговоренных случаев, получены диссертантом лично при научном руководстве д.ф.-м.н. В. В. Стрижова.

Структура и объем работы. Диссертация состоит из оглавления, введения, четырех разделов, заключения, списка иллюстраций, списка таблиц, перечня основных обозначений и списка литературы из 89 наименований. Основной текст занимает 103 страницы.

Краткое содержание работы по главам. В главе 1 вводятся основные определения и обозначения. В разделе 1.1 формулируются основные постановки задач в области обучения по предпочтениям, приводится обзор современных методов для решения задач восстановления предпочтения. В разделе 1.2 вводятся основные понятия, используемые в работе. В подразделе 1.2.1 приводится определение частично упорядоченного множества, графа, соответствующего конечному частично упорядоченному множеству и матрицы инцидентности этого графа. В подразделе 1.2.2 рассматриваются свойства матрицы предпочтений, приводятся обобщенные выражения для известных порядковых метрик. В подразделе 1.2.3 формулируется понятие конуса предпочтений. Приводится два вида описания конуса: полиэдральное описание и описание через конечный набор порождающих элементов. Кроме того, рас-

смачивается конус, задаваемый столбцами матрицы предпочтений в качестве порождающих элементов. Доказываются теоремы о взаимосвязи конусов полиэдрального и порождающего представления.

Глава 2 посвящена проблеме построения модели предпочтений объектов, заданных порядковым описанием. В разделе 2.1 формулируется общая постановка задачи с необходимыми ограничениями на монотонность и вид отображения. В разделе 2.2 ставится задача максимизации обобщенной ранговой корреляции и решается задача оценивания матрицы попарного доминирования объектов. На основе этой матрицы решается задача восстановления линейного порядка на множестве объектов.

Раздел 2.3 посвящен конусному представлению предпочтений и построению модели на основе суперпозиции конусов. В подразделе 2.3.1 вводятся функции потерь, использующие полиэдральное и порождающее представления конусов, а также целевое отношение предпочтения. В подразделе 2.3.2 определяется линейная полиэдральная модель, использующая в качестве множества допустимых значений сумму Минковского конусов, соответствующих порядковым признакам. Последующие подразделы посвящены оценке параметров линейной полиэдральной модели. В подразделе 2.3.3 исследуется метод прямой оценки параметров путем решения задачи неотрицательной линейной регрессии. В подразделе 2.3.4 предлагается метод сокращения признакового пространства путем использования центральных точек конусов. Доказывается теорема о сводимости задачи к задаче максимизации обобщенной τ -корреляции. В подразделе 2.3.5 предлагается метод построения модели на основе полиэдрального представления конусов.

В разделах 2.4 и 2.5 предлагаются альтернативные методы решения задачи. Так, в разделе 2.4 рассматривается модель криволинейной регрессии, выполняющая монотонное преобразование порядковых признаков. В разделе 2.5 рассматривается описание порядковой зависимости функцией копулы.

Глава 3 посвящена задаче построения оценок объектов с использованием экспертных оценок в линейных и порядковых шкалах. В разделе 3.1 приводится общая постановка задачи и описание базовых методов. В подразделе 3.1.1 рассматривается задача построения интегральных оценок объектов без учителя. В подразделе 3.1.2 рассматривается задача ранжирования с учителем, устанавливается различие между линейной и порядковой целевой переменной. Раздел 3.2 посвящен согласованию экспертных оценок и содержит два подраздела. В подразделе 3.2.1 приводятся методы согласования линейных экспертных оценок. В подразделе 3.2.2 приводятся методы согласования порядковых экспертных оценок с учетом понятий, приведенных в разделе 2.3.

Глава 4 посвящена анализу прикладных задач. В разделе 4.1 приводятся результаты решения задачи категоризации видов Красной книги РФ. В разделе 4.2 приводится описание программного комплекса, предназначенного для тестирования алгоритмов порядковой классификации объектов, заданных порядковым описанием, на различных наборах данных, в частности, на данных из репозитория UCI [6]. Приводится результат сравнения качества предлагаемых алгоритмов с известными методами. В разделе 4.3 приводятся результаты решения задачи ранжирования заповедников.

1 Основные определения и обозначения

1.1 Обзор предметной области и базовые методы

В связи с необходимостью решения задач информационного поиска [7], рекомендательных систем [73, 1], коллаборативной фильтрации [50], в последние годы получила развитие область, называемая *обучение по предпочтениям*. Эта область возникла на стыке двух направлений. Первым направлением является анализ данных и, в частности, задача обучения по прецедентам. Второе направление включает в себя теорию группового выбора и бинарных отношений. Необходимость соединения этих направлений возникла, в первую очередь, из-за того, что данные, получаемые исследователем для анализа, имеют нечисловую структуру. В частности, на множестве объектов, ответов или признаков может быть задано бинарное отношение порядка, выражающее степень попарного доминирования сущностей. На основании этой информации об объектах требуется решить задачу восстановления зависимости, числовой или порядковой, как, например, в случае задачи ранжирования альтернатив.

Искомая порядковая зависимость представляется линейным или частичным отношением порядка на заданном множестве сущностей. Рассмотрение частичных отношений порядка, являющееся характерной особенностью данной работы, вызвано наличием неполной экспертной информации о наборе объектов. В частности, не всегда у экспертов появляется возможность разметить большой объем данных (как, например, в задаче коллаборативной фильтрации), или же экспертных знаний оказывается недостаточно для оценки отношения доминирования между некоторой парой объектов. В этом случае, рассматривается задача агрегирования частичных порядков, приведения неполной экспертной информации об объектах или признаках к единому согласованному решению.

Определения и обозначения. Задано *множество объектов* $X = \{x_1, \dots, x_m\}$ размера m . В рамках данной работы предполагается, что рассматриваемые объекты имеют линейное или порядковое признаковое описание.

Определение 1. *Линейное признаковое описание:* объект x_i является элементом евклидова пространства \mathbb{R}^n .

В этом случае, объект является вектором $\mathbf{x}_i \in \mathbb{R}^n$, и определена *матрица плана* $\mathbf{X} = [\mathbf{x}_1^\top; \dots; \mathbf{x}_m^\top]$. Задачи обучения по предпочтениям для объектов, заданных линейным описанием, рассматриваются в главе 3.

Определение 2. Порядковое признаковое описание: на множестве объектов X задано n отношений предпочтения z_1, \dots, z_n , каждое из которых описывает отношение доминирования между парами объектов, $z_j(x_i, x_k) = [x_i \succeq_j x_k]$, где $[\cdot]$ — индикаторная функция.

В рамках данной работы рассматриваются отношения предпочтения, задаваемые отношением частичного порядка на множестве объектов, основные свойства которого приведены в разделе 1.2. Задачи обучения по предпочтениям для объектов, заданных порядковым описанием, рассматриваются в главе 2.

Определение 3. Целевое отношение предпочтения z_0 — отношение, описывающее искомое отношение доминирования между парами объектов из X , $z_0(x_i, x_k) = [x_i \succeq_0 x_k]$.

Ограничения на структуру целевого предпочтения z_0 определяют класс решаемых задач. Так, если z_0 задается отношением частичного порядка, решается задача восстановления предпочтения.

Кроме того, рассматриваются случаи, когда отношение z_0 определяется конечным набором меток Y . В этом случае, задана выборка объектов $D = \{(x_i, y_i)\}_{i=1}^m$ со значениями целевой переменной из Y , $y_i \in Y$, и отношение предпочтения z_0 определяется доминированием меток $z_0(x_i, x_k) = [y_i \succeq y_k]$. В рамках данной работы рассматриваются случаи двухклассовой классификации, $Y = \{0, 1\}$, и порядковой многоклассовой классификации, $Y = \{y_1, \dots, y_K \mid y_1 \succ \dots \succ y_K\}$.

Определение 4. Модель f — отображение множества объектов X во множество действительных чисел \mathbb{R} .

Если это оговорено дополнительно, то при решении задачи с конечным множеством меток Y рассматривается задача построения отображения $f : X \rightarrow Y$. В качестве частного случая, когда объекты имеют линейное признаковое описание $x_i \in \mathbb{R}^n$, рассматривается семейство *линейных моделей*,

$$f(\mathbf{w}, \mathbf{x}) = \mathbf{w}^\top \mathbf{x}, \quad (1)$$

где \mathbf{w} — вектор весов признаков.

Для определения оптимальных параметров модели определена *функция потерь* $S(f, X, z_0)$ модели f на множестве объектов X . Основной задачей является поиск отображения \hat{f} , минимизирующего функцию потерь на X ,

$$\hat{f} = \arg \min_{f \in \mathcal{F}} S(f, X, z_0),$$

где \mathcal{F} — класс допустимых отображений. Подробный вид функции потерь будет рассмотрен в главах 2 и 3 при постановке задач.

В следующих параграфах рассмотрим приведем краткий исторический обзор, относящийся к порядковому описанию объектов и целевой переменной.

Порядковое описание объектов. Как было сказано выше, на множестве объектов X задано n отношений предпочтения z_1, \dots, z_n , каждое из которых описывает отношение доминирования между парами объектов, $z_j(x_i, x_k) = [x_i \succeq_j x_k]$. В более частной постановке каждое из отношений z_1, \dots, z_n задается конечным множеством меток из множеств X_1, \dots, X_n , и множество X представляется декартовым произведением множеств X_1, \dots, X_n . В этом случае определена *матрица плана* \mathbf{X} , элементы которой являются элементами упорядоченных множеств.

Пример 1. *Матрица плана для трех объектов, описанных двумя порядковыми признаками.*

$$\mathbf{X} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 1 \end{pmatrix}$$

В примере 1 показана матрица плана для трех объектов, описанных двумя порядковыми признаками. Отметим, что цифры в этой матрице не являются действительными числами, а выражают степень превосходства $3 \succ 2 \succ 1$. Например, в некоторой интерпретации цифра «1» означает «плохо», «2» — «нормально», «3» — «хорошо». В более общей постановке, когда на множестве X задан набор отношений предпочтения z_1, \dots, z_n , не представляемых конечным набором меток, описание объектов не может быть выражено двумерной матрицей.

Впервые подобная постановка была рассмотрена в XVIII веке, когда было сформировано направление теории общественного выбора [5, 60, 72]. В терминах этого направления, объектами x_1, \dots, x_m являются кандидаты, участвующие в голосовании, а признаки соответствуют n избирателям. Первоначальная задача состояла в определении корректной системы голосования, при которой один из m кандидатов выбирается «наиболее честным» путем. В 1785 году маркизом Кондорсе был описан *парадокс Кондорсе*, согласно которому коллективное правило голосования может содержать транзитивные циклы, даже в случаях, когда исходные ранги избирателей не содержат циклов. Этот парадокс приводит к неразрешимым противоречиям при выборе системы голосования.

Позднее парадокс Кондорсе был обобщен *теоремой Эрроу о невозможности*, сформулированной К. Эрроу в 1951 году [4]. Теорема утверждает, что при $m \geq 3$ и $n \geq 2$ не

существует системы голосования, удовлетворяющей одновременно пяти базовым условиям: универсальности, полноты, монотонности, отсутствию диктатора и независимости от посторонних альтернатив. Другими словами, для любой системы голосования можно подобрать пример, когда хотя бы одно из пяти базовых требований не будет соблюдаться, и универсальной системы голосования не существует.

Для построения адекватной системы голосования были определены дополнительные критерии качества системы [89]. Так, одним из наиболее исследованных является *критерий Кондорсе*. Согласно этому критерию, *победителем по Кондорсе* является тот кандидат, который выигрывает у всех остальных кандидатов при всех попарных сравнениях. Говорят, что система голосования удовлетворяет критерию Кондорсе, если победителем по этой системе является победитель по Кондорсе, в том случае, если победитель по Кондорсе существует (победитель по Кондорсе существует не всегда, что напрямую следует из теоремы о невозможности). В число систем голосования, удовлетворяющих принципу Кондорсе, входят такие правила, как правило *ранжирования пар*, *правило Шульце* и *правило Кемени*.

Правило ранжирования пар [82] состоит из трех основных этапов: 1) подсчитать количество голосов при всех попарных сравнениях кандидатов, 2) отсортировать все пары по «величине победы» одного кандидата над другим, 3) построить конечное ранжирование по порядку отсортированных пар, пропуская те пары, которые при их рассмотрении образуют цикл в полученном ранжировании. Показано, что система голосования, устроенная таким образом, удовлетворяет принципу Кондорсе.

Метод Шульце [70], разработанный в 1997 году Маркусом Шульце, использует эвристику пути для ранжирования кандидатов. Согласно этой эвристике, путем от одного кандидата к другому является цепочка победителей-проигравших при парном сравнении, начинающаяся на первом кандидате и заканчивающаяся на втором. Для каждого кандидата рассчитывается «сила пути», и победителем считается кандидат с наибольшей суммарной силой.

Одним из наиболее применяемых и исследованных правил голосования является правило Кемени (англ.: Kemeny voting rule) [53, 72]. Правило Кемени формулируется следующим образом. На множестве объектов X задано n отношений предпочтения z_1, \dots, z_n . Необходимо найти отношение линейного порядка z^* , минимизирующее среднее расстояние до всех z_1, \dots, z_n :

$$z^* = \arg \min_{z \in \text{LO}} \sum_{j=1}^n \sum_{i,k=1}^m |z(x_i, x_k) - z_j(x_i, x_k)|, \quad (2)$$

где LO — множество всех линейных порядков.

Отметим, что задача нахождения медианы Кемени является NP-трудной для случая $n \geq 4$ [8]. Для ее вычисления предложены вероятностные методы [2], а также приближенные, основанные на методе ветвей и границ [25] и др. [74, 88, 21]. Правило Кемени также удовлетворяет принципу Кондорсе.

Позднее в работах [88, 87] была рассмотрена более общая остановка агрегирования предпочтений, согласно которой каждому предпочтению соответствовало отношение частичного порядка z_j , определенное на множестве X , и частично упорядоченное множество X_j . Был предложен вероятностный метод моделирования предпочтений, при котором истинное ранжирование на X является неизвестным, а наблюдаемые X_1, \dots, X_n — его зашумленными реализациями. Для нахождения оптимального ранжирования X^+ был предложен метод максимума правдоподобия

$$X^+ = \arg \max_X \prod_{j=1}^n \rho(X_j | X),$$

где ρ является условным распределением X_i при истинном X .

Наличие целевой переменной: обучение ранжированию. В предыдущем параграфе были рассмотрены задачи теории общественного выбора и агрегирования предпочтений без учителя. В задачах подобного типа отсутствует целевое отношение предпочтения, или зависимая переменная, и основной целью является поиск ранжирования, являющегося средним в некотором смысле.

Необходимость учета целевой переменной в задаче восстановления предпочтения появилась в конце XX века в связи с возникновением поисковых систем и появлением задачи *обучения ранжированию* [61, 43, 10]. Эта задача была впервые сформулирована в области информационного поиска и заключается в следующем. Поисковая система хранит внутри себя набор текстовых документов. На каждый запрос пользователя система возвращает список документов, отранжированный по степени релевантности, предсказанной системой. Помимо данных запрос-документ в системе также хранятся оценки ассессоров, бинарные величины, указывающие, является ли документ релевантным запросу. Оценки ассессоров являются обучающей информацией: параметры модели настраиваются таким образом, чтобы минимизировать расхождение между оценками ассессоров и оценками релевантности, возвращаемыми системой. В качестве меры расхождения рассматриваются меры *discounted cumulative gain* (DCG) и *normalized discounted cumulative gain* (NDCG).

Одной из первых работ взвешивания предпочтений с учетом целевой переменной стала работа Коэна и Шапиро [20]. Следуя нашим обозначениям, задано множество объектов X ,

на котором определено n отношений предпочтения z_1, \dots, z_n , таких, что

$$z_j(x_i, x_k) = [x_i \succeq_j x_k].$$

Кроме того, на множестве объектов задано целевое отношение z_0 . Была рассмотрена задача построения отображения $f : X \rightarrow Y$, где Y — множество, наделенное отношением линейного порядка. При этом отношение предпочтения z_f , построенное на основе отображения f , должно минимизировать расхождение с целевым отношением z_0 , $S(z_f, z_0) \rightarrow \min$. В качестве функции потерь, измеряющей расхождение между предпочтениями, была предложена функция, обратно пропорциональная τ -корреляции [54].

Для решения этой задачи в [20] было предложено оценивать функция доминирования $\text{PREF}(x_i, x_k)$ (в рамках данной работы будем использовать обозначение $\hat{z}(x_i, x_k)$), выражающую степень доминирования между каждой парой объектов. Было предложено оценивать эту функцию на основе линейной комбинации базовых функций предпочтения, $\hat{z}(x_i, x_k) = \sum_{j=1}^n w_j z_j(x_i, x_k)$, а параметры w_i оценивать адаптивно путем минимизации функции потерь. В качестве окончательной оценки объекта $f(x_i)$ было предложено использовать комбинацию вида

$$f(x_i) = \sum_{k=1}^n (\hat{z}(x_i, x_k) - \hat{z}(x_k, x_i)).$$

В общей постановке задачи восстановления предпочтения на множестве значений целевой переменной экспертами задано отношение порядка, определяющее отношение предпочтения $z_0(x_i, x_k) = [y_i \succeq y_k]$. В частном случае рассматривают конечное множество из K элементов с введенным на нем отношением полного линейного порядка,

$$Y = \{y_1, \dots, y_K, y_1 \succ y_2 \succ \dots \succ y_K\}. \quad (3)$$

Ряд задач, использующих подобный вид целевой переменной, определяемой конечным набором меток, принадлежит к категории обучения ранжированию, или порядковой (ранговой) регрессии [64, 63].

Основной трудностью для решения задач подобного типа является определение функции потерь S вследствие необходимости определения нормы на упорядоченном множестве Y и аппроксимации DCG-метрик [17]. В [43] рассматривают три класса подходов к решению задачи обучению ранжированию, каждый из которых характеризуется особым видом функции потерь: поточечный (pointwise), попарный (pairwise) и списочный (listwise) подходы.

Поточечный подход к задачам обучения ранжированию [23, 75] основан на обобщении стандартных методов классификации и регрессии, при котором числовая ось значений отоб-

ражения f разбивается на области, соответствующие упорядоченным меткам из множества Y . Так, в [64] было предложено обобщение метода логистической регрессии, заключающееся в последовательном решении промежуточных задач бинарной классификации, различающих объекты смежных классов y_k, y_{k+1} . Результирующий классификатор комбинирует результаты промежуточных. В дальнейшем такое представление задачи порядковой регрессии получило широкое развитие, в частности, в работах [56, 14].

В некоторых случаях, [56, 77] в рамках поточечного подхода на промежуточном этапе решения задачи возникает подзадача монотонизации выборки, решаемая методами линейного программирования [9]. Задача заключается в поиске переменных p_i , которые близки к значениям целевой переменной y_i , но согласованы по монотонности с объектами x_i :

$$\arg \min_{p_i} (y_i - p_i)^2, \quad \text{при выполнении } x_i \succeq x_j \Rightarrow p_i \geq p_j,$$

где неравенство $x_i \succeq x_j$ является покомпонентным.

В рамках попарного подхода задача ранжирования сводится к задаче двухклассовой классификации в пространстве пар объектов. Стандартная функция потерь штрафует пары объектов, не согласованных по экспертно заданному предпочтению z_0 . Одними из самых используемых алгоритмов для данного подхода служат модификация стандартного метода опорных векторов (RankSVM) [44, 12] и модификация бустинга (RankBoost) [40]. В частности, минимизируемая функция потерь для метода RankSVM выглядит следующим образом:

$$S(f, X, z_0) = \sum_{i,k=1}^m (1 - z_0(x_i, x_k) \mathbf{w}^\top (\mathbf{x}_i - \mathbf{x}_k))_+ + \lambda \|\mathbf{w}\|^2. \quad (4)$$

Такой вид функции потерь отличается от стандартной [85] видом функции отступа. Фактически, гиперплоскость «наибольшего зазора» отыскивается в пространстве пар объектов. При этом штрафуются те пары объектов, для которых не совпадают экспертное отношение доминирования и отношение доминирования, порожденного выражением $\mathbf{w}^\top (\mathbf{x}_i - \mathbf{x}_k)$.

Отметим, что приведенная выше функция потерь зависит от двух матриц: экспертной матрицы бинарного отношения с элементами $\mathbf{Z}_0(i, k) = z_0(x_i, x_k)$ и матрицы, элементы которой $\mathbf{w}^\top (\mathbf{x}_i - \mathbf{x}_k)$ порождают отношение доминирования для модели $f = \mathbf{w}^\top \mathbf{x}$. В данной работе будет подробно исследована взаимосвязь этих матриц, а также проведено обобщение подходов к решению задачи с помощью введения новой функции потерь, основывающейся на норме разности двух матриц.

В рамках списочного подхода к задаче ранжирования [13, 86] рассматривается функция потерь $S(X, \pi)$, которая зависит от перестановки объектов π на множестве объектов X . Ищется такая перестановка $\hat{\pi}$, которая минимизирует функцию потерь $\hat{\pi} = \arg \min_{\pi} S(X, \pi)$.

Отметим, что задача в подобной постановке обладает большой вычислительной сложностью, а функция потерь является дискретной и разрывной. Таким образом, основной задачей списочного подхода является поиск вычислительно эффективной аппроксимации функции потерь.

Приведенные выше методики обучения ранжированию в общем случае учитывают порядковую структуру зависимой переменной, но не учитывают порядковое описание объектов. Так, в рамках подхода RankSVM рассматриваются модели вида $f(\mathbf{w}, \mathbf{x}) = \mathbf{w}^T \mathbf{x}$, то есть подразумевается, что объекты представлены линейным описанием $\mathbf{x} \in \mathbb{R}^n$. Тем временем, недавние исследования [27, 79] показали, что при порядковой структуре данных переход к монотонным моделям позволяет существенно повысить качество алгоритмов ранжирования.

Построение монотонного отображения $f : X \rightarrow Y$ на декартовом произведении порядковых множеств $X = X_1 \times X_2 \times \dots \times X_n$ в области обучения предпочтений получило название задачи порядковой классификации (регрессии) с монотонными ограничениями [22, 56, 26, 18]. Основная идея построения такого отображения заключается во введении дополнительного множества параметров, которые отвечают за преобразование порядковой [65, 55] шкалы множества X_j , в линеаризации этого множества для поиска точки в линейном пространстве, наиболее согласованной с моделью $f(\mathbf{w}, \mathbf{x})$ в смысле функции потерь S . В частности, в [22] рассматривают монотонную аддитивную модель вида

$$f(\mathbf{x}) = \sum_{j=1}^n u_j(\mathbf{x}_j),$$

где u_j являются монотонными функциями, принимающими значения в промежутке $[0; 1]$. При этом параметры всех u_j являются независимыми величинами, то есть каждому из множеств X_j соответствует свое монотонное преобразование.

В работе [27] был предложен алгоритм монотонного ближайшего соседа для задачи порядковой классификации. Согласно этому методу, гиперплоскость, формируемая стандартным методом ближайшего соседа, подвергается дополнительной монотонной коррекции по исходным признаковым описаниям объектов. Эксперименты показали существенное улучшение качества классификации с использованием монотонных ограничений по сравнению со стандартным методом ближайшего соседа.

В работах К. В. Воронцова [3, 77] был предложен подход построения монотонной модели на основе использования локальных базисов. Подход строит монотонную композицию алгоритмов, устраняющую дефектные пары объектов, то есть пары, не согласованные по целевому и вычисляемому предпочтениям.

В последних работах [41] различные постановки задач с порядковой структурой описаний объектов и целевой переменной были объединены в область обучения по предпочтениям. Так, было рассмотрено три типа задач.

1. Ранжирование меток (англ.: label ranking), поиск отображения вида $X \rightarrow S_Y$, где S_Y — множество перестановок элементов Y .
2. Ранжирование сущностей (англ.: instance ranking), поиск отображения $X \rightarrow Y$, где Y является конечным множеством упорядоченных меток, $Y = \{y_1 \succ \dots \succ y_K\}$.
3. Ранжирование объектов (англ.: object ranking), поиск функции бинарного отношения на множестве объектов X .

Важность критериев. В рамках данной работы в главе 3 рассматриваются методы обучения по предпочтениям, использующие экспертную информацию о важности критериев, или признаков. Говоря формально, на множестве предпочтений $\{z_1, \dots, z_n\}$, определенных на множестве объектов X , определено отношение предпочтения критериев, например, $z_1 \succ \dots \succ z_n$. Другими словами, экспертами определено, какой из критериев является более важным, чем остальные. Для построения корректной модели, учитывающей порядок на множестве предпочтений, используются элементы теории важности критериев, разработанные В. В. Подиновским [47].

В случае линейного описания объектов, $\mathbf{x} \in \mathbb{R}^n$, и линейной модели $f = \mathbf{w}^\top \mathbf{x}$, важность критериев задается предпочтениям на множестве весов признаков,

$$\mathcal{W} = \{\mathbf{w} \mid w_1 \geq w_2 \geq \dots \geq w_n \geq 0\}. \quad (5)$$

В этом случае, модель (1) является моделью линейной регрессии с ограничениями [42]:

$$y = \mathbf{w}^\top \mathbf{x}, \quad \mathbf{w} \in \mathcal{W}, \quad (6)$$

где \mathcal{W} — полиэдральное множество [29]. Особенность представленной задачи состоит в том, что множество экспертных ограничений \mathcal{W} является конусом, то есть частным случаем полиэдрального множества.

Впервые понятие конуса, соответствующего набору предпочтений, было введено В. В. Стрижовым в работе [67]. Понятие конуса предпочтений также широко используется в данной работе. Так, одним из центральных фактов является установление связи между конусом, соответствующим частичному порядку, и матрицей инцидентности графа частичного порядка.

Она заключается в том, что каждая точка конуса раскладывается в неотрицательную комбинацию своих порождающих элементов, также являющихся столбцами матрицы частичного порядка. Благодаря этому факту, задача (6) может быть переписана в терминах известной задачи неотрицательной линейной регрессии (англ.: nonnegative least squares) [59, 76].

В данной работе также исследуется задача, объединяющая постановки (5) и (3) настоящего и предыдущего разделов. В частности, рассматриваются два конуса, соответствующие как экспертным оценкам целевой переменной, так и оценкам весов показателей. Предлагается процедура поиска согласованного решения [67] на основании минимизации расстояния между конусами.

1.2 Основные понятия, используемые для описания предпочтений

В рамках диссертационной работы рассматривается класс отношений предпочтения, задаваемых отношениями частичного порядка на объектах, признаках или множестве значений признака. В настоящем разделе будут введены определения частично упорядоченного множества, графа и конуса, соответствующих частичному порядку. Будет сформулирована теорема о разложении точки конуса, соответствующего частичному порядку.

1.2.1 Частично упорядоченное множество

Определение 5. *Частичным порядком на множестве X называется бинарное отношение \succeq , удовлетворяющее условиям*

- *рефлексивности, $\forall a \in X (a \succeq a)$,*
- *антисимметричности, $\forall a, b \in X, (a \succeq b) \wedge (b \succeq a) \Rightarrow (a = b)$,*
- *транзитивности $\forall a, b, c \in X (a \succeq b) \wedge (b \succeq c) \Rightarrow (a \succeq c)$.*

Для дальнейшего удобства будем обозначать за z бинарную функцию двух аргументов, выражающую отношение частичного порядка, заданного на множестве X :

$$z(a, b) = [a \succeq b], \quad a, b \in X,$$

где $[\cdot]$ — индикаторная функция. Функцию z также будем называть *отношением предпочтения*, определенным на множестве объектов X .

Определение 6. *Частично упорядоченное множество X — это множество, на котором задано отношение частичного порядка.*

Отметим, что определенное выше отношение частичного порядка является нестрогим. В некоторых случаях, мы будем использовать отношение *строгого* частичного порядка \succ , удовлетворяющего свойствам антирефлексивности, ($a \not\succeq a$), асимметричности, ($a \succ b \Rightarrow b \not\succeq a$), и транзитивности. Для частично упорядоченного множества различают пары сравнимых и несравнимых объектов. Объекты a и $b \in X$ являются сравнимыми, если для них выполнено $a \succeq b$ или $b \succeq a$. В противном случае, объекты являются несравнимыми.

Частично упорядоченному множеству поставим в соответствие *граф предпочтений* (*граф частичного порядка*) — невзвешенный ориентированный граф, в котором вершины являются объектами множества X , а ребро, соединяющее вершины a и b , указывает выполнение бинарного отношения z . Будет рассмотрено три важных примера отношений предпочтения, задаваемых отношениями частичного порядка:

1. Отношение предпочтения, заданное на целевой переменной y . В этом случае, для набора объектов эксперты предоставляют информацию предпочтения на множестве значений их целевой переменной $y_1 \succeq y_2 \succeq \dots \succeq y_K$.
2. Порядковое описание объектов x_i . В этом случае, признаковое описание объекта представляется в порядковой шкале.
3. Порядковые экспертные оценки весов признаков w . Эксперты предоставляют информацию о том, какие показатели являются наиболее важными.

Подробнее эти три случая будут рассмотрены в следующих разделах. Отметим, что перечисленные множества являются конечными. Введем определение *матрицы предпочтений* для множества X .

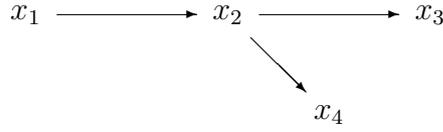
Определение 7. *Матрицей предпочтений (матрицей частичного порядка) \mathbf{Z} для конечного частично упорядоченного множества $X = \{x_1, \dots, x_m\}$ является матрица инцидентности графа, соответствующего частичному порядку, определенному на X :*

$$\mathbf{Z}(i, k) = z(x_i, x_k). \quad (7)$$

Пример 2. *Приведем пример частично упорядоченного множества X , состоящего из четырех объектов, таких что*

$$X = \{x_1, x_2, x_3, x_4 \mid z : x_1 \succeq x_2, x_2 \succeq x_3, x_2 \succeq x_4\}.$$

Отношение предпочтения z на объектах x_1, x_2, x_3, x_4 описывается графом частичного порядка



Этому графу соответствует матрица предпочтений \mathbf{Z} :

$$\mathbf{Z} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

1.2.2 Свойства матрицы предпочтений

В этом разделе приведем некоторые свойства отношений предпочтения. В частности, выразим в терминах матрицы предпочтений некоторые порядковые метрики (ранговую корреляцию Спирмена, Кендалла), правило Кемени, а также характеристику площади под кривой (AUC). Приведенные обобщения порядковых метрик будут использоваться в последующих разделах в качестве оптимизируемых функций потерь.

Ранговая корреляция. Пусть задано множество объектов $X = \{x_1, \dots, x_m\}$, и на этом множестве задано два различных отношения предпочтения, выраженных отношениями частичного порядка z_1, z_2 . Ниже введем два известных определения корреляции Спирмена и Кендалла между двумя векторами рангов, определенных на множестве X .

Определение 8. Коэффициентом корреляции Спирмена ρ_s называется величина

$$\rho_s = 1 - \frac{6}{m(m-1)(m+1)} \sum_{i=1}^m (R_i - S_i)^2,$$

где R_i — ранг объекта x_i во множестве X согласно отношению z_1 , S_i — ранг объекта x_i во множестве X согласно отношению z_2 .

Лемма 1. Для коэффициента корреляции Спирмена выполнено следующее соотношение:

$$\rho_s \propto - \sum_{i=1}^m \left(\sum_{k=1}^m (\mathbf{Z}_1(i, k) - \mathbf{Z}_2(i, k)) \right)^2, \quad (8)$$

где $\mathbf{Z}_1, \mathbf{Z}_2$ — матрицы предпочтений, соответствующих z_1, z_2 .

Доказательство. Доказательство этой леммы следует из того факта, что ранг наблюдения объекта x_i является суммой значений всех элементов матрицы \mathbf{Z} в строке i . \square

Определение 9. Коэффициентом корреляции Кендалла τ называется величина, подсчитывающая количество инверсий в двух порядковых рядах:

$$\tau = 1 - \frac{4}{m(m-1)}R, \quad \text{где } R = \sum_{i=1}^{m-1} \sum_{k=1}^m I[z_1(x_i, x_k) \neq z_2(x_i, x_k)].$$

Лемма 2. Для коэффициента корреляции Кендалла справедливо соотношение:

$$\tau \propto - \sum_{i=1}^m \sum_{k=1}^m (\mathbf{Z}_1(i, k) - \mathbf{Z}_2(i, k))^2 = -\|\mathbf{Z}_1 - \mathbf{Z}_2\|_F, \quad (9)$$

где $\|\cdot\|_F$ — норма Фробениуса.

Доказательство этой леммы следует непосредственно из определения корреляции Кендалла.

Отметим, что формулы (8) и (9) позволяют обобщить стандартные ранговые корреляции Спирмена и Кендалла на случай частичных порядков, а также на случай нечеткого отношения предпочтения $\mathbf{Z}(i, k) \in [0, 1]$, который будет рассмотрен в дальнейшем.

Медиана Кемени. Согласно формуле (2), правило Кемени заключается в поиске линейного порядка z на множестве объектов X , являющегося средним в смысле τ -корреляции для всех порядков z_1, \dots, z_n , определенных на X . Из формулы (9) непосредственно следует лемма об обобщении правила Кемени на случай частичных порядков, задаваемых матрицами \mathbf{Z}_i .

Лемма 3. Медианой в смысле Кемени называется линейный порядок, задаваемый матрицей \mathbf{Z}^* , минимизирующей выражение

$$\mathbf{Z}^* = \arg \min_{\mathbf{Z} \in LO} \sum_{i=1}^n \|\mathbf{Z} - \mathbf{Z}_i\|_F,$$

где LO — множество матриц, задающих линейный порядок.

Площадь под кривой. Используем понятие матрицы предпочтений для обобщения понятия площади под кривой (Area Under Curve, AUC), используемый для измерения качества алгоритма классификации, на случай задачи многоклассовой порядковой классификации.

Рассмотрим исходную формулу вычисления AUC для случая двух классов. Задано множество пар объект-ответ $\{(x_1, y_1), \dots, (x_m, y_m)\}$, где $y_i \in \{0, 1\}$. Обозначим за m_0 количество объектов класса 0, а за m_1 — количество объектов класса 1. За вектор \mathbf{p} обозначим вектор

ответов алгоритма классификации размера m , $p_i \in \mathbb{R}$. В этом случае, площадь под кривой вычисляется по формуле

$$AUC = \frac{1}{m_0 m_1} \sum_{k=1}^{m_0} \sum_{i=1}^{m_1} [p_i > p_k]. \quad (10)$$

Из определения матрицы предпочтения следует утверждение о представлении AUC:

Лемма 4.

$$AUC = \frac{1}{m_0 m_1} \sum_{k,i=1}^m [\mathbf{Z}(k, i) = 0][\hat{\mathbf{Z}}(k, i) = 0], \quad (11)$$

где $\hat{\mathbf{Z}}$ — матрица предпочтений, соответствующая вектору оценок алгоритма \mathbf{p} .

Доказательство. Формула (11) может быть переписана в виде

$$AUC = \frac{1}{m_0 m_1} \sum_{k,i=1}^m [y_k = 0, y_i = 1][p_k < p_i] = \frac{1}{m_0 m_1} \sum_{i=1}^{m_0} \sum_{k=1}^{m_1} [p_i > p_k].$$

□

Обобщим понятие площади под кривой на многоклассовый порядковый случай с использованием матрицы предпочтений. Пусть Y является конечным множеством меток, $Y = \{1, \dots, K\}$ с соответствующим количеством элементов в выборке m_1, \dots, m_k . Формула (11) обобщается следующим образом:

$$AUC = \frac{1}{M} \sum_{k,i=1}^m [\mathbf{Z}(k, i) = 0][\hat{\mathbf{Z}}(k, i) = 0], \quad (12)$$

где $M = \sum_{k_1 < k_2} m_{k_1} m_{k_2}$. Отметим, что в двухклассовом случае многоклассовая формула (12) переходит в (11). Кроме того, для задачи регрессии, когда $y = \mathbb{R}$, справедливо следующее утверждение.

Лемма 5. В случае задачи регрессии, AUC, определенный формулой (12), равен τ -корреляции между вектором ответов и вектором оценок алгоритма.

Доказательство. В случае задачи регрессии, $M = m(m-1)/2$ и

$$\begin{aligned} AUC &= \frac{2}{m(m-1)} \sum_{k,i=1}^m [\mathbf{Z}(k, i) = 0][\hat{\mathbf{Z}}(k, i) = 0] = \frac{2}{m(m-1)} \sum_{k,i=1}^m [y_k < y_i][p_k < p_i] = \\ &= 1 - \frac{4}{m(m-1)} \sum_{i=1}^{m-1} \sum_{k=i+1}^m [[y_k < y_i] \neq [p_k < p_i]] = \tau(\mathbf{y}, \mathbf{p}). \end{aligned}$$

□

1.2.3 Конусное представление предпочтений

В этом разделе введем понятие полиэдрального конуса \mathcal{X}_0 , соответствующего частично упорядоченному множеству X , рассмотрим его основные свойства и установим связь с матрицей предпочтений. Определение этого понятия необходимо для корректной линеаризации вектора предпочтений и построения модели предпочтений, которая будет рассмотрена в последующем разделе.

Определение 10. *Полиэдральный конус в \mathbb{R}^m — это множество \mathcal{X}_0 , такое что*

$$\mathcal{X}_0 = \{\boldsymbol{\chi} \mid \mathbf{A}\boldsymbol{\chi} \leq \mathbf{0}, \boldsymbol{\chi} \in \mathbb{R}^m\}, \quad (13)$$

для некоторой прямоугольной матрицы \mathbf{A} .

Пусть на множестве объектов $X = \{x_1, \dots, x_m\}$ задано отношение частичного порядка $z : z(x_i, x_k) = [x_i \succeq x_k]$. Введем определение конуса предпочтений, соответствующего множеству X с заданным на нем отношением z .

Определение 11. *Конусом предпочтений \mathcal{X}_0 для множества X с определенным на нем отношением частичного порядка $z : z = [x_i \succeq x_k]$ называется конус в пространстве \mathbb{R}_+^m , сохраняющий порядок относительно z :*

$$\mathcal{X}_0 = \{\boldsymbol{\chi} \in \mathbb{R}_+^m \mid x_i \succeq x_k \rightarrow \chi_i \geq \chi_k \quad \forall i, k = 1, \dots, m\}.$$

Замечание 1. *Среди всех матриц \mathbf{A} полиэдрального представления конуса \mathcal{X}_0 будем рассматривать прямоугольную матрицу размера $m^2 \times m$, где каждая строка соответствует паре элементов x_i, x_k и содержит максимум два ненулевых элемента:*

$$a_i = -1 \text{ и } a_k = 1, \text{ если } x_i \succeq x_k.$$

Согласно теореме Минковского-Вейла [15], каждый полиэдральный конус допускает представление через конечный набор порождающих. Другими словами, полиэдральный конус \mathcal{X}_0 представим в виде

$$\mathcal{X}_0 = \{\boldsymbol{\chi} \mid \boldsymbol{\chi} = \sum_{k=1}^r \lambda_k \boldsymbol{\zeta}_k, \lambda_k \geq 0\}, \quad (14)$$

где $\boldsymbol{\zeta}_k$ — порождающий элемент конуса \mathcal{X}_0 , а r — число порождающих элементов. В общем случае, количество порождающих r зависит от структуры отношения частичного порядка

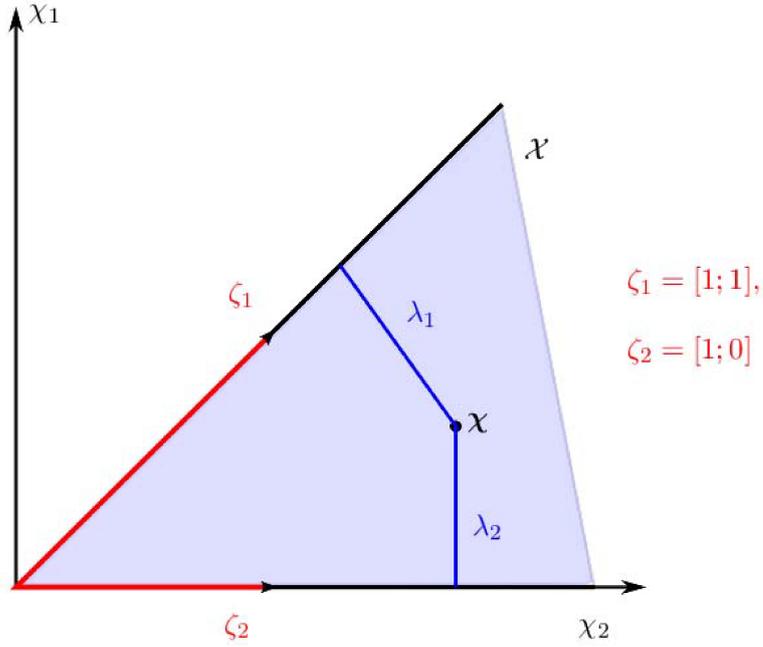


Рис. 1. Разложение точки по порождающим элементам конуса

множества X . На рис. 1 проиллюстрировано разложение точки χ по порождающим ζ_1, ζ_2 конуса \mathcal{X} . Точка $\chi \in \mathcal{X}$ является неотрицательной комбинацией направляющих конуса $\zeta_1 = [1; 1]$ и $\zeta_2 = [1; 0]$ с коэффициентами $\lambda_1, \lambda_2 \geq 0$.

Помимо конуса \mathcal{X}_0 , рассмотрим полиэдральный конус \mathcal{X} , представленный неотрицательной комбинацией столбцов матрицы предпочтений \mathbf{Z} , соответствующей множеству X :

$$\mathcal{X} = \{\mathbf{Z}\boldsymbol{\lambda} \mid \boldsymbol{\lambda} \geq \mathbf{0}\}. \quad (15)$$

Для конусов \mathcal{X}_0 и \mathcal{X} выполнена следующая теорема.

Теорема 1. $\mathcal{X} \subset \mathcal{X}_0$.

Доказательство. Для доказательства утверждения, что конус \mathcal{X} является подмножеством конуса \mathcal{X}_0 , надо доказать, что справедлива следующая импликация:

$$\boldsymbol{\lambda} \geq \mathbf{0} \quad \rightarrow \quad \mathbf{AZ}\boldsymbol{\lambda} \leq \mathbf{0}. \quad (16)$$

Для доказательства этого утверждения достаточно показать, что матрица \mathbf{AZ} не содержит положительных элементов.

Рассмотрим подробнее элемент (κ, k) матрицы \mathbf{AZ} . Согласно замечанию 1, строка κ матрицы \mathbf{A} соответствует паре элементов (i, i') и содержит два ненулевых элемента, -1 на месте i и 1 на месте i' , если $x_i \succeq x_{i'}$. При этом соответствующие элементы i и i' столбца k матрицы \mathbf{Z}

могут принимать следующие значения:

$$(\mathbf{Z}(i, k), \mathbf{Z}(i', k)) = \begin{cases} (1, 1), & \text{если } x_i \succeq x_k, x_{i'} \succeq x_k, \\ (0, 0), & \text{если } x_i \not\succeq x_k, x_{i'} \not\succeq x_k, \\ (1, 0), & \text{если } x_i \succeq x_k, x_{i'} \not\succeq x_k. \end{cases}$$

Таким образом, элемент (κ, k) матрицы \mathbf{AZ} может принимать лишь два значения:

$$[\mathbf{AZ}](\kappa, k) = \begin{cases} 0, & \text{если } x_i \succeq x_k, x_{i'} \succeq x_k \text{ или } x_i \not\succeq x_k, x_{i'} \not\succeq x_k, \\ -1, & \text{если } x_i \succeq x_k, x_{i'} \not\succeq x_k. \end{cases}$$

Таким образом, матрица \mathbf{AZ} не содержит положительных элементов, а, следовательно, справедлива импликация (16) и включение $\mathcal{X} \subset \mathcal{X}_0$. \square

Кроме того, для конусов \mathcal{X}_0 и \mathcal{X} выполнено следующее утверждение.

Лемма 6. *В случае линейного порядка на множестве X конусы \mathcal{X}_0 и \mathcal{X} совпадают.*

Доказательство. Для проведения доказательства перенумеруем элементы линейно упорядоченного множества X в порядке от большего к меньшему. Построим для отсортированного массива элементов $x_1, \dots, x_m \in X$ матрицу \mathbf{Z} , описывающую отношение порядка между каждой парой объектов. В этом случае, столбец матрицы \mathbf{Z} будет состоять из единичного столбца и, возможно, следующего за ним столбца из нулей, $[1, 1, \dots, 1, 1, 0, 0, \dots, 0, 0]^T$. Столбец, соответствующий наименьшему объекту из X , будет состоять из единиц. Допустим, что конусы \mathcal{X}_0 и \mathcal{X} не совпадают. Это означает, что существует бинарный вектор, являющийся порождающей конуса \mathcal{X}_0 и не являющийся столбцом матрицы \mathbf{Z} . Такой вектор обязательно имеет вид $[\dots, 0_i, 1_{i+1}, \dots]$, что нарушает отношение порядка на элементах x_i, x_{i+1} . Таким образом, порождающими конуса \mathcal{X}_0 являются только столбцы матрицы \mathbf{Z} . \square

Из доказанной теоремы 1 и леммы 6 следует, что конус предпочтений, представляемый разложением (15), позволяет описывать широкий класс преобразований упорядоченного множества. Разложение конуса (15) будет использовано в дальнейших разделах для построения моделей обучения по предпочтениям.

2 Построение модели предпочтений на наборе объектов, заданных порядковым описанием

В этом разделе рассматривается задача восстановления отношения предпочтения с использованием порядкового признакового описания. В качестве частного случая, рассматривается задача порядковой классификации с монотонными ограничениями [56, 22], которая заключается в следующем: необходимо построить монотонную функцию, отображающую множество объектов во множество меток классов, причем на множестве меток классов введено отношение строгого линейного порядка. Требование монотонности функции определяется экспертными соображениями о влиянии того или иного фактора на модель. Постановка подобного типа возникает в задачах теории принятия решений, информационного поиска [73, 83, 77], обучения по предпочтениям [41].

Новизна предлагаемого в данном разделе решения задачи состоит в следующем: порядковое признаковое описание объектов задается набором отношений частичного порядка [18]. Необходимо построить функцию, которая будет удовлетворять условию монотонности по каждому признаку.

Для построения такой функции предлагается несколько подходов. Первый подход основывается на построении линейной суперпозиции, соответствующих отношениям частичного порядка. Другими словами, строится взвешенный граф, каждое ребро которого является линейной комбинацией ребер заданных графов и выражает степень доминирования объектов [51, 11]. Веса линейной комбинации отыскиваются из соображений близости к экспертно заданному отношению предпочтения. На втором этапе по взвешенному графу определяется оптимальное отношение предпочтения на множестве объектов.

Второй подход основывается на вложении частично упорядоченного множества в линейное пространство. Для этого используется понятие конуса, соответствующего конечному частично упорядоченному множеству, определенному в подразделе 1.2.3. Для построения функции, областью определения которой является декартово произведение частично упорядоченных множеств, строится сумма Минковского конусов, соответствующих признакам. В качестве решения задачи восстановления предпочтения отыскивается точка, принадлежащая сумме конусов, и наилучшим образом коррелирующая с вектором целевой переменной. Для того, чтобы найти такую точку, предлагается параметрической и непараметрической методы.

Первый метод заключается в параметризации конусов, соответствующих порядковым признакам, согласно разложению (15). При таком разложении конусы представляются конеч-

ным набором порождающих элементов. С помощью такой параметризации задача формулируется в терминах неотрицательной линейной регрессии с большим количеством признаков. Предлагается метод сокращения размерности пространстве параметров на основе использования центральных точек конусов. Показано, что при таком подходе метод построения суммы конусов становится аналогичным методу построения взвешенного графа предпочтений, описанному ранее.

Второй метод основывается на непараметрическом представлении суммы Минковского конусов, соответствующих порядковым признакам [58]. Предлагается метод построения суммы Минковского конусов, заданных полиэдральным представлением. Решение задачи отыскивается в виде проекции на построенное множество.

Также в работе предлагаются альтернативные методы решения задачи. Одним из таких методов является метод криволинейной регрессии [30], заключающийся в поиске параметров, соответствующих элементам конечного частично упорядоченного множества. Другим методом является метод на основе копул [32], заключающийся в максимизации функции псевдоправдоподобия, которая остается неизменной при монотонных преобразованиях входящих в нее переменных.

2.1 Постановка задачи восстановления предпочтений

Рассматриваемая задача восстановления отношения предпочтения с использованием порядкового признакового описания заключается в следующем. Задано множество X , состоящее из m объектов, $X = \{x_1, \dots, x_m\}$. На множестве X задано n отношений предпочтения z_1, \dots, z_n , описывающих отношения доминирования между каждой парой объектов: $z_j(x_i, x_k) = [x_i \succeq_j x_k]$. В общем случае, каждое из отношений z_j является отношением частичного порядка.

Кроме того, на множестве X задано целевое отношение предпочтения $z_0 : z_0(x_i, x_k) = [x_i \succeq_0 x_k]$. В некоторых случаях, рассматриваемых отдельно, целевое отношение z_0 задается конечным набором меток Y , где каждому объекту x_i из X поставлена в соответствие метка $y_i \in Y : z_0(x_i, x_k) = [y_i \succeq y_k]$. В частности, рассматривается задача двухклассовой классификации, $Y = \{0, 1\}$, и задача порядковой классификации, $Y = \{l_1 \prec \dots \prec l_K\}$.

Требуется построить отображение $f : X \rightarrow \mathbb{R}$, сохраняющее порядок на парах объектов:

$$x_i \succeq_0 x_k \quad \rightarrow \quad f(x_i) \geq f(x_k). \quad (17)$$

Отметим, что задача в постановке (17) требует, чтобы отображение f восстанавливало порядок на всех парах объектов. В дальнейшем мы откажемся от этого строгого ограничения,

введя меру различия предпочтений $\rho(z_f, z_0)$, где z_f — порядок, задаваемый отображением f , $z_f(x_i, x_k) = [f(x_i) \geq f(x_k)]$, и наложим для отображения f требование минимизации функции различия на множестве X .

От отображения f требуется выполнение условия монотонности по всем отношениям z_1, \dots, z_n :

$$x_i \succeq x_k \quad \rightarrow \quad f(x_i) \geq f(x_k), \quad (18)$$

где отношение доминирования на объектах $x_i \succeq x_k$ выполняется, когда каждая из компонент объекта x_i доминируется соответствующей компонентой объекта x_k , $z_j(\mathbf{x}_i, \mathbf{x}_k) = 1$, $j = 1, \dots, n$.

2.2 Восстановление предпочтений с использованием матрицы попарного доминирования объектов

В этом разделе приведен базовый метод восстановления отношения предпочтения z_f , задаваемого отображением f на множестве объектов X , $z_f(x_i, x_k) = [f(x_i) \geq f(x_k)]$. Этот метод основывается на восстановлении матрицы предпочтений $\hat{\mathbf{Z}}$, каждый элемент которой является линейной комбинацией соответствующих элементов матриц $\mathbf{Z}_1, \dots, \mathbf{Z}_n$. При этом порядок, определяемый матрицей $\hat{\mathbf{Z}}$, оптимально приближает порядок, задаваемый зависимой переменной и соответствующим отношением порядка z_0 .

Мера различия предпочтений. Для того, чтобы определить понятие оптимальности отношения предпочтения z_f , задаваемого отображением f , введем понятие различия предпочтений $\rho(z_f, z_0)$. В качестве функции различия будем рассматривать значение отрицательной τ -корреляции между векторами предпочтений, которая, согласно формуле (9), пропорциональна норме Фробениуса между матрицами предпочтений \mathbf{Z}_f и \mathbf{Z}_0 :

$$S(f, X, z_0) = \rho(z_f, z_0) = \|\mathbf{Z}_f - \mathbf{Z}_0\|_F \propto -\tau(z_f, z_0). \quad (19)$$

Для оптимального восстановления предпочтений требуется найти такое отображение f , матрица \mathbf{Z}_f которого минимизирует отклонение $\rho(z_f, z_0)$. Для решения этой задачи предлагается приближенный алгоритм, описанный в следующем разделе.

Алгоритм оценивания матрицы предпочтений. Для нахождения отображения f , минимизирующего (19), предлагается построить матрицу попарного доминирования объектов $\hat{\mathbf{Z}}$ с использованием матриц $\mathbf{Z}_1, \dots, \mathbf{Z}_n$, соответствующих признаковым предпочтениям объектов.

Представим каждый элемент матрицы $\hat{\mathbf{Z}}$ линейной комбинацией элементов матриц $\mathbf{Z}_1, \dots, \mathbf{Z}_n$:

$$\hat{\mathbf{Z}} = \sum_{j=1}^n w_j \mathbf{Z}_j. \quad (20)$$

В этом случае, минимизация выражения (19) приводит к задаче нахождения оптимальных параметров $\hat{\mathbf{w}}$:

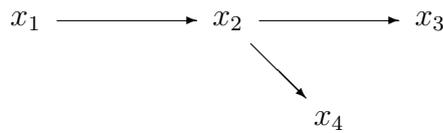
$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \sum_{i=1}^m \sum_{k=1}^m \left(\mathbf{Z}_0 - \sum_{j=1}^n w_j \mathbf{Z}_j \right)^2,$$

где \mathbf{Z}_0 — матрица частичного порядка, соответствующая вектору значений зависимой переменной \mathbf{y} . Другими словами, задача построения линейной комбинации матриц является задачей оптимального взвешивания графов предпочтений: матрица $\hat{\mathbf{Z}}$ является матрицей инцидентности взвешенного ориентированного графа предпочтений, вес каждого ребра которого определяет степень доминирования между соответствующей парой объектов.

Пример 3. Приведем пример линейной комбинации матриц $\mathbf{Z}_1, \mathbf{Z}_2$ согласованных предпочтений. Рассмотрим множество из четырех элементов $X = \{x_1, x_2, x_3, x_4\}$ и матрицу \mathbf{Z}_1 из примера 2,

$$\mathbf{Z}_1 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

соответствующую частичному порядку



В качестве матрицы \mathbf{Z}_2 , рассмотрим матрицу

$$\mathbf{Z}_2 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

соответствующую линейному порядку



Построим матрицу $\hat{\mathbf{Z}}$ как линейную комбинацию матриц $\mathbf{Z}_1, \mathbf{Z}_2$ с весами $w_1 = w_2 = \frac{1}{2}$:

$$\hat{\mathbf{Z}} = \frac{1}{2}\mathbf{Z}_1 + \frac{1}{2}\mathbf{Z}_2 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1/2 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Поскольку матрицы \mathbf{Z}_1 и \mathbf{Z}_2 соответствуют практически тождественным ранжированиям объектов, только один элемент матрицы $\hat{\mathbf{Z}}$ равен $\frac{1}{2}$, что может быть интерпретировано как неопределенность в доминировании лишь одной пары объектов x_3 и x_4 .

Пример 4. Приведем пример линейной комбинации матриц $\mathbf{Z}_1, \mathbf{Z}_2$ несогласованных предпочтений. Рассмотрим матрицу \mathbf{Z}_1 из предыдущего примера,

$$\mathbf{Z}_1 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

а в качестве матрицы \mathbf{Z}_2 рассмотрим матрицу

$$\mathbf{Z}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix},$$

соответствующую линейному порядку

$$x_1 \longleftarrow x_2 \longleftarrow x_3 \longleftarrow x_4$$

Аналогично предыдущему примеру, построим матрицу $\hat{\mathbf{Z}}$ как линейную комбинацию матриц $\mathbf{Z}_1, \mathbf{Z}_2$ с весами $w_1 = w_2 = \frac{1}{2}$:

$$\hat{\mathbf{Z}} = \frac{1}{2}\mathbf{Z}_1 + \frac{1}{2}\mathbf{Z}_2 = \begin{pmatrix} 1 & 1/2 & 1/2 & 1/2 \\ 1/2 & 1 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1 & 0 \\ 1/2 & 1/2 & 1/2 & 1 \end{pmatrix}.$$

Поскольку матрицы \mathbf{Z}_1 и \mathbf{Z}_2 соответствуют противоположным ранжированиям объектов, практически все элементы матрицы $\hat{\mathbf{Z}}$ равны $\frac{1}{2}$, что может быть интерпретировано как неопределенность в попарном доминировании объектов.

Восстановление отношения предпочтения. Для решения задачи (19) необходимо построить матрицу линейного порядка \mathbf{Z}_f , располагая оценкой матрицы $\hat{\mathbf{Z}}$, элементы которой принадлежат отрезку $[0, 1]$. Для того чтобы построить отображение f , воспользуемся идеей из [20]. Будем строить вектор оценок объектов \mathbf{f} в виде суммы столбцов матрицы $\hat{\mathbf{Z}} = [\hat{\mathbf{z}}_1, \dots, \hat{\mathbf{z}}_m]$

$$\mathbf{f} = \sum_{k=1}^m \hat{\mathbf{z}}_k. \quad (21)$$

В работе [20] доказан результат о том, что решение (21) в аналогичной постановке является 2-аппроксимацией исходной задачи поиска линейного порядка по матрице доминирования объектов.

Задача порядковой классификации. Используем оценку (20) матрицы $\hat{\mathbf{Z}}$ для задачи многоклассовой порядковой классификации, являющейся частным случаем задачи восстановления предпочтения с конечным множеством меток $Y = \{l_1, l_2, \dots, l_K\}$ с заданным на нем отношением строгого линейного порядка, $l_1 \prec l_2, \dots, \prec l_K$. Требуется построить монотонное отображение $f : X \rightarrow Y$, доставляющее минимум функции ошибки порядковой классификации.

Будем искать решение задачи в виде обобщения формулы (21) путем взвешивания столбцов матрицы доминирования объектов. Располагая оценкой матрицы частичного порядка $\hat{\mathbf{Z}}$, оценим параметры λ_k в линейной комбинации $\mathbf{f} = \hat{\mathbf{Z}}\boldsymbol{\lambda}$.

Воспользуемся методом порядковой логистической регрессии для оценки параметров λ_k . Классифицируем объект x_i следующим образом:

$$\hat{y}_i = \begin{cases} y_1, & \text{если } f(x_i) \leq \mu_1, \\ y_2, & \text{если } \mu_1 < f(x_i) \leq \mu_2, \\ \dots, & \\ y_K, & \text{если } \mu_K < f(x_i), \end{cases}$$

где $f(x_i)$ является оценкой объекта x_i , компонентой i вектора оценок \mathbf{f} , построенного по правилу $\mathbf{f} = \hat{\mathbf{Z}}\boldsymbol{\lambda}$.

Здесь μ_1, \dots, μ_K являются параметрами решающего правила, разделяющими числовую ось \mathbb{R} на области принадлежности меткам классов y_k . Оптимальные параметры подбираются путем минимизации функции потерь S ,

$$S(\boldsymbol{\lambda}, \boldsymbol{\mu}) = - \sum_{i=1}^m \log (\sigma(\mu_{y_i} - \boldsymbol{\lambda}^\top \hat{\mathbf{z}}_i) - \sigma(\mu_{y_{i-1}} - \boldsymbol{\lambda}^\top \hat{\mathbf{z}}_i)) \rightarrow \min,$$

где σ — сигмоидная функция,

$$\sigma(t) = 1/(1 + \exp(-t)),$$

а \mathbf{z}_i — вектор-столбец i матрицы $\hat{\mathbf{Z}}$.

Применение данного подхода будет продемонстрировано в экспериментальной части работы в разделе решения задачи порядковой классификации на данных UCI. Результаты классификации превосходят результаты многих известных алгоритмов.

2.3 Восстановление предпочтений с использованием конусов

В данном разделе будем строить отображение f из множества объектов во множество действительных чисел, $f : X \rightarrow \mathbb{R}$. Для удовлетворения условиям монотонности (18) будем строить множество значений отображения f в виде суперпозиции конусов частичного порядка $\mathcal{X}_1, \dots, \mathcal{X}_n$ (11), соответствующим отношениям порядка z_1, \dots, z_n . Кроме того, определим конус целевого предпочтения \mathcal{Y}_0 , соответствующий целевому отношению z_0 , определенному на X , и будем искать решение задачи в виде ближайших точек конусов.

2.3.1 Определение функции потерь с использованием конусного представления предпочтений

Предложим метод построения оптимизируемой функции потерь для решения задачи восстановления предпочтения (17). Для построения функции потерь, определенной на элементах множества X с заданным на нем отношением частичного порядка z_0 , используем представление этого множества в линейном пространстве путем использования конуса \mathcal{Y}_0 , соответствующего целевому отношению предпочтения z_0 .

Используем понятие конуса частичного порядка (11) для определения конуса целевого предпочтения.

Определение 12. *Конусом целевого предпочтения \mathcal{Y}_0 для множества X с определенным на нем целевым отношением частичного порядка z_0 ,*

$$z_0 = [x_i \succeq_0 x_k],$$

называется конус, сохраняющий порядок относительно z_0 :

$$\mathcal{Y}_0 = \{\chi \in \mathbb{R}_+^m \mid x_i \succeq_0 x_k \rightarrow \chi_i \geq \chi_k \quad \forall i, k = 1, \dots, m\}.$$

Будем строить отображение f из множества X во множество действительных чисел \mathbb{R} . Обозначим вектор значений функции f на множестве X за $\mathbf{f} \in \mathbb{R}^m$.

Определение 13. Функцией потерь $S(f, X, z_0)$ для задачи восстановления предпочтений (17) будем называть расстояние от вектора \mathbf{f} до конуса \mathcal{Y}_0 :

$$S(f, X, z_0) = \min_{\mathbf{y} \in \mathcal{Y}_0} \|\mathbf{f} - \mathbf{y}\|_2,$$

где $\|\cdot\|_2$ — стандартная евклидова норма.

Определение 14. Решением задачи восстановления предпочтений (17) будем называть вектор $\hat{\mathbf{f}}$, минимизирующий значений функции потерь $S(f, X, z_0)$:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f} \in \mathcal{F}} S(f, X, z_0). \quad (22)$$

Отметим, что конус целевого предпочтения \mathcal{Y}_0 , как и любой полиэдральный конус, допускает два представления: полиэдральное (13) и задаваемое конечным набором порождающих элементов (14). Порождающее представление определяется конусом \mathcal{Y} , который является подмножеством конуса \mathcal{Y}_0 , и задается разложением (15):

$$\mathcal{Y}_0 \supset \mathcal{Y} = \{\mathbf{Z}_0 \boldsymbol{\lambda}_0 \mid \boldsymbol{\lambda}_0 \geq \mathbf{0}\}, \quad (23)$$

где \mathbf{Z}_0 — матрица частичного порядка, соответствующая целевому отношению предпочтения z_0 , определенному на X .

В этом случае, функция потерь $S(f, X, z_0)$ записывается следующим образом:

$$S(f, X, z_0) = \min_{\boldsymbol{\lambda}_0 \geq \mathbf{0}} \|\mathbf{f} - \mathbf{Z}_0 \boldsymbol{\lambda}_0\|_2. \quad (24)$$

Функция потерь (24) является справедливой в случае, когда на множестве меток Y задано отношение нестрогого частичного порядка. В случае, когда на Y задано отношение строгого порядка (например, в задаче порядковой классификации, $Y = \{y_1 \succ \dots \succ y_K\}$), необходимо рассматривать лишь внутренние точки конуса \mathcal{Y} , то есть все элементы вектора $\boldsymbol{\lambda}_0$ должны быть положительными, $\boldsymbol{\lambda}_0 > \mathbf{0}$. Выполнение этого условия можно достигнуть введением дополнительного ограничения в задачу (24), например

$$S(f, X, z_0) = \min_{\boldsymbol{\lambda}_0 \geq \tau \mathbf{1}, \tau > 0} \|\mathbf{f} - \mathbf{Z}_0 \boldsymbol{\lambda}_0\|_2.$$

В данной работе мы будем рассматривать упрощенный подход на основе использования центральной точки конуса \mathcal{Y} , то есть случай $\boldsymbol{\lambda}_0 = \mathbf{1}$, и оптимизировать функцию потерь

$$S(f, X, z_0) = \|\mathbf{f} - \mathbf{Z}_0 \mathbf{1}\|_2. \quad (25)$$

Полиэдральное представление конуса \mathcal{Y}_0 . Помимо порождающего представления (23), согласно определению 13, конус целевого предпочтения \mathcal{Y}_0 допускает также полиэдральное представление следующего вида:

$$\mathcal{Y}_0 = \{\mathbf{y} \mid \mathbf{A}\mathbf{y} \leq 0\}, \quad (26)$$

где матрица \mathbf{A} имеет размеры $m^2 \times m$ и описывает отношение предпочтения между каждой парой объектов: в строке матрицы \mathbf{A} , соответствующей паре объектов i, k содержится не более двух ненулевых элементов: -1 на i -м месте и 1 на k -м месте, если для объектов i, k выполняется $x_i \succeq_0 x_k$.

Цель оптимизационной задачи (22) заключается в том, чтобы найти вектор \mathbf{f} , принадлежащий конусу \mathcal{Y}_0 . Согласно полиэдральному представлению (26), вектор \mathbf{f} должен выполнять условие

$$\mathbf{A}\mathbf{f} \leq \mathbf{0}.$$

При этом в случае строгого отношения порядка, определенного на элементах множества Y , рассматривается строгий вариант неравенства (26),

$$\mathbf{A}\mathbf{f} < \mathbf{0}. \quad (27)$$

Отметим, что векторное неравенство (27) эквивалентно предъявляемым условиям монотонности (18). Условие (27) переформулировывается в виде задачи минимизации количества положительных элементов вектора $\mathbf{A}\mathbf{f}$, что приводит к стандартной оптимизационной задаче:

$$S(f, X, z_0) = \sum_{x_i \succ_0 x_k} [f(x_i) \leq f(x_k)]. \quad (28)$$

В работе [62] для выполнения условия (27) рассматривается оптимизационная задача с использованием дополнительного вектора переменных \mathbf{t} :

$$\min_{\mathbf{t} \geq \mathbf{0}} \|\mathbf{t}\|_2, \quad \mathbf{A}\mathbf{f} < \mathbf{t}.$$

Кроме того, при использовании кусочно-линейной аппроксимации функции потерь и введении регуляризатора на вектор параметров модели, задача (28) переписывается в терминах стандартного функционала RankSVM (4).

Путем использования полиэдрального (13) и порождающего (14) представлений конуса целевого предпочтения \mathcal{Y}_0 сформулировано две оптимизационные задачи. Первая задача (24) заключается в минимизации расстояния до конуса \mathcal{Y} . Вторая задача (28) минимизирует количество невыполняемых ограничений-неравенств. Отметим, что стандартные функции потерь,

рассматриваемые в задачах попарного ранжирования [12, 83] являются аппроксимацией выражения (28) и соответствуют *попарному подходу* к задаче ранжирования. В свою очередь, задача поиска ближайших точек в конусах (24) соответствует и обобщает *поточечный подход* к задаче: параметры неотрицательной комбинации λ_0 в конусе \mathcal{U} играют роль разделителей числовой оси на конечное множество меток. В последующих разделах в основном будет использоваться вариант (25).

2.3.2 Линейная полиэдральная модель восстановления предпочтений

В этом подразделе определим линейную полиэдральную модель для решения задачи восстановления предпочтения. Модель основывается на построении суммы конусов предпочтений, соответствующих порядковым признакам.

Как и ранее, множество объектов $X = \{x_1, \dots, x_m\}$ задано порядковым описанием: на X задано n отношений частичного порядка z_1, \dots, z_n . Поставим в соответствие каждому из отношений z_j конус частичного порядка $\mathcal{X}_j \subset \mathbb{R}^m$, определяемый формулой (11).

Напомним определение *суммы Минковского* двух множеств.

Определение 15. *Суммой Минковского двух подмножеств $\mathcal{X}_1, \mathcal{X}_2$ линейного пространства является множество, состоящее из сумм всевозможных векторов из \mathcal{X}_1 и \mathcal{X}_2 :*

$$\mathcal{X}_1 \oplus \mathcal{X}_2 = \{\chi \mid \chi = \chi_1 + \chi_2, \chi_1 \in \mathcal{X}_1, \chi_2 \in \mathcal{X}_2\}.$$

Будем строить множество значений отображения f в виде суммы Минковского конусов частичных порядков $\mathcal{X}_1, \dots, \mathcal{X}_n$. Для этого определим множество значений линейной полиэдральной модели.

Определение 16. *Множеством значений линейной полиэдральной модели f на объектах, $\mathbf{f}(x_1, \dots, x_m)$, назовем сумму Минковского конусов частичного порядка $\mathcal{X}_1, \dots, \mathcal{X}_n$:*

$$\mathbf{f}(x_1, \dots, x_m) \in \mathcal{X}_1 \oplus \dots \oplus \mathcal{X}_n \subset \mathbb{R}^m. \quad (29)$$

Согласно разложению (15), каждый вектор конуса \mathcal{X}_j представим в виде неотрицательной комбинации столбцов матрицы частичного порядка \mathbf{Z}_j . Используя этот факт, множество значений линейной полиэдральной модели на объектах x_1, \dots, x_m представляется следующим образом.

Лемма 7. *Значением линейной полиэдральной модели f на объектах x_1, \dots, x_m является неотрицательная линейная комбинация вида*

$$\mathbf{f}(x_1, \dots, x_m) = \sum_{j=1}^n \mathbf{Z}_j \lambda_j, \lambda_j \geq 0. \quad (30)$$

Доказательство этой леммы следует непосредственно из формулы (15) и определения (29). Важным следствием разложения (30) является возможность определения линейной полиэдральной модели на произвольном объекте $x \in \mathbb{X}$.

Определение 17. *Линейной полиэдральной моделью $f : \mathbb{X} \rightarrow \mathbb{R}$ является отображение вида*

$$f(x) = \sum_{j=1}^n \mathbf{z}_j^\top \boldsymbol{\lambda}_j, \quad \boldsymbol{\lambda}_j \geq \mathbf{0}, \quad (31)$$

где \mathbf{z}_j — вектор индикаторов доминирования объектом x объектов x_1, \dots, x_m согласно отношению предпочтения z_j :

$$\mathbf{z}_j(k) = [x \succeq_j x_k],$$

где $\mathbf{z}_j(k)$ — k -й элемент вектора \mathbf{z}_j .

Следующие подразделы посвящены оценкам параметров модели (31) с использованием целевого отношения предпочтения z_0 и соответствующего ему конуса частичного порядка \mathcal{U}_0 . В подразделе 2.3.3 рассматривается прямой метод оценки параметров на основе решения задачи неотрицательной линейной регрессии. В подразделе 2.3.4 рассматривается метод регуляризации модели с использованием центральных точек конусов. Наконец, в подразделе 2.3.5 рассматривается метод построения суммы Минковского (29) с использованием полиэдрального представления конусов частичного порядка.

2.3.3 Прямая оценка параметров полиэдральной модели

Рассматривается задача минимизации функции потерь $S(f, X, z_0)$ вида (25), где f является линейной полиэдральной моделью вида (30):

$$\mathbf{f}(x_1, \dots, x_m) = \sum_{j=1}^n \mathbf{Z}_j \boldsymbol{\lambda}_j, \quad \boldsymbol{\lambda}_j \geq \mathbf{0},$$

для определения оптимальных параметров $\boldsymbol{\lambda}_j$.

В данном разделе предлагается прямая оценка параметров модели (30), построенной на основе суммы Минковского конусов частичных порядков. Для решения оптимизационной задачи разработан алгоритм оценки параметров модели.

Оптимальные параметры $\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_n$ должны минимизировать функцию потерь (25). С учетом явного вида модели \mathbf{f} задача поиска оптимальных параметров записывается следующим образом:

$$(\hat{\boldsymbol{\lambda}}_1, \dots, \hat{\boldsymbol{\lambda}}_n) = \arg \min_{\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_n \geq \mathbf{0}} \left\| \mathbf{Z}_0 \mathbf{1} - \sum_{j=1}^n \mathbf{Z}_j \boldsymbol{\lambda}_j \right\|_2. \quad (32)$$

Задача (32) является задачей *неотрицательной линейной регрессии* с матрицей $[\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n]$, являющейся присоединением всех матриц $\mathbf{Z}_1, \dots, \mathbf{Z}_n$, соответствующих порядковым признакам. Для решения этой задачи в [59, 16] предложен подход на основе *метода активного множества*. В данной работе предлагается использовать этот метод для итеративного решения задачи неотрицательной линейной регрессии последовательно для всех матриц $\mathbf{Z}_1, \dots, \mathbf{Z}_n$.

Обозначим за $\hat{\lambda}_j^t$ оценку j -го вектора параметров, полученную на шаге t . Алгоритм итеративной оценки параметров заключается в следующем.

1. На начальном шаге присвоить всем элементам вектора параметров равные значения $\hat{\lambda}_{jk}^0 = 1/m$. В этом случае каждый вектор $\hat{\lambda}_j^0$ является центральной точкой конуса \mathcal{X}_j .
2. На шаге t провести последовательную оценку параметров $\hat{\lambda}_j^t$ для всех $j = 1, \dots, n$. Для этого предполагаются фиксированными все векторы $\hat{\lambda}_1^t, \dots, \hat{\lambda}_{j-1}^t, \hat{\lambda}_{j+1}^{t-1}, \dots, \hat{\lambda}_n^{t-1}$, и вектор $\hat{\lambda}_j^t$ отыскивается путем минимизации функции потерь:

$$\hat{\lambda}_j^t = \arg \min_{\lambda_j \geq \mathbf{0}} \left\| \mathbf{Z}_0 \mathbf{1} - \sum_{j'=1}^{j-1} \mathbf{Z}_{j'} \hat{\lambda}_{j'}^t - \sum_{j'=j+1}^m \mathbf{Z}_{j'} \hat{\lambda}_{j'}^{t-1} - \mathbf{Z}_j \lambda_j \right\|_2. \quad (33)$$

Каждая из задач (33) является задачей неотрицательной линейной регрессии и решается методом активного множества.

2.3.4 Оценка параметров с использованием центральной точки конуса

В данном подразделе рассматривается задачи минимизации функции потерь (32) на сокращенном пространстве параметров. Для сокращения размерности пространства параметров используем понятие центральной точки полиэдрального конуса. Показано, что использование центральных точек приводит к задаче максимизации τ -корреляции (19).

Определение 18. *Центральной точкой полиэдрального конуса \mathcal{X} , задаваемого конечным набором порождающих элементов ζ_k ,*

$$\mathcal{X} = \left\{ \mathbf{x} \mid \mathbf{x} = \sum_{k=1}^r \lambda_k \zeta_k, \lambda_k \geq 0 \right\},$$

является точка $\bar{\mathbf{x}}$ вида

$$\bar{\mathbf{x}} = \frac{1}{r} \sum_{k=1}^r \zeta_k.$$

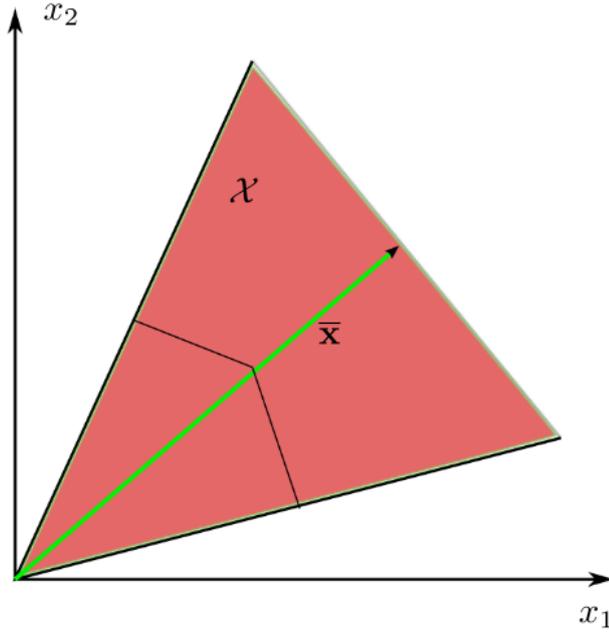


Рис. 2. Центральная точка конуса

Другими словами, центральной точкой конуса является центральная точка выпуклой комбинации его порождающих элементов. Изображение центральной точки показано на рис. 2.

Перепишем формулу (30) следующим образом:

$$\mathbf{f}(x_1, \dots, x_m) = \sum_{j=1}^n \mathbf{Z}_j \lambda_j = \sum_{k=1}^m \sum_{j=1}^n \mathbf{z}_{jk} \lambda_{jk},$$

где \mathbf{z}_{jk} — столбец k матрицы \mathbf{Z}_j . Выделив в этой комбинации переменные $w_j = \sum_{k=1}^m \lambda_{jk}$, получим

$$\mathbf{f}(x_1, \dots, x_m) = \sum_{k=1}^m \sum_{j=1}^n w_j \mathbf{z}_{jk} \lambda'_{jk},$$

где $\lambda'_{jk} = \frac{\lambda_{jk}}{w_j}$ для всех ненулевых w_j . При этом выполняются условия нормированности параметров λ'_{jk} , а также неотрицательности всех параметров:

$$\sum_{j=1}^n \lambda'_{jk} = 1, \quad \lambda'_j \geq \mathbf{0}, \quad w_j \geq 0.$$

Для дальнейшего удобства переобозначим параметры $\lambda_{jk} := \lambda'_{jk}$.

Рассмотрим конус \mathcal{X}_k^w , порождаемый элементами $w_j \mathbf{z}_{jk}$. Для снижения размерности пространства параметров, заменим в модели f конус \mathcal{X}_k^w его центральной точкой. В этом случае для модели \mathbf{f} получим следующее выражение:

$$\mathbf{f}(\mathbf{x}_1, \dots, \mathbf{x}_m) = \sum_{k=1}^m \lambda_k \mathbf{z}_{jk}^w,$$

где

$$\mathbf{z}_{jk}^w = \sum_{j=1}^n w_j \mathbf{z}_{jk}, \quad \sum_{k=1}^m \lambda_k = 1.$$

Таким образом, доказана следующая теорема о регуляризации линейной полиэдральной модели с помощью центральных точек конусов.

Теорема 2. При замене конуса \mathcal{X}_k^w его центральной точкой $\bar{\mathbf{x}}_k^w = \frac{1}{n} \sum_{j=1}^n w_j \mathbf{z}_{jk}$ линейная полиэдральная модель (30) принимает следующий вид:

$$\mathbf{f}(\mathbf{x}_1, \dots, \mathbf{x}_m) = \hat{\mathbf{Z}}\boldsymbol{\lambda}, \quad \hat{\mathbf{Z}} = \sum_{j=1}^n w_j \mathbf{Z}_j, \quad (34)$$

при ограничениях

$$w_j \geq 0, \quad \sum_{k=1}^m \lambda_k = 1, \quad \boldsymbol{\lambda} \geq \mathbf{0}.$$

Матрица $\hat{\mathbf{Z}}$ описывает отношение частичного порядка между каждой парой объектов выборки. Элемент этой матрицы $\hat{\mathbf{Z}}(i, k)$ может быть интерпретирован как степень достоверности того, что объект x_i доминирует объект x_k , $x_i \succeq x_k$.

Минимизация функции потерь для поиска оптимальных параметров. Оптимизационная задача (32) с учетом регуляризации (34) представляется следующим образом:

$$(\hat{\mathbf{w}}, \hat{\boldsymbol{\lambda}}) = \arg \min_{\mathbf{w}, \boldsymbol{\lambda}} \left\| \mathbf{Z}_0 \mathbf{1} - \left(\sum_{j=1}^n w_j \mathbf{Z}_j \right) \boldsymbol{\lambda} \right\|_2, \quad (35)$$

$$w_j \geq 0, \quad \sum_{k=1}^m \lambda_k = 1, \quad \boldsymbol{\lambda} \geq \mathbf{0}.$$

Справедлива следующая лемма о том, функция потерь (35) ограничена сверху функцией потерь (19), являющейся отрицательной τ -корреляцией.

Лемма 8. $\min_{\boldsymbol{\lambda} \geq \mathbf{0}, \|\boldsymbol{\lambda}\|_1=1} \|\mathbf{Z}_0 \mathbf{1} - m \hat{\mathbf{Z}} \boldsymbol{\lambda}\|_2 \leq \|\mathbf{Z}_0 - \hat{\mathbf{Z}}\|_F.$

Доказательство.

$$\min_{\boldsymbol{\lambda} \geq \mathbf{0}, \|\boldsymbol{\lambda}\|_1=1} \|\mathbf{Z}_0 \mathbf{1} - m \hat{\mathbf{Z}} \boldsymbol{\lambda}\|_2 = \min_{\boldsymbol{\lambda} \geq \mathbf{0}, \|\boldsymbol{\lambda}\|_1=m} \|\mathbf{Z}_0 \mathbf{1} - \hat{\mathbf{Z}} \boldsymbol{\lambda}\|_2 \leq \|\mathbf{Z}_0 \mathbf{1} - \hat{\mathbf{Z}} \mathbf{1}\|_2 \leq \|\mathbf{Z}_0 - \hat{\mathbf{Z}}\|_F,$$

где последнее выражение является аппроксимацией отрицательной τ -корреляции, согласно формуле (9). \square

Доказанное утверждение позволяет свести задачу (35) к задаче восстановления отношения предпочтения на основе максимизации τ -корреляции из раздела 2.2.

Оценку параметров модели (35) проведем в два этапа. На первом этапе оценим матрицу частичного порядка $\hat{\mathbf{Z}}$, являющуюся линейной комбинацией матриц $\mathbf{Z}_1, \dots, \mathbf{Z}_n$,

$$\hat{\mathbf{Z}} = \sum_{j=1}^n w_j \mathbf{Z}_j,$$

путем минимизации функции ошибки

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \sum_{i=1}^m \sum_{k=1}^m \left(\mathbf{Z}_0 - \sum_{j=1}^n w_j \mathbf{Z}_j \right)^2.$$

На втором этапе оценим параметры λ линейной комбинации $\hat{\mathbf{Z}}\lambda$ методом неотрицательной линейной регрессии.

2.3.5 Построение модели с использованием полиэдрального представления конусов

В данном разделе рассмотрим метод построения суммы Минковского (29) конусов $\mathcal{X}_1, \dots, \mathcal{X}_n$, задаваемых полиэдральным представлением с прямоугольными матрицами $\mathbf{A}_1, \dots, \mathbf{A}_n$. Как было показано выше, каждый полиэдральный конус определяется системой линейных неравенств

$$\mathcal{X}_j = \{\mathbf{x} | \mathbf{A}_j \mathbf{x} \leq 0\}, \quad j = 1, \dots, m,$$

задаваемых матрицей \mathbf{A}_j размера $m^2 \times m$, где каждая строка \mathbf{a} матрицы \mathbf{A}_j соответствует паре элементов x_i, x_k и содержит максимум два ненулевых элемента: $a_i = -1$ и $a_k = 1$, если $x_i \succeq x_k$. В общем случае множество \mathbf{X}_j является многогранником, определяемым матрицей \mathbf{A}_j . В случае, когда отношение предпочтения z_j , задающее конус \mathcal{X}_j , является отношением линейного порядка, матрица \mathbf{A}_j имеет следующий вид:

$$\mathbf{A}_j = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -1 \end{pmatrix}.$$

Рассматривается линейная полиэдральная модель (29) восстановления отношения предпочтения. Это означает, что множество значений модели \mathcal{X} представляется суммой Минковского конусов $\mathcal{X}_1, \dots, \mathcal{X}_n$:

$$\mathcal{X} = \mathcal{X}_1 \oplus \mathcal{X}_2 \oplus \dots \oplus \mathcal{X}_n.$$

Для построения вектора оценок объектов требуется построить допустимое множество значений, являющееся суммой Минковского выпуклых многогранников $\mathcal{X}_1, \dots, \mathcal{X}_n$. Воспользуемся методом точного построения суммы Минковского двух многогранников [84, 37], заданных системами линейных неравенств. Ниже кратко опишем рассматриваемый метод.

Пусть дано два выпуклых многогранника, задающихся системами неравенств:

$$\mathcal{X}_1 = \{\mathbf{x}_1 | \mathbf{A}_1 \mathbf{x}_1 \leq \mathbf{b}_1\}, \quad \mathcal{X}_2 = \{\mathbf{x}_2 | \mathbf{A}_2 \mathbf{x}_2 \leq \mathbf{b}_2\}.$$

Их суммой Минковского является вектор \mathbf{x} , являющийся решением системы:

$$\begin{cases} \mathbf{x} - \mathbf{x}_1 - \mathbf{x}_2 = 0, \\ \mathbf{A}_1 \mathbf{x}_1 \leq \mathbf{b}_1, \\ \mathbf{A}_2 \mathbf{x}_2 \leq \mathbf{b}_2, \end{cases}$$

которая заменой переменной $\mathbf{x}_1 = \mathbf{x} - \mathbf{x}_2$ преобразуется в систему:

$$\begin{cases} \mathbf{A}_1 \mathbf{x} - \mathbf{A}_2 \mathbf{x}_2 \leq \mathbf{b}_1, \\ \mathbf{A}_2 \mathbf{x}_2 \leq \mathbf{b}_2. \end{cases} \quad (36)$$

Справедливо следующее утверждение: вектор $\mathbf{x} \in \mathcal{X}$ тогда и только тогда, когда найдется вектор \mathbf{x}_2 , удовлетворяющий системе (36).

Таким образом, задача поиска вектора \mathbf{x} сводится к решению системы линейных неравенств:

$$\mathbf{C} \mathbf{x}_2 \leq \mathbf{d}, \quad \text{где } \mathbf{C} = \begin{pmatrix} -\mathbf{A}_1 \\ \mathbf{A}_2 \end{pmatrix}, \quad \text{а } \mathbf{d} = \begin{pmatrix} \mathbf{b}_1 - \mathbf{A}_1 \mathbf{x} \\ \mathbf{b}_2 \end{pmatrix}. \quad (37)$$

Для решения этой системы используем следующий вариант леммы Минковского-Фаркаша [84].

Лемма 9. *Для того чтобы система линейных неравенств $\mathbf{C} \mathbf{x}_2 \leq \mathbf{d}$ имела решение, необходимо и достаточно, чтобы для любого такого вектора \mathbf{u} , что*

$$\mathbf{C}^T \mathbf{u} = 0, \quad \mathbf{u} \geq \mathbf{0},$$

выполнялось условие

$$\mathbf{d}^T \mathbf{u} \geq 0.$$

Обозначим за \mathbf{V} фундаментальную систему решений для уравнения с матрицей \mathbf{C}^T , причем

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \end{pmatrix},$$

где \mathbf{V}_1 — часть матрицы \mathbf{V} , строки которой соответствуют неравенствам матрицы $-\mathbf{A}_1$, а \mathbf{V}_2 — часть матрицы \mathbf{V} , строки которой соответствуют неравенствам матрицы \mathbf{A}_2 . Тогда условие $\mathbf{d}^\top \mathbf{u} \geq 0$ переписывается в виде:

$$\mathbf{V}_1^\top (\mathbf{b}_1 - \mathbf{A}_1 \mathbf{x}) + \mathbf{V}_2^\top \mathbf{b}_2 \geq 0.$$

Таким образом, полагая

$$\mathbf{A} = \mathbf{V}_1^\top \mathbf{A}_1, \quad \mathbf{b} = \mathbf{V}_1^\top \mathbf{b}_1 + \mathbf{V}_2^\top \mathbf{b}_2,$$

мы получим параметры \mathbf{A}, \mathbf{b} системы неравенств, описывающей сумму Минковского $\mathcal{X}_1 \oplus \mathcal{X}_2$. Отдельной задачей является нахождение фундаментальной системы решений \mathbf{V} с неотрицательными столбцами. Метод отыскания такой системы описан в [84].

Принимая во внимание то, что вектором \mathbf{b} для конуса является нулевой вектор, получаем следующий результат для построения суммы Минковского.

Алгоритм построения суммы Минковского полиэдральных конусов. Суммой Минковского полиэдральных конусов $\mathcal{X}_1 \oplus \mathcal{X}_2 \oplus \dots \oplus \mathcal{X}_n$, заданных матрицами $\mathbf{A}_1, \dots, \mathbf{A}_n$, является конус

$$\mathcal{X} = \{\mathbf{x} \mid \mathbf{A}^{(n)} \mathbf{x} \leq \mathbf{0}\},$$

задаваемый матрицей

$$\mathbf{A}^{(n)} = \mathbf{V}_{n-1}^\top \mathbf{A}^{(n-1)},$$

где матрица $\mathbf{A}^{(n-1)}$ определяет конус, являющийся суммой Минковского конусов $\mathcal{X}_1 \oplus \mathcal{X}_2 \oplus \dots \oplus \mathcal{X}_{n-1}$, а матрица \mathbf{V}_{n-1} является частью фундаментальной системы решений для уравнения с матрицей $\begin{pmatrix} -\mathbf{A}^{(n-1)} \\ \mathbf{A}_n \end{pmatrix}$, соответствующей неравенствам матрицы $-\mathbf{A}^{(n-1)}$.

В качестве решения задачи восстановления предпочтения будем рассматривать в множестве \mathcal{X} точку $\hat{\mathbf{y}}$, ближайшую к $\mathbf{y}_0 = \mathbf{Z}_0 \mathbf{1}$. Проекцией точки \mathbf{y} на множество \mathcal{X} называется вектор $P(\mathbf{y}) \in \mathcal{X}$, удовлетворяющий условию

$$P_{\mathcal{X}}(\mathbf{y}) = \arg \min_{\mathbf{x} \in \mathcal{X}} \|\mathbf{y} - \mathbf{x}\|_2.$$

Построив область значения модели \mathcal{X} , определим вектор предпочтений $\hat{\mathbf{y}}$ как проекцию экспертной оценки \mathbf{y}_0 на \mathcal{X} :

$$\hat{\mathbf{y}} = P_{\mathcal{X}}(\mathbf{y}_0). \quad (38)$$

Проекция вектора \mathbf{y}_0 на допустимое множество значений \mathcal{X} проиллюстрирована на рис. 3.

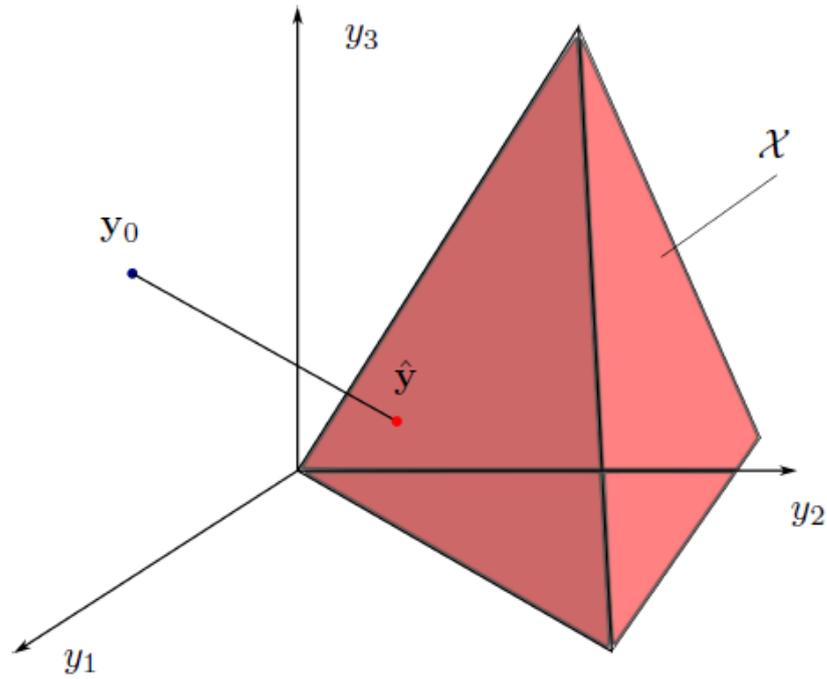


Рис. 3. Проекция точки y_0 на конус \mathcal{X}

2.4 Восстановление предпочтений методом криволинейной регрессии

В этом разделе предлагается альтернативный метод решения задачи восстановления предпочтения с порядковыми признаками. Метод основывается на монотонной коррекции исходного признакового описания. Исходному порядковому признаку, заданному на наборе объектов, ставится в соответствие вектор в евклидовом пространстве размерности m , отыскиваемый из соображений его монотонности по компонентам и оптимальности функции потерь. Метод, изложенный в этом разделе, описан в работе [30].

Постановка задачи. Как и ранее, задано множество объектов $X = \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$, на котором заданы отношения предпочтения z_1, \dots, z_n и целевое отношение предпочтения z_0 . В этом разделе будем рассматривать случай, когда каждое из отношений z_j задано конечным множеством меток $X_j = \{l_1 \succ \dots \succ l_{k_j}\}$. При этом пространство объектов \mathbb{X} является декартовым произведением множеств X_1, \dots, X_n , $x_{ij} \in X_j$, а отношение предпочтения z_j представляется в виде $z_j(x_i, x_k) = [x_{ij} \succeq x_{kj}]$. В этом случае, определена матрица плана $\mathbf{X} = [x_{ij}]$. Целевое отношение предпочтения z_0 задано конечным множеством меток Y , и выполняется соотношение $z_0(x_i, x_k) = [y_i \succeq y_k]$.

Требуется построить отображение вида $f : X_1 \times \dots \times X_n \rightarrow Y$.

Модель криволинейной регрессии. Криволинейная модель $f(\mathbf{w}, \mathbf{x}_i)$ имеет вид

$$f(\mathbf{w}, \mathbf{x}_i) = \nu(\mathbf{b}_0, h(\mathbf{w}, \mathbf{x}_i)), \quad (39)$$

$$h(\mathbf{w}, \mathbf{x}_i) = \sum_{j \in \mathcal{J}} u_j g(\mathbf{b}_j, x_{ij}). \quad (40)$$

где вектор параметров $\mathbf{w} = [\mathbf{b}_0; \mathbf{b}_1; \dots; \mathbf{b}_n; \mathbf{u}] = [\mathbf{b}_0^\top, \mathbf{b}_1^\top, \dots, \mathbf{b}_n^\top, \mathbf{u}^\top]^\top$ состоит из векторов \mathbf{b}_j — параметров монотонной коррекции j -го признака χ_j и весовых коэффициентов признаков $\mathbf{u} = [u_1, \dots, u_j, \dots, u_n]^\top$. Функция g монотонной коррекции задана следующим образом:

$$g(\mathbf{b}_j, X_j) : X_j \mapsto \mathbf{b}_j = \begin{cases} l_1 \mapsto b_{j1}, \\ l_2 \mapsto b_{j2}, \\ \dots \\ l_{k_j} \mapsto b_{jk_j}. \end{cases}$$

При этом соблюдается условие монотонности параметров,

$$\text{Ord}(\mathbf{b}_j) : 0 < b_{j1} < b_{j2} < \dots < b_{jk_j} < 1 \quad \text{для } j = 1, \dots, n \quad \text{и} \quad (41)$$

$$\text{Ord}(\mathbf{b}_0) : b_{01} < b_{02} < \dots < b_{0k_0}.$$

Функция $\nu(\mathbf{b}_0, h(\mathbf{w}, \mathbf{x}_i))$ определяет для числа $h(\mathbf{w}, \mathbf{x}_i)$ ближайшую по модулю компоненту вектора \mathbf{b}_0 :

$$\nu(\mathbf{b}_0, h(\mathbf{w}, \mathbf{x}_i)) = \arg \min_{j \in \mathcal{J}} |b_{0j} - h(\mathbf{w}, \mathbf{x}_i)|.$$

Введя обозначение для матрицы скорректированных значений признаков

$$\mathbf{G} = [g_{ij}] = [g(\mathbf{b}_j, x_{ij})], \quad i \in \mathcal{I}, j \in \mathcal{J},$$

перепишем (39) и (40) следующим образом:

$$f(\mathbf{w}, \mathbf{x}_i) = \nu(\mathbf{b}_0, [\mathbf{G}\mathbf{u}]_i). \quad (42)$$

Назначим функцией ошибки модели сумму квадратов регрессионных остатков,

$$S(f, X, z_0) = \|\mathbf{G}\mathbf{u} - \mathbf{b}_0\|_2^2.$$

Оценивание параметров модели. Оценивание параметров модели выполняется итеративно в три этапа. Сначала при фиксированных значениях векторов $\hat{\mathbf{b}}_0, \dots, \hat{\mathbf{b}}_n$ оцениваются весовые коэффициенты

$$\hat{\mathbf{u}} = \arg \min_{\mathbf{u} \in \mathbb{R}^n} S([\hat{\mathbf{b}}_0; \dots; \hat{\mathbf{b}}_n; \mathbf{u}]^\top).$$

Затем при фиксированных значениях коэффициентов $\hat{\mathbf{u}}$ оцениваются параметры монотонной коррекции

$$[\mathbf{b}_1; \dots; \mathbf{b}_n] = \arg \min_{\text{Ord}(\mathbf{b}_1), \dots, \text{Ord}(\mathbf{b}_n)} S([\hat{\mathbf{b}}_0; \dots; \hat{\mathbf{b}}_n; \mathbf{u}]^\top)$$

с учетом требования монотонности значений этих параметров. На последнем этапе оценивается вектор \mathbf{b}_0 :

$$\mathbf{b}_0 = \arg \min_{\text{Ord}(\mathbf{b}_0)} S([\hat{\mathbf{b}}_0; \dots; \hat{\mathbf{b}}_n; \mathbf{u}]^\top).$$

Итерации выполняются до стабилизации функции ошибки S .

Рассмотрим эти три этапа более подробно. За начальное приближение примем столбцы матрицы \mathbf{X} , вложенной в линейное пространство

$$\hat{\mathbf{G}} = [\mathbf{g}(\hat{\mathbf{b}}_1, X_1), \dots, \mathbf{g}(\hat{\mathbf{b}}_n, X_n)] = \mathbf{X},$$

поскольку, как было сказано выше, $g = \text{id}$. Таким образом, векторы $\hat{\mathbf{b}}_0, \hat{\mathbf{b}}_1, \dots, \hat{\mathbf{b}}_n$ в начальном приближении в качестве элементов содержат элементы множеств X_1, \dots, X_n .

Шаг 1. Найдем $\hat{\mathbf{u}}$ при фиксированных $\hat{\mathbf{b}}_0, \dots, \hat{\mathbf{b}}_n$:

$$\hat{\mathbf{u}} = \arg \min_{\mathbf{u}} \|\hat{\mathbf{b}}_0 - \hat{\mathbf{G}}\mathbf{u}\|.$$

Решение на шаге 1 имеет вид:

$$\hat{\mathbf{u}} = (\hat{\mathbf{G}}^T \hat{\mathbf{G}})^{-1} \hat{\mathbf{G}}^T \hat{\mathbf{b}}_0.$$

Шаг 2. При фиксированных $\hat{\mathbf{b}}_0, \hat{\mathbf{u}}$ оценим скорректированную матрицу описаний

$$\mathbf{G} = [\mathbf{g}(\mathbf{b}_1, X_1), \dots, \mathbf{g}(\mathbf{b}_n, X_n)] = [\mathbf{g}_1, \dots, \mathbf{g}_n].$$

Для каждого $\mathbf{g}_j \in \mathbb{R}^m$ будем вычислять вектор $\hat{\mathbf{g}}_j$, являющийся монотонной коррекцией исходного вектора \mathbf{g}_j :

$$\left\{ \begin{array}{l} [\hat{\mathbf{g}}_1, \dots, \hat{\mathbf{g}}_n] = \arg \min \|\hat{\mathbf{b}}_0 - \mathbf{G}\hat{\mathbf{u}}\|_2, \\ \text{из } g_{ij_1} \leq g_{ij_2} \text{ следует } \hat{g}_{ij_1} \leq \hat{g}_{ij_2} \quad i \in \mathcal{I}, j_1, j_2 \in \mathcal{J}, \\ g_{ij} \in [0, 1] \quad i \in \mathcal{I}, j \in \mathcal{J}, \text{ согласно (41)}. \end{array} \right.$$

За \mathcal{I} и \mathcal{J} обозначены индексные множества объектов и признаков, соответственно: $\mathcal{I} = \{1, \dots, m\}$, $\mathcal{J} = \{1, \dots, n\}$. По векторам $\hat{\mathbf{g}}_1, \dots, \hat{\mathbf{g}}_n$ однозначно восстанавливаются векторы $\hat{\mathbf{b}}_1, \dots, \hat{\mathbf{b}}_n$ как упорядоченные векторы, содержащие различные элементы $\hat{\mathbf{g}}_1, \dots, \hat{\mathbf{g}}_n$. Для решения этой задачи используется алгоритм градиентного спуска, описанный в [9].

Шаг 3. Наконец, при фиксированных $\hat{\mathbf{b}}_1, \dots, \hat{\mathbf{b}}_n, \hat{\mathbf{u}}$ оценим вектор \mathbf{b}_0 :

$$\hat{\mathbf{b}}_0 = \arg \min_{\text{Ord}(\mathbf{b}_0)} \|\mathbf{b}_0 - \hat{\mathbf{G}}\hat{\mathbf{u}}\|_2.$$

Снижение размерности пространства параметров. Для снижения размерности пространства предлагается метод выбора наиболее информативных признаков. Множество индексов признаков, включенных в модель, назовем активным набором и обозначим $\mathcal{A} \subseteq \mathcal{J}$.

Поставим задачу выбора наиболее информативных признаков следующим образом. Разобьем выборку объектов X на две подвыборки, обучающую и тестовую. Обозначим индексы элементов этих подвыборок соответственно $\mathcal{L} \sqcup \mathcal{T} = \mathcal{I}$. Для некоторого активного набора признаков \mathcal{A} найдем на обучающей подвыборке $X_{\mathcal{L}}$ оптимальные, согласно заданной функции ошибки S , параметры $\hat{\mathbf{w}}_{\mathcal{A}}$,

$$\hat{\mathbf{w}}_{\mathcal{A}} = \arg \min_{\mathbf{w}} S(\mathbf{w}_{\mathcal{A}} | X_{\mathcal{L}}).$$

Затем выберем наиболее информативные признаки — активный набор $\hat{\mathcal{A}}$ по всем поднаборам индексов признаков $\mathcal{A} \subseteq \mathcal{J}$, доставляющий на тестовой выборке $X_{\mathcal{T}}$ минимум функции ошибки:

$$\hat{\mathcal{A}} = \arg \min_{\mathcal{A} \subseteq \mathcal{J}} S(\hat{\mathbf{w}}_{\mathcal{A}} | X_{\mathcal{T}}).$$

Для выбора наиболее информативного подмножества признаков используется итеративный алгоритм добавления признаков.

На первом шаге этого алгоритма принимается активное множество информативных признаков $\hat{\mathcal{A}} = \emptyset$. На каждом следующем шаге к множеству $\hat{\mathcal{A}}$ добавляется признак с индексом \hat{j} , такой что

$$\hat{j} = \arg \min_{j \in \mathcal{J} \setminus \hat{\mathcal{A}}} S(\hat{\mathbf{w}}_{\hat{\mathcal{A}} \cup \{j\}} | X_{\mathcal{T}}).$$

Эта процедура продолжается итеративно до тех пор, пока значение функции ошибки S на контрольной выборке $X_{\mathcal{T}}$ не достигнет минимума.

2.5 Восстановление предпочтений с использованием функции копулы

Аналогично предыдущему разделу, рассмотрим случай, когда каждое из отношений z_j задано конечным множеством меток $X_j = \{l_1 \succ \dots \succ l_{k_j}\}$. Пространство объектов X является декартовым произведением множеств X_1, \dots, X_n , $x_{ij} \in X_j$, а отношение предпочтения z_j представляется в виде $z_j(x_i, x_k) = [x_{ij} \succeq x_{kj}]$. Определена матрица плана $\mathbf{X} = [x_{ij}]$. Будем обозначать объект из множества X вектором $\mathbf{x} \in X$. Целевое отображение z_0 задается конечным набором меток Y .

В данном разделе принимается вероятностная структура данных. Случайные величины, соответствующие признакам, будем обозначать за ξ_1, \dots, ξ_n . Для решения задачи восстановления предпочтения предлагается альтернативный алгоритм на основе копулы [66, 39] — функции, являющейся многомерными параметрическими функциями распределения равномерно распределенных случайных величин. Использование копул для оценки распределений помогает справиться с проблемой ранговости критериев. Для оценки параметра копулы необходимо знать только ранговые соотношения между величинами, а не их абсолютные значения.

Предлагается алгоритм выбора наиболее информативных критериев, использующий параметры копулы в качестве показателя информативности. Эти параметры обладают свойством монотонности: чем больше параметр копулы, тем больше ранговая связь между двумя случайными величинами. Предлагается выбирать те критерии, которые имеют наибольшую ранговую корреляцию с целевым отношением предпочтения.

Требуется построить отображение вида $f : X_1 \times \dots \times X_n \rightarrow Y$, минимизирующее функционал среднего риска. Минимум среднего риска достигается отображением

$$f(\mathbf{x}) = \arg \max_{y \in Y} P(y|\mathbf{x}_i), \quad (43)$$

где $P(y|\mathbf{x}_i)$ — апостериорная вероятность класса $y \in Y$ для объекта \mathbf{x} . Эта вероятность является условной по \mathbf{x} . Для оценки апостериорной вероятности $P(y|\mathbf{x}_i)$ будем использовать копулы.

Свойства копул, используемые для оценки условной вероятности.

Определение 19. *Функция $C : [0, 1]^n \rightarrow [0, 1]$ называется копулой размерности n , если выполняются следующие условия:*

$$C(u_1, \dots, u_{j-1}, 0, u_{j+1}, \dots, u_n) = 0,$$

$$C(1, \dots, 1, u, 1, \dots, 1) = u,$$

$$\int_B dC(u) \geq 0, \quad \text{где } B = \prod_{i=1}^d [a_i, b_i] \subseteq [0, 1]^d.$$

Выполнение этих свойств означает, что функция C является функцией распределения многомерной случайной величины $[u_1, \dots, u_n]^T$, такой, что одномерное распределение каждого из u_j равномерно на интервале $[0, 1]$.

Важным фактом, позволяющим применять копулы для построения регрессионных моделей, является следующая теорема.

Теорема Склера. Многомерная функция распределения случайной величины:

$$H(x_1, \dots, x_n) = P[\xi_1 \leq x_1, \dots, \xi_n \leq x_n]$$

случайного вектора (ξ_1, \dots, ξ_n) с одномерными функциями распределения

$$F_i(x) = P[\xi_i \leq x_i]$$

может быть записана в виде:

$$H(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)).$$

Таким образом, для оценивания совместного распределения H случайных величин ξ_1, \dots, ξ_n достаточно оценить их одномерные распределения $F_i(x_i)$ и функцию копулы, связывающую эти случайные величины.

Следующая теорема утверждает, что функция копулы не изменяется при действии на случайные величины любых монотонных преобразований.

Пусть ξ_1, ξ_2 — две случайные величины с совместной функцией распределения $H(x, y)$. Пусть также ϕ, ψ — две монотонных функции, преобразующие случайные величины ξ_1 и ξ_2 в

$$\mu_1 = \phi(\xi_1), \quad \mu_2 = \psi(\xi_2)$$

с совместной функцией распределения $H'(\mu_1, \mu_2)$. Тогда копула, связывающая случайные величины μ_1 и μ_2 :

$$C'(F'(t_1), G'(t_2)) = H'(t_1, t_2) = C(F(t_1), G(t_2)),$$

то есть,

$$C' = C.$$

Таким образом, чтобы оценить функцию копулы, описывающую связь между случайными величинами ξ_1, \dots, ξ_n , достаточно знать только ранговые соотношения этих случайных величин. Абсолютные значения величин ξ_1, \dots, ξ_n используются только при оценивании их одномерных распределений.

Для решения задачи восстановления предпочтения необходимо знать апостериорную вероятность (43). Эта вероятность выражается через частную производную функции копулы C , о чем утверждает следующая теорема.

Пусть ξ_1, ξ_2 — две случайные величины с одномерными функциями распределения $F(x_1), G(x_2)$. Тогда условная вероятность $P(\xi_2 \leq x_2 | \xi_1 = x_1)$ равна частной производной копулы:

$$P(\xi_2 \leq x_2 | \xi_1 = x_1) = \frac{\partial}{\partial v} C(u, v) |_{(G(x_2), F(x_1))},$$

взятой в точке

$$u = G(x_2), \quad v = F(x_1).$$

В нашей задаче имеется n случайных величин ξ_1, \dots, ξ_n , соответствующих признакам, и случайная величина μ , соответствующая вектору меток \mathbf{y} .

Для оценки условной вероятности воспользуемся следующими обозначениями. Обозначим одномерные функции распределения y и всех компонент многомерной случайной величины $(\mu, \boldsymbol{\xi})$:

$$G_\mu^0(y), G_{\xi_1}^1(x_1), \dots, G_{\xi_n}^n(x_n).$$

Обозначим совместные функции распределения упорядоченных поднаборов - векторов $\mathbf{x}_k = [x_1, \dots, x_k]^\top$ размерности от 1 до n :

$$F_{\boldsymbol{\xi}^k}^k(\mathbf{x}^k), \quad \mathbf{x}_k = [x_1, \dots, x_k]^\top, \quad k = 1, \dots, n.$$

Для нахождения условной вероятности $P(\mu \leq y | \mathbf{x}_i)$ воспользуемся частной производной копулы $C(u, v)$ по переменной u :

$$P(\mu \leq y | \mathbf{x}_i) = \frac{\partial}{\partial u} C(u, v) |_{F_{\boldsymbol{\xi}^n}^n(\mathbf{x}_i^n), G_\mu(y)},$$

взятой в точке

$$u = F_{\boldsymbol{\xi}^n}^n(\mathbf{x}_i^n), \quad v = G_\mu(y).$$

Неизвестной в этой формуле является функция совместного распределения $F_{\boldsymbol{\xi}^n}^n$. Чтобы найти эту функцию, воспользуемся теоремой о частной производной копулы:

$$F_{\boldsymbol{\xi}^n}^n(\mathbf{x}^n) = C^{n-1}(u, v) |_{F_{\boldsymbol{\xi}^{n-1}}^{n-1}(\mathbf{x}^{n-1}), G_{\xi_n}^n(x_n)},$$

...

$$F_{\boldsymbol{\xi}^i}^i(\mathbf{x}^i) = C^{i-1}(u, v) |_{F_{\boldsymbol{\xi}^{i-1}}^{i-1}(\mathbf{x}^{i-1}), G_{\xi_i}^i(x_i)},$$

...

$$F_{\xi_1, \xi_2}^2(x^1, x^2) = C^1(u, v) |_{G_{\xi_1}^1(x^1), G_{\xi_2}^2(x^2)}.$$

Таким образом, чтобы оценить апостериорную вероятность $P(\mu \leq y | \mathbf{x}_i)$, необходимо оценить все $n + 1$ одномерные распределения y и компонент случайного вектора \mathbf{x} , а также n копул C, C^1, \dots, C^{n-1} .

Копулы, используемые при построении оценок объектов. Для решения задачи (43) предлагается использовать архимедовскую копулу.

Определение 20. Копула $C(u_1, \dots, u_n)$ называется архимедовской, если для нее выполнены следующие условия:

$$C(u_1, \dots, u_n) = \psi(\psi^{-1}(u_1) + \dots + \psi^{-1}(u_n)),$$

где функция ψ называется генератором, и для нее должны быть выполнены:

$$(-1)^k \psi^{(k)}(x) \geq 0$$

для всех $x \geq 0$ и $k = 0, 1, \dots, n - 2$. А также, функция

$$(-1)^{n-2} \psi^{n-2}(x)$$

должны быть невозрастающей и выпуклой.

Будем использовать частные случаи архимедовской копулы, задаваемые следующими функциями-генераторами:

- копула Клейтона,

$$\psi(t) = (1 + \theta t)^{-\frac{1}{\theta}}, \quad \theta \in \Theta = (0, \infty), \quad (44)$$

- копула Гумбеля,

$$\psi(t) = \exp(-t^{\frac{1}{\theta}}), \quad \theta \in \Theta = [1, \infty). \quad (45)$$

Отметим, что эти семейства копул зависят только от одного параметра θ , что делает задачу менее вычислительно трудоемкой.

В случае копулы Гумбеля, частная производная имеет следующий вид:

$$\frac{\partial}{\partial u} C(u, v) = \left(\frac{\ln u}{\ln C} \right)^{\theta-1} \frac{C}{u}.$$

Оценка параметров копулы. Как было сказано выше, для оценки параметра $\theta \in \Theta$ копулы используются не сами случайные величины ξ, μ , а последовательности рангов этих величин. Векторам \mathbf{x} и \mathbf{y} соответствуют последовательности рангов:

$R_x = (R_{x_1}, \dots, R_{x_m})$, где R_{x_i} – ранг i -го объекта в вариационном ряду вектора \mathbf{x} ,

$R_y = (R_{y_1}, \dots, R_{y_n})$, где R_{y_i} – ранг i -го объекта в вариационном ряду вектора \mathbf{y} .

Отметим, что наиболее часто используемым методом оценки параметров распределения является метод максимизации правдоподобия, который в случае копул записывается следующим образом:

$$L(\theta) = \sum_{i=1}^m \log \left(c_{\theta}(F(x_i), G(y_i)) \right),$$

$$c_{\theta}(u, v) = \frac{\partial^2}{\partial u \partial v} C_{\theta}(u, v).$$

Вместо значений функций одномерных распределений $F(x_i), G(y_i)$ можно подставить их эмпирические значения, получив таким образом функцию псевдоправдоподобия:

$$L'(\theta) = \sum_{i=1}^m \left(\log c_{\theta} \left(\frac{R_i}{m+1}, \frac{S_i}{m+1} \right) \right).$$

Заметим, что функция L' зависит только от самой копулы C_{θ} , то есть, в нашем случае, только от параметра θ , и максимизация этой функции не представляет собой большой вычислительной сложности.

Благодаря этому способу, задача оценки распределений $F_{\xi^i}^i(\mathbf{x}^i)$ распадается на два независимых этапа: оценка параметра θ_i копул C^i путем максимизации псевдоправдоподобия и оценка параметров одномерных распределений $G_{\mu}^0(y), G_{\xi^1}^1(x^1), \dots, G_{\xi^n}^n(x^n)$ с помощью метода максимума правдоподобия.

Алгоритм оценки апостериорного распределения. Приведем подробный алгоритм оценки распределений $F_{\xi^i}^i(\mathbf{x}^i)$. Как было сказано выше, необходимо оценить $n+1$ одномерное распределение $G_{\mu}^0(y), G_{\xi^1}^1(x^1), \dots, G_{\xi^n}^n(x^n)$ и n функций копулы C, C^1, \dots, C^n .

1. Оцениваются одномерные распределения $G_{\xi^1}^1(x^1), G_{\xi^2}^2(x^2)$. Все функции $G_{\xi^i}^i(x^i)$ будем искать в классе бета-распределений. То есть, распределение случайной величины ξ задается плотностью вероятности g_{ξ} , имеющей вид:

$$\begin{cases} g_{\xi}(x) &= \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, \\ B(\alpha, \beta) &= \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx. \end{cases}$$

Параметры α и β для этого распределения оцениваются методом моментов. Для этого численно решается система уравнений:

$$\begin{cases} E(X) &= \frac{\alpha}{\alpha+\beta}, \\ D(X) &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \end{cases}$$

2. Оценим копулу $C^1(u, v)$, связывающую переменные x_1 и x_2 , максимизируя функцию псевдоправдоподобия:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L'_{12}(\theta) = \sum_{i=1}^m \left(\log c_{\theta} \left(\frac{R_{x_1}}{m+1}, \frac{R_{x_2}}{m+1} \right) \right).$$

3. Оценив одномерные распределения $G_{\xi_1}^1(x^1)$, $G_{\xi_2}^2(x^2)$ и копулу $C^1(u, v)$, получаем оценку функции совместного распределения $F_{\xi_1, \xi_2}^2(x^1, x^2)$. Повторяем шаги 1-2, каждый раз прибавляя по одному новому признаку x^i и оценивая на шаге 3 функцию $F_{\xi^i}^i(\mathbf{x}^i)$.

4. Повторив n раз шаги 1-2, получим функцию совместного распределения всех признаков $F_{\xi^n}^n$. На последнем шаге оценим функцию распределения $G_{\mu}^0(y)$, копулу $C(u, v)$, связывающую μ и ξ , и найдем \hat{y} , доставляющий максимум апостериорной вероятности:

$$\hat{y} = \arg \max_y P(\mu \leq y | \mathbf{x}_i) = \arg \max_y \frac{\partial}{\partial u} C(u, v) \Big|_{F_{\xi^n}^n(\mathbf{x}_i^n), G_{\mu}^0(y)},$$

взятой в точке

$$u = F_{\xi^n}^n(\mathbf{x}_i^n), \quad v = G_{\mu}^0(y),$$

где

$$F_{\xi^i}^i(\mathbf{x}_i^i) = C^{i-1}(u, v) \Big|_{F_{\xi^{i-1}}^{i-1}(\mathbf{x}^{i-1}), G_{\xi_i}^i(x_i)},$$

взятой в точке

$$u = F_{\xi^{i-1}}^{i-1}(\mathbf{x}^{i-1}), \quad v = G_{\xi_i}^i(x_i)$$

для всех

$$i = 2, \dots, n.$$

Выбор признаков. Для повышения качества классификации на контрольной подвыборке, предлагается выбрать наиболее информативные признаки. Множество индексов признаков, включенных в функцию вероятности (43), назовем активным набором и обозначим $\mathcal{A} \subseteq \mathcal{J}$.

Для того, чтобы выбрать наиболее информативные признаки, предлагается использовать следующий эвристический алгоритм. Информационными будем считать те признаки, которые имеют наибольшую ранговую связь с вектором меток \mathbf{y} . Чтобы понять, какие признаки имеют наибольшую связь, рассмотрим некоторые свойства копул о ранговой связи.

Лемма 10. *Случайные величины ξ и μ являются независимыми тогда и только тогда, когда*

$$C(u, v) = uv, \quad u, v \in [0, 1],$$

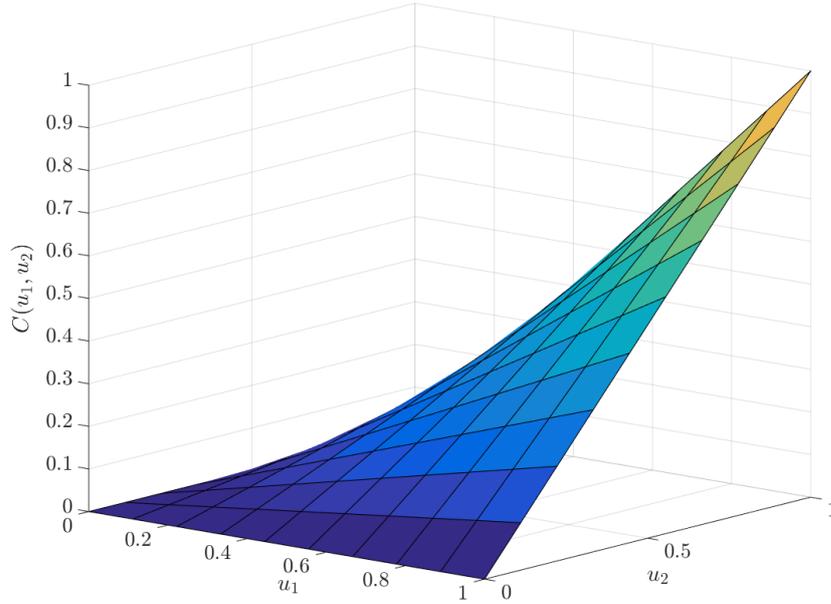


Рис. 4. Копула Клейтона

где

$$C(F(x), G(y)) = H(x, y),$$

где $H(x, y)$ — совместная функция распределения случайных величин ξ и μ .

Лемма 11. Границы Фреше для копулы:

$$W(u, v) \leq C(u, v) \leq M(u, v), \quad u, v \in [0, 1],$$

где

$$W(u, v) = \max(0, u + v - 1) -$$

минимальная копула,

$$M(u, v) = \min(u, v) -$$

максимальная копула.

Причем, если $C(u, v) = W(u, v)$, то μ — монотонно убывающая функция ξ , если $C(u, v) = M(u, v)$, то μ — монотонно возрастающая функция ξ .

В качестве примера рассмотрим копулу Гумбеля (45):

$$C_\theta(u, v) = \exp \left[\left((-\log(u))^\theta + (-\log(v))^\theta \right)^{\frac{1}{\theta}} \right] \quad \theta \geq 1.$$

При стремлении параметра копулы $\theta \rightarrow 1$, $C_\theta(u, v) \rightarrow uv$, то есть, случайные величины являются независимыми. При стремлении параметра $\theta \rightarrow \infty$, ранговая связь между случайными

величинами возрастает. Таким образом, ранговая связь изменяется монотонно при варьировании параметра копулы. Для решения задачи отбора признаков будем отбирать те из них, для которых параметр копулы с вектором меток \mathbf{u} является наибольшим.

Исходя из этого рассуждения, предлагается следующий алгоритм.

1. Примем пустое множество активных признаков

$$\mathcal{A} = \emptyset.$$

2. Для всех $j = 1, \dots, n$ вычислим параметры θ_j для копул $C_{\theta_j}(F_i(x^j), G(y))$ и включим в набор

$$\mathcal{A} = \mathcal{A} \cup \{k\}$$

тот признак k , для которого

$$k = \arg \max_{j \in \mathcal{J}} \theta_j.$$

Обозначим множество оставшихся признаков

$$\mathcal{J}' = \mathcal{J} \setminus \mathcal{A}.$$

3. Для всех признаков $j \in \mathcal{A}$ и всех $k_1, \dots, k_{\mathcal{A}}$ вычислим параметры θ_j для копул

$$C_{\theta_j}(F_i(x_i), H_{k_1, \dots, k_{\mathcal{A}}}(x^{k_1}, \dots, x^{k_{\mathcal{A}}}))$$

и включим в набор

$$\mathcal{A} = \mathcal{A} \cup \{k\}$$

тот признак k , для которого

$$k = \arg \max_{j \in \mathcal{J}'} \theta_j.$$

4. Будем повторять шаг 3, пока значение ошибки на контрольной выборке не стабилизируется.

3 Построение модели предпочтений с использованием экспертных оценок

В этой главе рассмотрим задачу построения интегральной оценки качества объектов или, другими словами, задачу ранжирования объектов с использованием экспертной информации. Будут рассмотрены случаи линейной и порядковой зависимой переменной, а также линейные и порядковые экспертные ограничения элементы вектора параметров модели. Основные результаты этой главы опубликованы в работе [58].

Для принятия решений при администрировании объектов управления, например, государственных заповедников, регионов государств или объектов финансирования, возникает необходимость упорядочить объекты в соответствии с интегральной оценкой их качества, иначе, построить на наборе объектов бинарную функцию предпочтения. Для построения интегральной оценки качества объектов выполняются следующие шаги.

1. Выбирается критерий качества $S(f, X)$.
2. Формируется набор объектов $X = \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$, сравнимых в контексте выбранного критерия.
3. Формируется набор из тех признаков, которые эксперт считает необходимыми для описания этого критерия. Считается, что значения признаков выставлены в линейной шкале $\mathbf{x} \in \mathbb{R}^n$.
4. На основании выбранных критериев и признаков составляется матрица описаний объектов \mathbf{X} .
5. Собираются вектор значений целевой переменной \mathbf{y} и вектор экспертных оценок весов признаков \mathbf{w} .

Ранее было предложено несколько подходов к построению интегральных оценок объектов. Обзор этих подходов представлен в разделе 3.1. Различают методы построения оценок без учителя и с учителем. Методы с учителем характеризуются наличием заданной на наборе объектов целевой переменной.

Подход без учителя заключается в нахождении оценок с помощью описаний объектов и выбранного метода их построения. Одним из таких методов является метод главных компонент [49], согласно которому интегральная оценка является проекцией векторов-описаний

объектов на первую главную компоненту матрицы плана. Также в качестве методов без учителя рассматриваются Парето-расслоение и метрический метод [69].

Подход с учителем использует кроме описаний объектов значения целевой переменной, которые могут быть заданы линейным или порядковым описанием. В линейном случае ранее был предложен подход, в котором восстанавливается регрессия описаний объектов на экспертные оценки качества объектов [80]. Случай порядковой целевой переменной соответствует задаче обучения ранжированию [61], основная сложность которой заключается в определении функции потерь.

Задача согласования экспертных оценок, рассматриваемая в разделе 3.2, характеризуется наличием экспертных ограничений на вектор весов признаков, линейных или порядковых. В общем случае, оптимальное решение задачи не удовлетворяет экспертным ограничениям, то есть экспертные данные противоречат друг другу. Предлагается метод устранения разногласия в данных экспертов. Рассматриваются задачи согласования линейных и порядковых экспертных оценок.

В первом случае рассматриваются линейные экспертные оценки и значения целевой переменной. Предлагается следующая процедура их уточнения. Множество допустимых экспертных оценок представляется отрезком, ограниченным наибольшим и наименьшим значением оценки. Вводится структурный параметр, позволяющий выбрать точку на этом отрезке и отдать предпочтение оценкам объектов или оценкам весов признаков.

Во втором случае экспертные оценки весов признаков и значения целевой переменной являются порядковыми. Порядковые экспертные оценки весов признаков задают конус предпочтений, свойства которого приведены в подразделе 1.2.3. Матрица плана задает линейное отображение этого конуса в пространство оценок объектов. Полученный в результате отображения конус может пересекаться с конусом значений целевой переменной. В этом случае экспертные оценки показателей и объектов считаются непротиворечивыми, и отыскивается наиболее устойчивый вектор оценок. В противном случае, выполняется процедура уточнения оценок. Уточнение оценок выполняется с помощью минимизации расстояния между векторами, располагающимися в двух конусах одного пространства.

Предлагаемые алгоритмы в этом разделе будут проиллюстрированы модельными данными, показанными в таблице 1 (подробное описание этих данных приведено в разделе 4.3 и в работе [69]). Данные представляют собой 10 объектов-заповедников, описанных восемью признаками, значения которых принадлежат отрезку $[0; 1]$. Помимо этого, заданы экспертные оценки весов признаков (выделены курсивом) и значения целевой переменной (выделены

Таблица 1. Задача ранжирования заповедников, используемые данные

Объект	Оценки качества объектов	Признак 1	Признак 2	Признак 3	Признак 4	Признак 5	Признак 6	Признак 7	Признак 8
<i>Оценки весов</i>		<i>1.00</i>	<i>0.85</i>	<i>0.78</i>	<i>0.70</i>	<i>0.69</i>	<i>0.58</i>	<i>0.48</i>	<i>0.29</i>
\mathbf{x}_1	1.00	3	0	1	4.5	3	3	2.5	0
\mathbf{x}_2	0.83	1	7	1	8	2	8.5	5	40
\mathbf{x}_3	0.67	2	1	1.5	9	1.5	9.5	6.5	66
\mathbf{x}_4	0.63	1	3	3	5	4.5	18	11	7
\mathbf{x}_5	0.58	0	12	8	19	11	7.5	11	62
\mathbf{x}_6	0.50	0	0	2	5	2.5	2.5	3.5	11
\mathbf{x}_7	0.44	1	5	4	11	20	16.5	15.5	4
\mathbf{x}_8	0.38	0	0	3	7	1	4.5	2.5	0
\mathbf{x}_9	0.33	0	6	0	7	1.5	7	4.5	1
\mathbf{x}_{10}	0.17	0	0	1	4	2	2.5	3.5	14.5

жирным). Задача заключается в восстановлении интегральной оценки \mathbf{u} с учетом экспертных данных.

3.1 Постановка задачи и базовые методы

Задано множество X , состоящее из m объектов, $X = \mathbf{x}_1, \dots, \mathbf{x}_m$. Каждый объект \mathbf{x}_i является элементом n -мерного евклидова пространства, $\mathbf{x}_i \in \mathbb{R}^n$. Поскольку объекты заданы линейным описанием, определена матрица плана \mathbf{X} размера $m \times n$, $\mathbf{X}(i, j) = x_{i,j}$, где элемент матрицы $x_{i,j}$ соответствует значению j -го признака i -го объекта.

Задача построения интегральной оценки, или задача ранжирования, заключается в поиске отображения $f : X \rightarrow \mathbb{R}$, минимизирующего функцию потерь S .

В данной главе рассматривается линейная модель, согласно которой значение оценки $f(\mathbf{x}_i)$ является линейной комбинацией признаков-элементов вектора \mathbf{x}_i :

$$f(\mathbf{w}, \mathbf{x}_i) = \mathbf{w}^\top \mathbf{x}_i. \quad (46)$$

За вектор \mathbf{f} оценок качества объектов обозначим вектор значений функции f на всем наборе объектов,

$$\mathbf{f} = [f(\mathbf{w}, \mathbf{x}_1), \dots, f(\mathbf{w}, \mathbf{x}_m)]^\top = \mathbf{X}\mathbf{w}, \quad (47)$$

где \mathbf{w} — вектор весов признаков.

Будем считать, что значения каждого признака приведены в единую шкалу по правилу

$$x_{ij} \mapsto \frac{x_{ij} - \min_i x_{ij}}{\max_i x_{ij} - \min_i x_{ij}},$$

и все элементы x_{ij} принадлежат отрезку $[0, 1]$.

Кроме того, будем рассматривать случай матрицы \mathbf{X} полного ранга, $\text{rank}(\mathbf{X}) = n$, при $m \geq n$.

3.1.1 Ранжирование объектов без учителя

Данная задача отличается от предыдущих постановок тем, что на множестве X не задана целевая переменная, или целевое отношение предпочтения z_0 .

Метод главных компонент. Для получения интегральных оценок качества объектов без учителя рассмотрим метод главных компонент. Данным методом строятся такие проекции объектов на новую систему координат, называемую главными компонентами, чтобы сумма квадратов расстояний от объектов до их проекций на первую главную компоненту была бы минимальной. Рассмотрим ортогональную матрицу \mathbf{W} в линейной комбинации $\mathbf{U}^T = \mathbf{X}^T \mathbf{W}$ таких векторов-строк матрицы \mathbf{X} , что векторы-столбцы $\mathbf{u}_1, \dots, \mathbf{u}_n$ матрицы \mathbf{U} будут иметь максимальную сумму дисперсий,

$$\sum_{j=1}^n \sigma^2(\mathbf{u}_j) \rightarrow \max,$$

где

$$\sigma^2(\mathbf{u}_j) = \frac{1}{m} \sum_{i=1}^m (u_i - \bar{u})^2, \quad \bar{u} = \frac{1}{m} \sum_{i=1}^m u_i.$$

Векторы-столбцы матрицы \mathbf{W} есть собственные векторы ковариационной матрицы $\Sigma = \mathbf{X}^T \mathbf{X}$. Последняя может быть найдена с помощью сингулярного разложения матрицы $\mathbf{X}^T \mathbf{X}$. Поскольку $\Sigma = \mathbf{X}^T \mathbf{X} = \mathbf{W} \Lambda \mathbf{W}^T$ (в сингулярном разложении матрица \mathbf{W} — ортогональная, а Λ — диагональная с невозрастающими неотрицательными элементами), то $\Sigma \mathbf{W} = \mathbf{W} \Lambda$ является собственной системой матрицы $\Sigma \mathbf{W}$.

Вектор оценок объектов $\mathbf{f} = \mathbf{X} \mathbf{w}$ есть проекция векторов-строк матрицы \mathbf{X} на первую главную компоненту, \mathbf{w} — первая вектор-строка матрицы \mathbf{W} . Этот вектор соответствует максимальному собственному значению матрицы Σ .

На рис. 5 показана иллюстрация метода главных компонент для вычисления оценок объектов методом без учителя. Для иллюстрации взяты все 10 объектов и первые два призна-

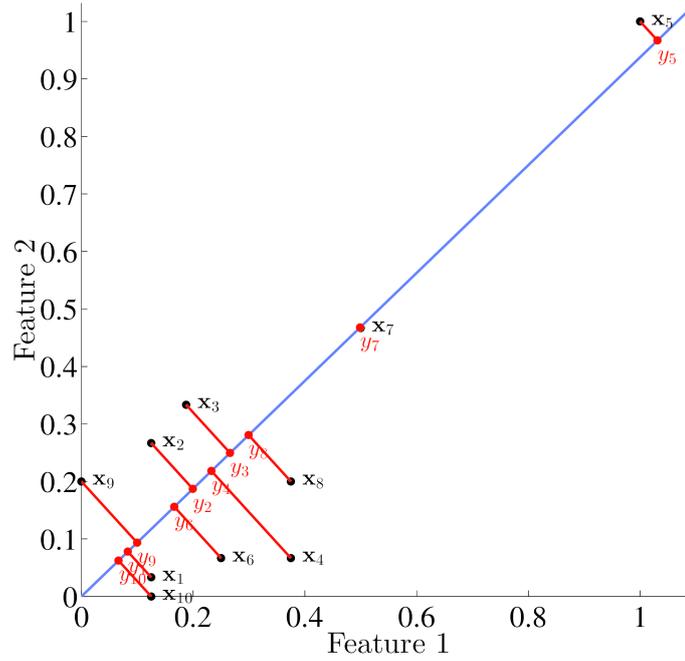


Рис. 5. Иллюстрация метода главных компонент

ка из табл. 1. Оценки $y_i = f(\mathbf{x}_i)$ объектов \mathbf{x}_i являются проекциями точек, соответствующих объектам, на первую главную компоненту, обозначенную синей линией.

Взвешенная сумма. Дополнительно к исходной постановке задачи экспертным путем задаются оценки весов признаков $\mathbf{w}_0 \in \mathbb{R}^n$. Вектор оценок объектов $\hat{\mathbf{y}}$ вычисляется как линейная комбинация столбцов матрицы \mathbf{X} с весами \mathbf{w}_0 :

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{w}_0.$$

Основной недостаток этого метода заключается в том, что при большой обусловленности матрицы \mathbf{X} небольшое возмущение вектора экспертных оценок вызовет значительное изменение вектора оценок объектов.

На рис. 6 показано сравнение экспертной оценки $\mathbf{y}_1 = \mathbf{X}\mathbf{w}_0$, полученной методом взвешенной суммы, с оценкой $\hat{\mathbf{y}}_{\text{PCA}}$, полученной методом главных компонент. Каждая точка на этом рисунке является объектом-заповедником. По оси абсцисс отложено значение оценки $\hat{\mathbf{y}}_{\text{PCA}}$, по оси ординат — значение оценки \mathbf{y}_1 .

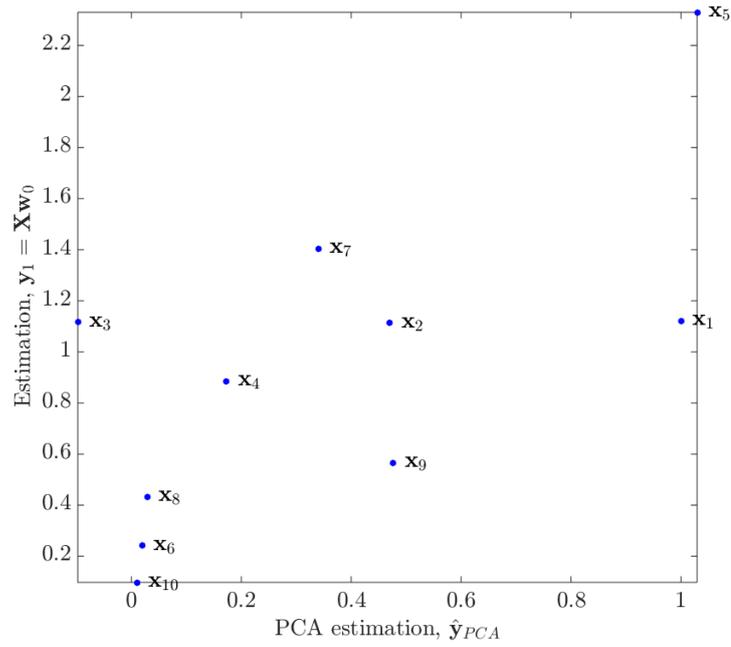


Рис. 6. Сравнение метода взвешенной суммы с оценкой \hat{y}_{PCA} , полученной методом главных компонент

3.1.2 Ранжирование объектов с учителем

Методы ранжирования объектов с учителем отличаются тем, что на множестве объектов задана целевая переменная, соответствующая экспертным оценкам объектов. В этом подразделе будем рассматривать два типа целевой переменной.

- **Линейная целевая переменная:** на множестве объектов задан вектор оценок объектов $\mathbf{y}_0 \in \mathbb{R}^n$.
- **Порядковая целевая переменная:** на множестве X определено целевое отношение предпочтения z_0 , задаваемое конечным набором меток Y , $z_0(\mathbf{x}_i, \mathbf{x}_k) = [y_i \geq_0 y_k]$. За вектор \mathbf{y}_0 , будем обозначать вектор меток, заданный на множестве Y .

Задача ранжирования заключается в минимизации функции ошибки $S(f, X, \mathbf{y}_0)$.

Линейная целевая переменная: экспертно-статистический метод. Пусть целевая переменная задана в линейном виде, $\mathbf{y}_0 \in \mathbb{R}^m$. Отыскиваются такие веса $\hat{\mathbf{w}}$, которые доставляют минимум сумме квадратов невязок $\mathbf{X}\mathbf{w} - \mathbf{y}_0$:

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w} \in \mathbb{R}^n} \|\mathbf{X}\mathbf{w} - \mathbf{y}_0\|^2.$$

Решение этой задачи для невырожденной матрицы $\mathbf{X}^\top \mathbf{X}$ получается методом наименьших квадратов,

$$\hat{\mathbf{w}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}_0. \quad (48)$$

Вектор оценок объектов вычисляется как $\hat{\mathbf{y}} = \mathbf{X} \hat{\mathbf{w}}$. Вычисленный таким образом вектор оценок объектов является ближайшим к \mathbf{y}_0 из пространства столбцов матрицы \mathbf{X} .

Порядковая целевая переменная: обучение ранжированию. Случай, когда на множестве объектов определено целевое отношения предпочтения z_0 , задаваемое конечным набором меток Y и вектором \mathbf{y}_0 , относится к категории задач обучения ранжированию. Задача состоит в минимизации функции потерь $S(f, X, \mathbf{y}_0)$, определяемой специальным образом. Как было сказано в разделе 1.1, исторически рассматривают три класса подходов к решению задачи, каждый из которых характеризуется особым видом функции потерь: поточечный, попарный и списочный подходы.

Рассмотрим попарный подход к задаче ранжирования. В рамках этого подхода задача сводится к задаче двухклассовой классификации в пространстве пар объектов. Стандартная функция потерь штрафует пары объектов, не согласованных по экспертно заданному предпочтению z_0 . Одним из самых используемых алгоритмов для данного подхода служит ранговый метод опорных векторов RankSVM (4). Минимизируемая функция потерь для метода RankSVM выглядит следующим образом:

$$S(f, X, z_0) = \sum_{i,k=1}^m (1 - [y_{0i} \succeq y_{0k}]) \mathbf{w}^\top (\mathbf{x}_i - \mathbf{x}_k)_+ + \lambda \|\mathbf{w}\|^2.$$

Такой вид функции потерь отличается от стандартной [85] видом функции отступа. Гиперплоскость «наибольшего зазора» отыскивается в пространстве пар объектов. При этом штрафуются те пары объектов, для которых не совпадают экспертное отношение доминирования и отношение доминирования, порожденного выражением $\mathbf{w}^\top (\mathbf{x}_i - \mathbf{x}_k)$.

3.2 Согласование экспертных оценок в задаче восстановления предпочтений

В данном разделе рассмотрим задачу согласования экспертных оценок. Рассматривается задача построения интегральных оценок (ранжирования) объектов с учителем с дополнительными экспертными ограничениями на множество параметров модели. Необходимо построить вектор оценок объектов, наилучшим образом приближающий целевую переменную и учитывающий экспертные ограничения.

Будем рассматривать две постановки задачи согласования. Задача согласования *линейных* экспертных оценок, рассматриваемая в подразделе 3.2.1, использует линейную целевую переменную $\mathbf{y}_0 \in \mathbb{R}^m$ и линейную экспертную оценку весов признаков $\mathbf{w}_0 \in \mathbb{R}^n$. Результаты подраздела 3.2.1 частично повторяют и развивают результаты, полученные В. В. Стрижовым в работе [69]. Задача согласования *порядковых* экспертных оценок (подраздел 3.2.2) использует порядковую целевую переменную $\mathbf{y}_0 \in Y^m$ и порядковую экспертную оценку важности признаков $w_{j_1} \succeq w_{j_2}$.

3.2.1 Согласование линейных экспертных оценок

Задано множество объектов $X = \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$, матрица плана \mathbf{X} , целевая переменная $\mathbf{y}_0 \in \mathbb{R}^m$ и экспертная оценка весов признаков $\mathbf{w}_0 \in \mathbb{R}^n$ для линейной модели $f(\mathbf{w}, \mathbf{x}) = \mathbf{w}^\top \mathbf{x}$.

Рассмотрим оценки объектов $\mathbf{y}_1 = \mathbf{X}\mathbf{w}_0$, вычисленные по экспертным оценкам весов признаков, а также веса признаков $\mathbf{w}_1 = \mathbf{X}^+\mathbf{y}_0$, определенные по экспертным оценкам объектов \mathbf{y}_0 . Веса признаков вычисляются с помощью псевдообратного линейного отображения $\mathbf{X}^+ = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$. Другими словами, отображение \mathbf{X} переводит вектор экспертных оценок показателей \mathbf{w}_0 в вектор \mathbf{y}_1 , а псевдообратное отображение \mathbf{X}^+ — вектор экспертных оценок объектов \mathbf{y}_0 в вектор \mathbf{w}_1 . В общем случае, вычисленные и выставленные оценки объектов, а также оценки весов признаков, различаются: $\mathbf{y}_1 \neq \mathbf{y}_0$, $\mathbf{w}_1 \neq \mathbf{w}_0$.

Определение 21. *Экспертные оценки \mathbf{y} и \mathbf{w} называются согласованными, если для них выполняются следующие условия:*

$$\mathbf{y} = \mathbf{X}\mathbf{w}, \quad \mathbf{w} = \mathbf{X}^+\mathbf{y}. \quad (49)$$

Будем искать экспертные оценки, для которых выполняются условия согласованности (49). Обозначим за $\mathbf{y}'_0 = \mathbf{X}\mathbf{X}^+\mathbf{y}_0$ вектор, являющийся проекцией вектора экспертных оценок \mathbf{y}_0 на пространство столбцов матрицы \mathbf{X} .

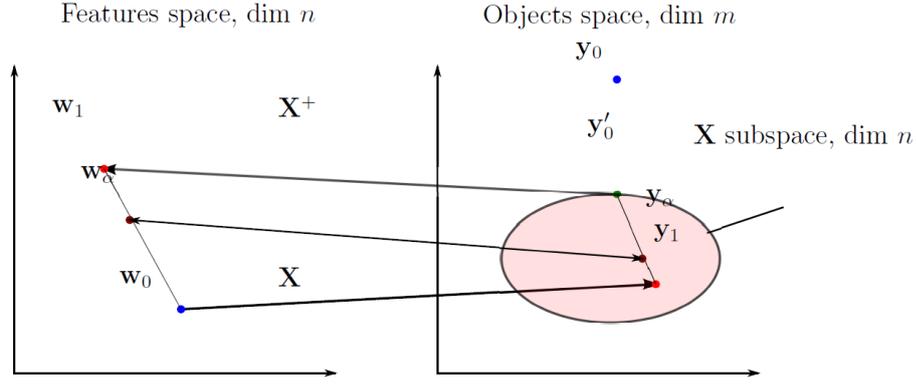
Метод α -согласования экспертных оценок. Для разрешения противоречия экспертных данных используются компромиссные оценки:

$$\mathbf{w}_\alpha \in [\mathbf{w}_0, \mathbf{w}_1] \text{ и } \mathbf{y}_\alpha \in [\mathbf{y}_1, \mathbf{y}'_0]. \quad (50)$$

Пара векторов $\mathbf{w}_\alpha, \mathbf{y}_\alpha$ для заданного α определяется выражениями

$$\mathbf{w}_\alpha = \alpha \mathbf{w}_0 + (1 - \alpha) \mathbf{X}^+ \mathbf{y}'_0, \quad \mathbf{y}_\alpha = (1 - \alpha) \mathbf{y}'_0 + \alpha \mathbf{X} \mathbf{w}_0.$$

Выполнена следующая теорема.

Рис. 7. Иллюстрация метода α -согласования

Теорема 3. Для заданных таким образом векторов $\mathbf{w}_\alpha, \mathbf{y}_\alpha$ выполняются условия согласованности (49).

Доказательство. В самом деле,

$$\mathbf{X}\mathbf{w}_\alpha = \alpha\mathbf{X}\mathbf{w}_0 + (1 - \alpha)\mathbf{X}\mathbf{X}^+\mathbf{y}'_0 = \alpha\mathbf{X}\mathbf{w}_0 + (1 - \alpha)\mathbf{y}'_0 = \mathbf{y}_\alpha,$$

$$\mathbf{X}^+\mathbf{y}_\alpha = (1 - \alpha)\mathbf{X}^+\mathbf{y}'_0 + \alpha\mathbf{X}^+\mathbf{X}\mathbf{w}_0 = (1 - \alpha)\mathbf{X}^+\mathbf{y}'_0 + \alpha\mathbf{w}_0 = \mathbf{w}_\alpha.$$

□

На рис. 7 показана иллюстрация метода согласования экспертных оценок в линейных шкалах. \mathbf{y}_0 и \mathbf{w}_0 — экспертные оценки объектов и весов признаков. \mathbf{y}'_0 — ближайшая к \mathbf{y}_0 точка в подпространстве столбцов матрицы \mathbf{X} размерности n . Пара векторов \mathbf{y}_α и \mathbf{w}_α является согласованной парой экспертных оценок.

Параметр α определяет предпочтение экспертных оценок. При $\alpha \rightarrow 0$ отдается предпочтение оценкам объектов, при $\alpha \rightarrow 1$ — оценкам показателей. Параметр α может быть задан самим экспертом или найден как решение задачи минимизации сумм квадратов остатков,

$$Q = \frac{\|\mathbf{w}_0 - \mathbf{w}_\alpha\|}{n} + \frac{\|\mathbf{y}'_0 - \mathbf{y}_\alpha\|}{m} \rightarrow \min. \quad (51)$$

Метод γ -согласования экспертных оценок. При отказе от ограничений (50) и отыскании компромиссных оценок в окрестностях векторов $\mathbf{w}_0, \mathbf{y}'_0$ решается задача оптимизации (51) с регуляризирующим параметром $\gamma^2 \in [0, +\infty)$. Перепишем эту задачу в виде

$$\mathbf{w}_\gamma = \arg \min_{\mathbf{w} \in \mathbb{R}^n} (\varepsilon^2 + \gamma^2 \delta^2),$$

где $\varepsilon^2 = \|\mathbf{w}_0 - \mathbf{w}_\gamma\|^2$ and $\delta^2 = \|\mathbf{y}'_0 - \mathbf{y}_\gamma\|^2$. Решением этой задачи являются уточненные экспертные оценки весов признаков

$$\mathbf{w}_\gamma = (\mathbf{X}^T \mathbf{X} + \gamma^2 I_n)^{-1} (\mathbf{X}^T \mathbf{y}'_0 + \gamma^2 \mathbf{w}_0) \quad (52)$$

и уточненные оценки весов показателей $\mathbf{y}_\gamma = \mathbf{X}\mathbf{w}_\gamma$. Здесь параметр γ^2 так же, как и параметр α , определяет доверие экспертным оценкам объектов либо весов признаков.

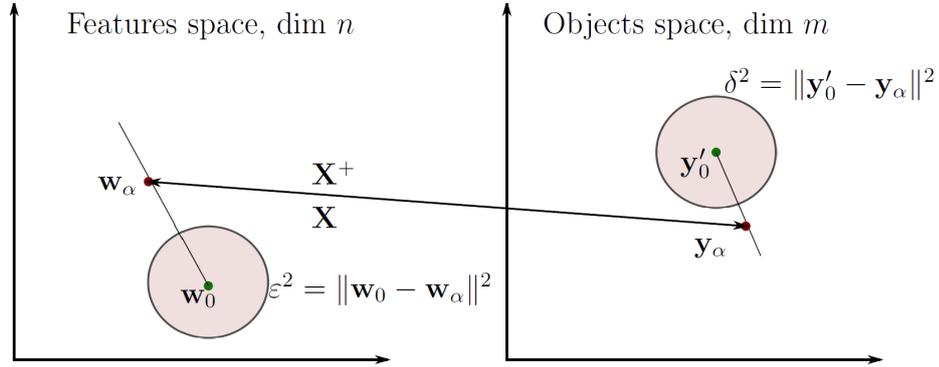


Рис. 8. Иллюстрация метода γ -согласования

На рис. 8 показана иллюстрация метода γ -согласования. Радиусы кругов, описанных около точек \mathbf{w}_0 и \mathbf{y}'_0 равны, соответственно, ε и δ .

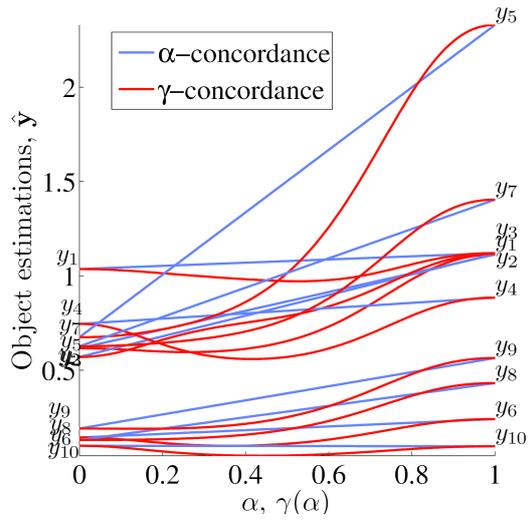
На рис. 9 показано сравнение методов α - и γ -согласования. По оси абсцисс отложены значения параметра α от 0 до 1, при этом значение параметра γ изменяется по закону $\gamma = \frac{\alpha}{1-\alpha}$, то есть γ изменяется от 0 до ∞ . На графике слева показано изменение оценок объектов. Синими линиями показано изменение оценок для α -согласования, красными линиями показано изменение оценок для γ -согласования. Крайние случаи соответствуют значениям оценок \mathbf{y}'_0 и $\mathbf{y}_1 = \mathbf{X}\mathbf{w}_0$, соответственно. На графике справа показано изменение значения функции ошибки (51). Видно, что в случае γ -согласования эта функция имеет явно выраженный минимум, соответствующий оптимальному значению параметра γ .

На рис. 10 показано сравнение экспертно-статистического метода и метода γ -согласования для случая $\gamma = 1$.

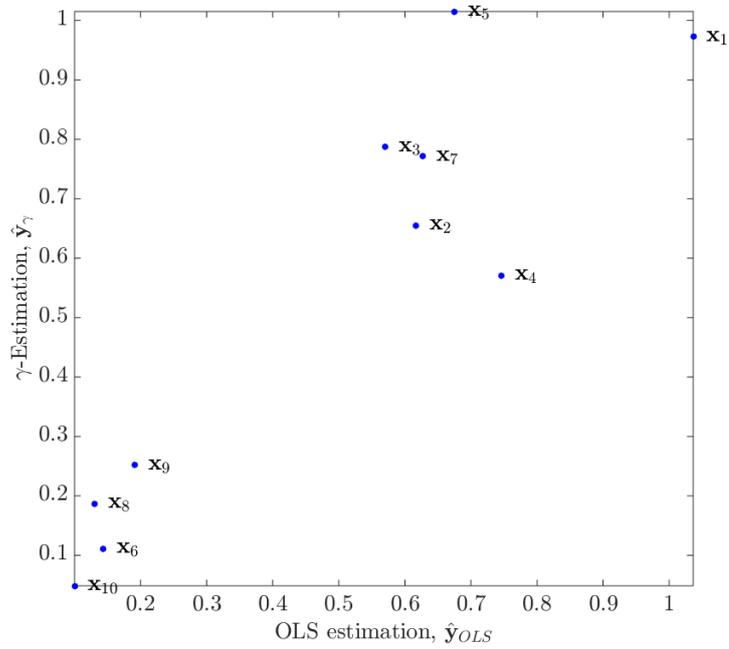
3.2.2 Согласование порядковых экспертных оценок

В этом разделе рассматривается случай порядковых экспертных оценок $\mathbf{y}_0, \mathbf{w}_0$. Задано множество объектов $X = \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ и матрица плана \mathbf{X} . На множестве X определено целевое отношение предпочтения z_0 , задаваемое вектором $\mathbf{y}_0 \in Y^m$: $z_0(\mathbf{x}_i, \mathbf{x}_k) = [y_{0i} \geq y_{0k}]$. Кроме того, задано отношение предпочтения на множестве признаков, выражающее степени важности критериев. Это отношение предпочтения задано в виде систем линейных неравенств на вектор весов признаков \mathbf{w} линейной модели $f(\mathbf{w}, \mathbf{x}) = \mathbf{w}^\top \mathbf{x}$: $w_j \geq w_t$.

Другими словами, определено два конуса частичного порядка, выражающие 1) целевое предпочтение \mathbf{y}_0 и 2) важность критериев. Обозначим эти конусы, соответственно, \mathcal{Y}_0 и \mathcal{W}_0 .



(a) Изменение оценок объектов

Рис. 9. Иллюстрация α - и γ -согласованияРис. 10. Сравнение экспертно-статистического метода и метода γ -согласования

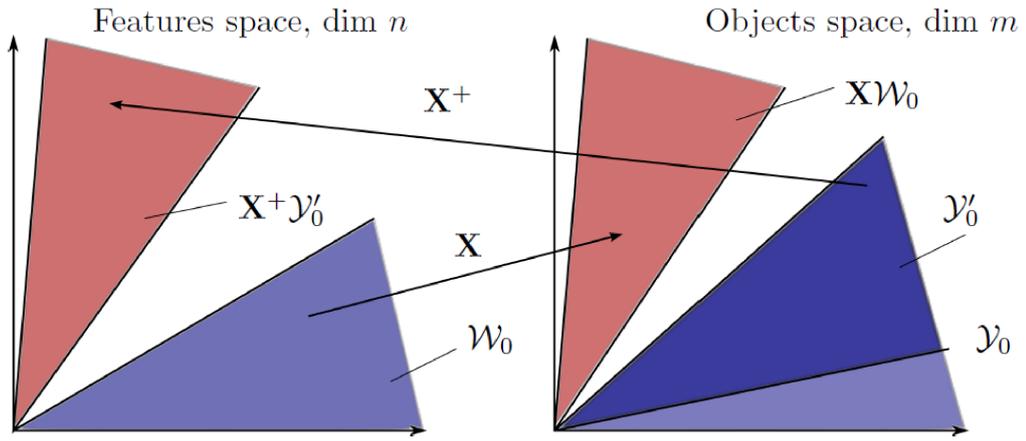


Рис. 11. Иллюстрация конусов, соответствующих экспертным оценкам

Будем рассматривать полиэдральное представление этих конусов с матрицами A^m и A^n , соответственно:

$$Y'_0 = \{x | A^m x \leq 0\}, \quad W_0 = \{w | A^n w \leq 0\}.$$

Таким образом, рассматривается семейство линейных моделей, $f(x) = w^T x$, где вектор w принадлежит конусу экспертных оценок, $w \in W_0$. Линейное отображение, задаваемое матрицей X , переводит конус W_0 экспертных оценок весов признаков в конус XW_0 . В свою очередь, линейное отображение, задаваемое матрицей XX^+ , переводит конус Y'_0 экспертных оценок объектов в конус $Y'_0 = XX^+Y'_0$, состоящий из векторов, лежащих в подпространстве, образованном столбцами матрицы X (X^+ является псевдообратным отображением). Перечисленные конусы проиллюстрированы на рис. 11.

Порядковые экспертные оценки будем называть *согласованными*, если конусы XW_0 и Y'_0 имеют непустое пересечение, отличное от точки 0 . В этом случае, найдутся векторы $\hat{y} \in Y'_0$, $\hat{w} \in W_0$, удовлетворяющие условиям согласованности (49). Если же конусы XW_0 и Y'_0 в качестве пересечения содержат только точку 0 , будем называть порядковые экспертные оценки *несогласованными*.

Для обоснования предложенных ниже методов приведем некоторые свойства полиэдральных конусов.

Лемма 12. *Пересечение полиэдральных конусов с вершиной в начале координат является полиэдральным конусом.*

Доказательство. Рассмотрим два полиэдральных конуса. Выпуклый полиэдральный конус с вершиной в начале координат — это область решения некоторой системы однородных неравенств. Пусть первому конусу соответствует система неравенств $A_1 x \leq 0$, а второму —

$\mathbf{A}_2\mathbf{x} \leq \mathbf{0}$. Пересечение двух конусов — область решений системы, составленной из неравенств обеих систем, соответствующих конусам. Другими словами, пересечение двух данных конусов задается системой однородных неравенств с матрицей

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{pmatrix}.$$

□

Лемма 13. *Геометрическое место точек, в которое отображение $\mathbf{X} : \mathcal{W} \rightarrow \mathcal{Y}$ переводит конус, является конусом.*

Доказательство. Для любого вектора \mathbf{w} , принадлежащего конусу \mathcal{W} , вектор $\lambda\mathbf{w}$ также ему принадлежит, а $\mathbf{y} = \mathbf{X}\mathbf{w}$. Поэтому, если вектор \mathbf{y} принадлежит рассматриваемому геометрическому месту точек, то и вектор $\lambda\mathbf{y} = \lambda\mathbf{X}\mathbf{w} = \mathbf{X}(\lambda\mathbf{w})$ ему принадлежит.

Таким образом, если \mathcal{W} — полиэдральный конус, то отображение \mathbf{X} переводит его в полиэдральный конус $\mathbf{X}\mathcal{W}$. Соответствующее псевдообратное отображение \mathbf{X}^+ переводит конус \mathcal{Y} в конус $\mathbf{X}^+\mathcal{Y}$. □

Лемма 14. $\mathcal{W}_0 \cap \mathbf{X}^+\mathcal{Y}'_0 = \{\mathbf{0}\} \iff \mathbf{X}\mathcal{W}_0 \cap \mathcal{Y}'_0 = \{\mathbf{0}\}$.

Доказательство. Отметим сначала, что конусы \mathcal{W}_0 и $\mathcal{Y}'_0 = \mathbf{X}\mathbf{X}^+\mathcal{Y}_0$ имеют одинаковую размерность n , поскольку $\text{rank}(\mathbf{X}) = n$. Это означает, что отображения \mathbf{X}, \mathbf{X}^+ осуществляют взаимно однозначное преобразование из пространства признаков размерности n в пространство столбцов матрицы \mathbf{X} .

Пусть вектор $\mathbf{w} \in \mathcal{W}_0 \cap \mathbf{X}^+\mathcal{Y}'_0$, $\mathbf{w} \neq \mathbf{0}$. Тогда соответствующий вектор $\mathbf{X}\mathbf{w} \in \mathbf{X}\mathcal{W}_0$, а также $\mathbf{X}\mathbf{w} \in \mathbf{X}\mathbf{X}^+\mathcal{Y}'_0 = \mathcal{Y}'_0$. То есть, конусы $\mathbf{X}\mathcal{W}_0$ и \mathcal{Y}'_0 пересекаются по ненулевому вектору $\mathbf{X}\mathbf{w}$.

В другую сторону доказательство аналогично: пусть вектор $\mathbf{y} \in \mathcal{Y}'_0 \cap \mathbf{X}\mathcal{W}_0$, $\mathbf{y} \neq \mathbf{0}$. Тогда соответствующий вектор $\mathbf{X}^+\mathbf{y} \in \mathbf{X}^+\mathcal{Y}'_0$, а также $\mathbf{X}^+\mathbf{y} \in \mathbf{X}^+\mathbf{X}\mathcal{W}_0 = \mathcal{W}_0$. То есть, конусы \mathcal{W}_0 и $\mathbf{X}^+\mathcal{Y}'_0$ пересекаются по ненулевому вектору $\mathbf{X}^+\mathbf{y}$. □

Из доказанного выше утверждения следует, что для каждого вектора \mathbf{w}_p , принадлежащего конусу $\mathcal{W}_p = \mathcal{W}_0 \cap \mathbf{X}^+\mathcal{Y}'_0$, найдется вектор $\mathbf{y}'_p \in \mathcal{Y}'_0 = \mathbf{X}\mathcal{W}_0 \cap \mathcal{Y}'_0$, такой, что для векторов $\mathbf{w}_p, \mathbf{y}'_p$ выполняются условия согласованности (49). При этом, для вектора $\mathbf{y}'_p \in \mathcal{Y}'_0$ найдется вектор $\mathbf{y}_p \in \mathcal{Y}_0$, такой, что $\mathbf{y}'_p = \mathbf{X}\mathbf{X}^+\mathbf{y}_p$, что следует из определения конуса \mathcal{Y}'_0 .

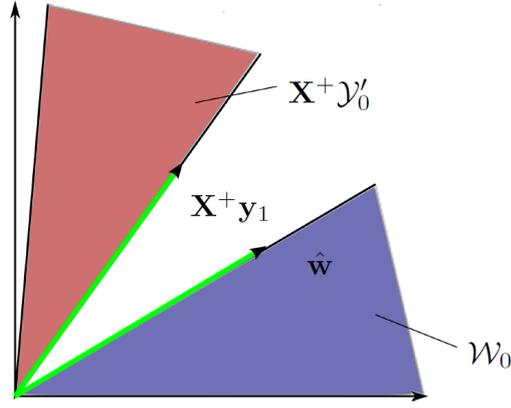


Рис. 12. Нахождение ближайших векторов в конусах

Исходя из вышесказанного, согласованная пара $\hat{\mathbf{w}}, \hat{\mathbf{y}}$ должна удовлетворять следующим ограничениям:

$$\begin{cases} \mathbf{A}^n \hat{\mathbf{w}} \leq \mathbf{0}, \\ \hat{\mathbf{y}} = \mathbf{X}\mathbf{X}^+ \mathbf{y}, \\ \mathbf{A}^m \mathbf{y} \leq \mathbf{0}. \end{cases}$$

Постановка задачи согласования экспертных оценок с использованием полиэдральных конусов. Поставим задачу согласования порядковых экспертных оценок. Будем искать уточненные экспертные оценки как ближайшие векторы в двух конусах.

Требуется найти в конусах \mathcal{W}_0 и \mathcal{Y}_0 векторы $\hat{\mathbf{w}}$ и \mathbf{y}_1 , такие что

$$(\hat{\mathbf{w}}, \mathbf{y}_1) = \arg \min_{\mathbf{w} \in \mathcal{W}_0, \mathbf{y} \in \mathcal{Y}_0} \|\mathbf{X}^+ \mathbf{y} - \mathbf{w}\|_2, \quad \|\mathbf{X}^+ \mathbf{y}\|_2 = 1, \quad \|\mathbf{w}\|_2 = 1, \quad (53)$$

где $\|\cdot\|_2$ — евклидова метрика в пространстве \mathbb{R}^m .

Таким образом, отыскивается вектор весов $\hat{\mathbf{w}}$, элементы которого имеют ранговый порядок, согласованный с конусом \mathcal{W}_0 . При этом оценка объектов $\hat{\mathbf{y}} = \mathbf{X}\mathbf{X}^+ \mathbf{y}_1$, приведенная в порядковую шкалу, принадлежит пространству столбцов матрицы \mathbf{X} и является ближайшей к \mathcal{Y}_0 . Иллюстрация задачи нахождения ближайших векторов в конусах показана на рис. 12.

Алгоритм решения задачи минимизации расстояния между векторами в конусах.

Перепишем задачу (53) в следующем виде:

$$\begin{aligned} & \text{minimize} && \|\mathbf{X}^+ \mathbf{y} - \mathbf{w}\|_2 \\ & \text{subject to} && (\mathbf{X}^+ \mathbf{y})^\top \mathbf{X}^+ \mathbf{y} = 1, \quad \mathbf{w}^\top \mathbf{w} = 1, \\ & && \mathbf{A}^n \mathbf{w} \leq \mathbf{0} \quad \mathbf{A}^m \mathbf{y} \leq \mathbf{0}. \end{aligned}$$

Построим итерационный алгоритм, последовательно находящий приближения векторов $\mathbf{y}^{(2k)}$, $\mathbf{w}^{(2k+1)}$ на четном и нечетном шаге. Векторы $\mathbf{a} = \mathbf{y}^{(2k)}$ и $\mathbf{b} = \mathbf{w}^{(2k+1)}$ будем считать решениями двух последовательно решаемых оптимизационных задач, полагая вектор $\mathbf{w}^{(0)} = \mathbf{w}_0$ на шаге $k = 0$, где \mathbf{w}_0 — центральная точка конуса \mathcal{W}_0 .

Задача $2k$:	Задача $2k + 1$:
minimize $\ \mathbf{X}^+\mathbf{a} - \mathbf{w}^{(2k)}\ _2$	minimize $\ \mathbf{X}^+\mathbf{y}^{(2k+1)} - \mathbf{b}\ _2$
subject to $(\mathbf{X}^+\mathbf{a})^\top \mathbf{X}^+\mathbf{a} = 1,$	subject to $\mathbf{b}^\top \mathbf{b} = 1,$
$\mathbf{A}^m \mathbf{a} \leq 0.$	$\mathbf{A}^n \mathbf{b} \leq 0.$

При решении задач на каждом шаге значения констант $\mathbf{w}^{(2k)}$ и $\mathbf{y}^{(2k+1)}$ принимаются равными значениям соответствующих решений \mathbf{a} и \mathbf{b} предыдущего шага.

Минимизация расстояния между векторами в конусах, задаваемых порождающим представлением. Задачу (53) можно также переформулировать с использованием конусов, задаваемых представлением через набор порождающих элементов. Конусы \mathcal{W}_0 и \mathcal{Y}_0 могут быть представлены в виде

$$\mathcal{W}_0 = \{\mathbf{Z}_w \boldsymbol{\lambda}_w \mid \boldsymbol{\lambda}_w \geq \mathbf{0}\}, \quad \mathcal{Y}_0 = \{\mathbf{Z}_0 \boldsymbol{\lambda}_0 \mid \boldsymbol{\lambda}_0 \geq \mathbf{0}\},$$

где \mathbf{Z}_w — матрица предпочтений, определенного на элементах вектора весов показателей \mathbf{w} . В этом случае задача нахождения оптимальных параметров (53) переписется следующим образом:

$$(\hat{\boldsymbol{\lambda}}_0, \hat{\boldsymbol{\lambda}}_w) = \arg \min_{\boldsymbol{\lambda} \geq \mathbf{0}, \|\boldsymbol{\lambda}\|_1=1} \|\mathbf{Z}_0 \boldsymbol{\lambda}_0 - \mathbf{X} \mathbf{Z}_w \boldsymbol{\lambda}_w\|_2 = \arg \min_{\boldsymbol{\lambda} \geq \mathbf{0}, \|\boldsymbol{\lambda}\|_1=1} \|\mathbf{Z}_0 \boldsymbol{\lambda}_0 - \mathbf{X}_w \boldsymbol{\lambda}_w\|_2, \quad (54)$$

где матрица \mathbf{X}_w имеет следующую структуру. Пусть, например, исходная матрица \mathbf{X} состоит из четырех признаков, $\mathbf{X} = [\boldsymbol{\chi}_1, \boldsymbol{\chi}_2, \boldsymbol{\chi}_3, \boldsymbol{\chi}_4]$, при этом, согласно экспертным предпочтениям, на элементах вектора \mathbf{w} задан линейный порядок: $w_1 \geq w_2 \geq w_3 \geq w_4 \geq 0$. Тогда матрица предпочтений является верхней треугольной,

$$\mathbf{Z}_w = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

и матрица \mathbf{X}_w представляется частичной суммой столбцов матрицы \mathbf{X} :

$$\mathbf{X}_w = [\boldsymbol{\chi}_1, \boldsymbol{\chi}_1 + \boldsymbol{\chi}_2, \boldsymbol{\chi}_1 + \boldsymbol{\chi}_2 + \boldsymbol{\chi}_3, \boldsymbol{\chi}_1 + \boldsymbol{\chi}_2 + \boldsymbol{\chi}_3 + \boldsymbol{\chi}_4].$$

Для решения задачи (54) предлагается следующий итеративный алгоритм, основывающийся на последовательном решении задач неотрицательной линейной регрессии:

$$\begin{array}{c|c} \text{Задача } 2k : & \text{Задача } 2k + 1 : \\ \hline \hat{\boldsymbol{\lambda}}_0 = \min_{\boldsymbol{\lambda}_0 \geq \mathbf{0}, \|\boldsymbol{\lambda}_0\|=1} \|\mathbf{Z}_0 \boldsymbol{\lambda}_0 - \mathbf{X}_w \hat{\boldsymbol{\lambda}}_w^{(2k-1)}\|_2^2 & \hat{\boldsymbol{\lambda}}_w = \min_{\boldsymbol{\lambda}_w \geq \mathbf{0}, \|\boldsymbol{\lambda}_w\|=1} \|\mathbf{Z}_0 \hat{\boldsymbol{\lambda}}_0^{(2k)} - \mathbf{X}_w \boldsymbol{\lambda}_w\|_2^2 \end{array}$$

Минимизация расстояния с использованием монотонной интерполяции. Рассмотрим частный случай задачи (53), при котором вектор оценок объектов \mathbf{y}_0 считается представленным в линейной шкале, а оценки весов признаков являются порядковыми. В этом случае, задача (53) может быть переформулирована в терминах задачи монотонной интерполяции [81, 45].

Обозначим $\tilde{\mathbf{w}} = \mathbf{X}^+ \mathbf{y}_0$. Требуется найти такую монотонную последовательность $w_1 \leq \dots \leq w_n$, что она лучше всего приближает вектор $\tilde{\mathbf{w}}$ в смысле среднего квадрата ошибки:

$$\begin{cases} \hat{\mathbf{w}} = \arg \min_{\mathbf{w} \in \mathbb{R}^n} \sum_{i=1}^n (\tilde{w}_i - w_i)^2, \\ w_n \geq \dots \geq w_1. \end{cases}$$

Чтобы добиться согласования экспертных оценок, введем в модель параметр ν . При $\nu \rightarrow 0$ предпочтение отдается экспертным оценкам объектов, при $\nu \rightarrow \infty$ — экспертным оценкам весов признаков.

Требуется найти такой вектор $\hat{\mathbf{w}}$, что:

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w} \in \mathbb{R}^n} \left(\frac{1}{2} \sum_{i=1}^n (\tilde{w}_i - w_i)^2 + \nu \sum_{i=1}^{n-1} (w_i - w_{i+1})_+ \right).$$

Алгоритм решения такой задачи представлен в [81].

Согласование оценок с использованием структурного параметра. В случае нетривиального пересечения конусов \mathcal{W}_0 и $\mathbf{X}^+ \mathcal{Y}_0$, решением задачи (53) является вектор $\hat{\mathbf{w}}$, лежащий в пересечении этих конусов, и вектор $\hat{\mathbf{y}}$, для которого выполняется равенство $\hat{\mathbf{y}} = \mathbf{X} \mathbf{X}^+ \mathbf{y}_1 = \mathbf{X} \hat{\mathbf{w}}$, то есть, для векторов $\hat{\mathbf{y}}, \hat{\mathbf{w}}$ выполнены условия согласованности (49). Если пересечение конусов — тривиальное, алгоритм найдем ближайшие друг к другу векторы, для которых не выполнено условие согласования. Аналогично случаю линейных шкал, предлагается процедура согласования экспертных оценок со структурным параметром α .

Получив решения задачи, векторы $\hat{\mathbf{w}}$ и $\hat{\mathbf{y}} = \mathbf{X} \mathbf{X}^+ \mathbf{y}_1 \neq \mathbf{X} \hat{\mathbf{w}}$, выполним процедуру линейного уточнения вектора оценок объектов,

$$\mathbf{y}_\alpha = (1 - \alpha) \hat{\mathbf{y}} + \alpha \mathbf{X} \hat{\mathbf{w}}.$$

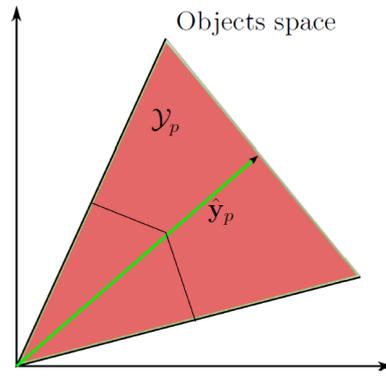


Рис. 13. Иллюстрация вектора \mathbf{y}_p как наиболее устойчивого решения

Вектор \mathbf{y}_α и соответствующий ему вектор $\mathbf{w}_\alpha = \mathbf{X}^+ \mathbf{y}_\alpha$ задают в соответствующих пространствах конусы \mathcal{Y}_α и \mathcal{W}_α , причем пересечение $\mathcal{Y}_\alpha \cap \mathbf{X}\mathcal{W}_\alpha$ не пусто. Так же, как в случаях уточнения оценок в линейных шкалах, при $\alpha \rightarrow 0$ предпочтение отдается экспертным оценкам качества объектов, при $\alpha \rightarrow 1$ — экспертным оценкам весов признаков. Ниже покажем, как по уточненным оценкам, выставленным в порядковых шкалах, можно получить оценки в линейных шкалах.

Оценки, устойчивые к возмущению матрицы описаний. Рассмотрим найденный конус $\mathcal{Y}_p = \mathcal{Y}_\alpha \cap \mathcal{W}_\alpha$ и матрицу плана \mathbf{X} . Возмутим элементы этой матрицы, представив $\mathbf{X} \rightarrow \mathbf{X} + \Delta\mathbf{X}$, приняв гипотезу нормального распределения матрицы $\Delta = \delta\mathbf{I}$, $\delta \sim \mathcal{N}(0, \sigma^2)$. Образ линейного отображения $\mathbf{y} = (\mathbf{X} + \Delta\mathbf{X})\mathbf{w}$ будет также иметь нормальное распределение. Согласно принятой гипотезе, будем считать устойчивым к малому возмущению матрицы \mathbf{X} такой вектор \mathbf{y}_p , который наиболее удален от всех граней конуса \mathcal{Y}_p при условии нормировки $\|\mathbf{y}_p\| = 1$. Вектор \mathbf{y}_p является центром сферы, вписанной в конус \mathcal{Y}_p , и называется точкой Чебышева.

Иллюстрация вектора \mathbf{y}_p показана на рис. 13. Расстояние от искомого вектора \mathbf{y}_p до граней \mathbf{b} конуса отыскивается как решение оптимизационной задачи

$$\hat{\mathbf{y}}_p = \arg \max_{\mathbf{y}_p \in \mathcal{Y}_p} \{ \|\mathbf{y}_p - \mathbf{b}\|_2 : \mathbf{b} \in \mathbb{R}^m \setminus \mathcal{Y}_p, \|\mathbf{y}_p\|_2 \leq 1 \}. \quad (55)$$

На рис. 14 показано сравнение метода с порядковыми экспертными данными, заключающегося в отыскании устойчивого решения внутри конуса (55), и метода с порядковыми признаками (38).

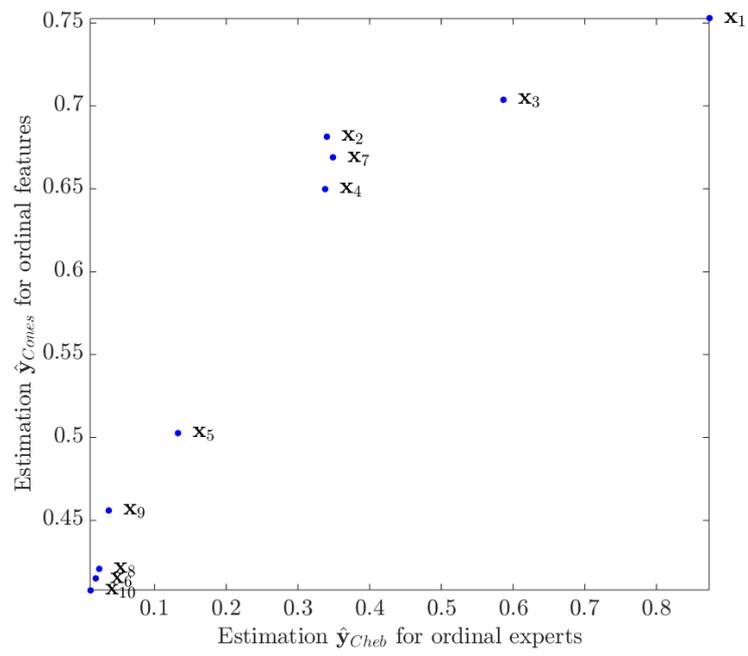


Рис. 14. Сравнение оценок, полученных методом с порядковыми экспертными данными и методом с порядковыми признаками

4 Анализ прикладных задач

Эта глава посвящена анализу и иллюстрации предложенных методов на прикладных задачах. Будут рассмотрены задача категоризации Красной книги РФ [79], задача ранжирования заповедников [58] и задача порядковой классификации на наборах данных из репозитория UCI [6].

4.1 Категоризация объектов Красной книги РФ

Редкие и находящиеся под угрозой исчезновения объекты животного мира играют важную роль в экосистемах, во многих случаях являясь надежными индикаторами их состояния. Сохранение генофонда редких видов — одна из главных задач охраны окружающей среды. Красная книга Российской Федерации — важнейший правовой механизм сохранения таких видов. Она необходима для организации исследований и мониторинга состояния этих животных и растений и их местообитаний, для разработки и осуществления специальных мер по их охране, восстановлению и научно обоснованному использованию. Занесение того или иного вида дикого животного или растения в Красную книгу является юридически значимым действием, формализующим признаком, выделяющим редкий вид, как объект правовой охраны, от других видов.

Процесс ведения Красных книг в России имеет более чем 30-летнюю историю. Первая Красная книга СССР была издана в 1978 г. В соответствии с Положением о Красной книге СССР, занесение в неё какого-либо вида означало установление запрета на его добывание, возлагало на уполномоченные государственные органы обязательства по охране как самого вида, так и его местообитаний. Виды были отнесены лишь к двум категориям:

1. Категория А — виды, находящиеся под угрозой исчезновения.
2. Категория Б — редкие виды.

В Красной книге Российской Федерации приняты следующие категории статуса редкости таксонов (видов, подвидов, популяций) животных:

0. Вероятно исчезнувшие. Таксоны, известные ранее с территории (или акватории) Российской Федерации и нахождение которых в природе не подтверждено (для беспозвоночных — в последние 100 лет, для позвоночных животных — в последние 50 лет).

1. Находящиеся под угрозой исчезновения. Таксоны, численность особей которых уменьшилась до критического уровня таким образом, что в ближайшее время они могут исчезнуть.
2. Сокращающиеся в численности. Таксоны с неуклонно сокращающейся численностью, которые при дальнейшем воздействии факторов, снижающих численность, могут в короткие сроки попасть в категорию находящихся под угрозой исчезновения.
3. Редкие. Таксоны, которые имеют малую численность и распространены на ограниченной территории (или акватории) или спорадически распространены на значительных территориях (или акваториях).
4. Неопределенные по статусу. Таксоны, которые, вероятно, относятся к одной из предыдущих категорий, но достаточных сведений об их состоянии в природе в настоящее время нет, либо они не в полной мере соответствуют критериям всех остальных категорий.
5. Восстанавливаемые и восстанавливающиеся. Таксоны, численность и распространение которых под воздействием естественных причин или в результате принятых мер охраны начали восстанавливаться и приближаются к состоянию, когда не будут нуждаться в срочных мерах по сохранению и восстановлению.

Действующий Перечень объектов животного мира, занесенных в Красную книгу Российской Федерации (с изменениями на 28 апреля 2011 г.), включает 414 видов животных: 155 видов беспозвоночных и 259 видов позвоночных: 40 видов круглоротых и рыб, 8 видов земноводных, 21 вид пресмыкающихся, 123 вида птиц, 65 видов млекопитающих.

Вместе с тем некоторые виды животных занесены в Красную книгу Российской Федерации на уровне подвида или даже популяции, при этом для разных таксонов (подвидов, популяций) одного вида могут быть установлены разные категории статуса редкости. Поэтому для заполнения таблицы в части объектов животного мира в качестве базовой единицы брался не вид, а таксон (вид, подвид, популяция). Таким образом, действующий Перечень объектов животного мира, занесенных в Красную книгу Российской Федерации, включает 437 таксонов животных: 155 таксонов беспозвоночных и 282 таксона позвоночных: 48 таксонов круглоротых и рыб, 8 таксонов земноводных, 21 таксон пресмыкающихся, 128 таксонов птиц, 77 таксонов млекопитающих.

Биологическое состояние	Суммарные угрозы	Значимость	Защитенность	Готовность
Численность	Климатические изменения	Возможные потери	Конвенции, договора, законодательства	Искусственное воспроизведение популяций
Тенденция изменения численности	Геологические катастрофы	Для ареала в России	применение, достаточность, эффективность	Репродукция в природе
Темп изменения численности	Биогеографические факторы	Биогеографическая роль	Стратегии, менеджмент-планы	Сохранение EX-SITU
Тенденция изменения температуры численности	Разрушение местообитаний	Ресурсная значимость	применение, достаточность, эффективность	Необходимые затраты на обеспечение сохранения/восстановления
Встречаемость/ Плотность	Добыча (сбор)	Научная значимость	ООПТ	Степень изученности
Тенденция изменения плотности	Загрязнения	Индикаторная значимость	применение, достаточность, эффективность	Надежность мониторинга
Общая площадь ареала	Интродукция чужеродных видов		применение, достаточность, эффективность	
Тенденция изменения площади ареала	Случайная гибель		Территории международного значения, ключевые территории	
Структура ареала	Бесплодность		применение, достаточность, эффективность	
Тенденция изменения структуры ареала	Эксплуатация кормовой базы		Восстановление и сохранение местообитаний	
Популяционная структура вида			применение, достаточность, эффективность	
Тенденция изменения популяционной структуры			Искусственное воспроизводство популяций	
Генетическое разнообразие			применение, достаточность, эффективность	
Тенденция изменения генетического разнообразия			Репродукция (реакклиматизация)	
Половозрастная и социальная структура			применение, достаточность, эффективность	
Тенденция изменения половозрастной структуры			Создание новых популяций	
Физиологическое состояние			применение, достаточность, эффективность	
Тенденция изменения физиологического состояния			Регуляция использования и торговли	
Состояние местообитаний			применение, достаточность, эффективность	
Тенденция изменения состояния местообитания			Управление воспроизводством	
			применение, достаточность, эффективность	
			Борьба с болезнями и паразитами	
			применение, достаточность, эффективность	
			Регуляция численности	
			применение, достаточность, эффективность	
			Сохранение генетических материалов	
			применение, достаточность, эффективность	
			Содержание и разведение в неволе	
			применение, достаточность, эффективность	
			Введение в культуру	
			применение, достаточность, эффективность	
			Экологическое образование	
			применение, достаточность, эффективность	
			Эколого-просветительская работа	
			применение, достаточность, эффективность	
			Saracu Building/Training	
			применение, достаточность, эффективность	
			Экологическая пропаганда	
			применение, достаточность, эффективность	
			Эколого-художественная деятельность	
			применение, достаточность, эффективность	

Рис. 15. Признаки в задаче категоризации таксонов

Таблица 2. Экспертное признаковое описание объектов Красной книги РФ

Названия таксонов	Азовская белуга	Сибирский осетр	Сахалинский осетр	Волховский сиг	Длинноперая палля	Кильдинская треска	Белый медведь	Желтозобик	Кулик-лопачень
Численность	1	2	2	2	3	2	3	2	2
Плотность	1	2	1	2	3	2	3	2	2
Площадь ареала	3	2	2	2	1	1	3	1	1
Структура ареала	5	5	4	5	5	5	5	2	2
Структура популяции	1	2	0	2	0	1	1	1	1
Генетическое разнообразие	1	0	0	1	0	1	0	0	0
Физиологическое состояние	2	3	3	3	3	3	3	3	2
Изменение размера популяции	1	2	2	3	3	0	4	3	2
Экспертная категоризация	1	2	1	2	3	1	5	3	3

Описание выборки. Каждый таксон описан набором признаков, отражающих его состояние. Список используемых для построения модели признаков представлен на рис. 15. Предполагается, что текущая версия Красной книги РФ составлена «в целом» непротиворечиво, то есть, существует соответствие между описанием вида и его статусом. Эксперт, владеющий информацией о таксоне, выставляет оценку для каждого признака в ранговой шкале. Таким образом, задана матрица «объект-признак» (табл. 2), состоящая из описаний таксонов, и вектор меток классов таксонов. Требуется построить отображение, восстанавливающую класс таксона из Красной книги РФ по его описанию. Это отображение должно принимать значения на линейно упорядоченном множестве статусов таксона.

Задача ревизии Красной книги РФ и построения модели оценок таксонов является актуальной из-за постоянного пополнения книги новыми записями о таксонах.

Используются следующие предположения о составе и свойствах признаков:

- 1) оцениваемый экспертами состав признаков считается исчерпывающим для получения модели;
- 2) учитывается структура на признаках;

Таблица 3. Доминирование признаков: биологические

	Численность	Изменение численности	Плотность	Площадь ареала	Структура ареала	Структура популяции	Генетическое разнообразие	Социальная структура	Физиологическое состояние	Состояние местообитаний
Численность	1	1	1	1	1	1	1	1	1	1
Изменение численности	0	1	1	1	1	1	1	1	1	1
Плотность	0	0	1	1	1	1	1	1	1	1
Площадь ареала	0	0	0	1	1	1	1	1	1	1
Структура ареала	0	0	0	0	1	1	1	1	1	0
Структура популяции	0	0	0	0	0	1	1	1	1	0
Генетическое разнообразие	0	0	0	0	0	0	1	0	0	0
Социальная структура	0	0	0	0	0	0	1	1	1	0
Физиологическое состояние	0	0	0	0	0	0	1	0	1	0
Состояние местообитаний	0	0	0	0	1	1	1	1	1	1

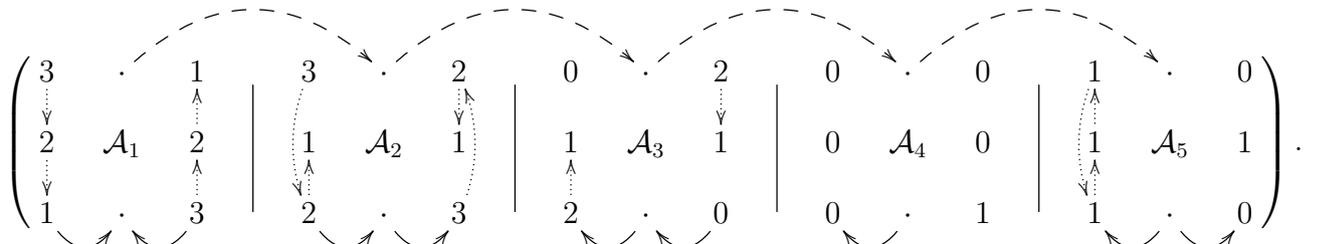
- 3) на значениях признаков задано отношение полного порядка, кроме пропущенного значения;
- 4) выполняется правило «the bigger the better», то есть большему (благоприятному) значению признака соответствует больший (благоприятный) статус вида;
- 5) различные экспертные оценки по одному и тому же виду приветствуются;
- 6) каждый из признаков принимает на выборке все допустимые значения и только их, при нарушении условия ставится вопрос о корректности анкеты;
- 7) признак отвергается, если более половины значений пропущены.

Как показано на рис. 15, используемое множество признаков разделено на пять подмножеств: биологическое состояние \mathcal{A}_1 , воздействия и угрозы \mathcal{A}_2 , значимость \mathcal{A}_3 , защищенность \mathcal{A}_4 и готовность \mathcal{A}_5 . Внутри каждого подмножества на множестве признаков экспертами задано отношение доминирования. Для построения корректной модели необходимо учитывать экспертную информацию о доминировании признаков. В некоторых случаях (табл. 3) это отношение является отношением линейного порядка, в других (табл. 4) — отношением частичного порядка. Более того, для подмножества «воздействие и угрозы» (табл. 4) эти экспертные оценки являются противоречивыми, то есть не задают отношение порядка ввиду невыполнения условия транзитивности.

Таблица 4. Доминирование признаков: воздействие и угрозы

	Климатические изменения	Геологические катастрофы	Биоценологические факторы	Разрушение местообитаний	Добыча (сбор)	Загрязнения	Интродукция чужеродных видов	Случайная гибель	Беспокойство	Эксплуатация кормовой базы
Климатические изменения	1	1	0	0	0	0	1	1	1	0
Геологические катастрофы	0	1	0	0	0	0	0	1	0	0
Биоценологические факторы	1	1	1	0	0	0	1	1	1	0
Разрушение местообитаний	1	1	1	1	0	1	1	1	1	0
Добыча (сбор)	1	1	1	0	1	1	1	1	1	0
Загрязнения	1	1	1	0	0	1	1	1	1	0
Интродукция чужеродных видов	0	0	0	0	0	0	1	1	1	0
Случайная гибель	0	0	0	0	0	0	0	1	0	0
Беспокойство	0	1	0	0	0	0	1	1	1	0
Эксплуатация кормовой базы	1	1	1	0	0	1	1	1	1	1

Помимо отношения доминирования внутри каждой из групп $\mathcal{A}_1, \dots, \mathcal{A}_5$, экспертами задано отношение предпочтения между группами. Это отношение является линейным, $\mathcal{A}_1 \succ \dots \succ \mathcal{A}_5$. Схематично экспертная анкета показана на следующей схеме:



Штриховыми пунктирными стрелками показано отношение доминирования между подмножествами $\mathcal{A}_1, \dots, \mathcal{A}_5$. Сплошными стрелками показано отношение доминирования внутри каждого из подмножеств $\mathcal{A}_1, \dots, \mathcal{A}_5$. Точечным пунктиром показано отношение доминирования по каждому признаку.

Для решения этой задачи были использованы следующие подходы: подходы на основе оценивания матрицы предпочтений, подход на основе конусов, метод криволинейной регрессии и подход на основе копул (все подходы описаны в главе 2), а также метод решающих деревьев (Trees), порядковый метод опорных векторов (SVM) [38] и метод классификации на основе ближайших соседей с монотонными ограничениями (KNN), описанный в [27].

Эксперимент проводился на выборке, состоящей из экспертных оценок видов Красной книги РФ и экспертных оценок важности признаков. В выборку вошли 102 объекта из трех

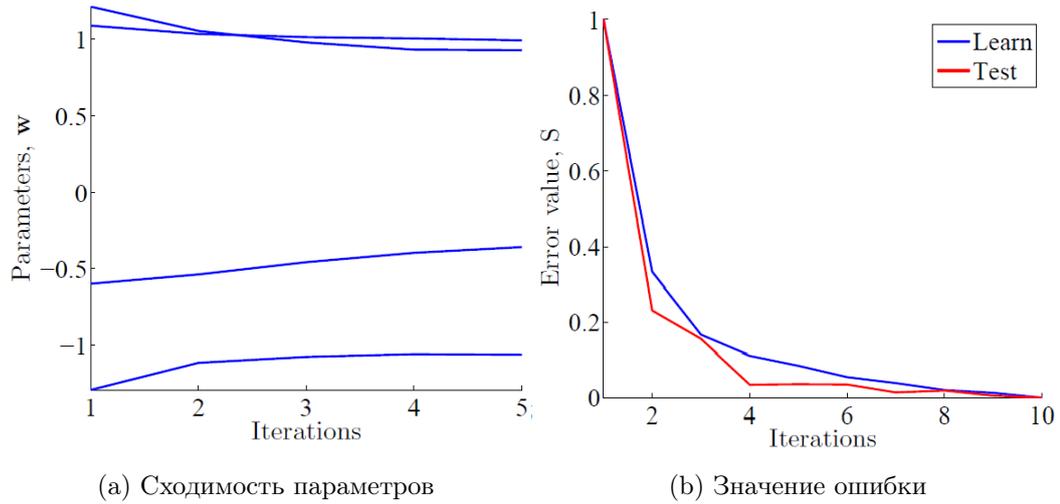


Рис. 16. Сходимость параметров полиэдральной модели, случай малого количества параметров

категорий риска, описанные 102-мя признаками (равенство числа объектов и признаков является случайным совпадением). Признаки представлены в таблице на рис. 15.

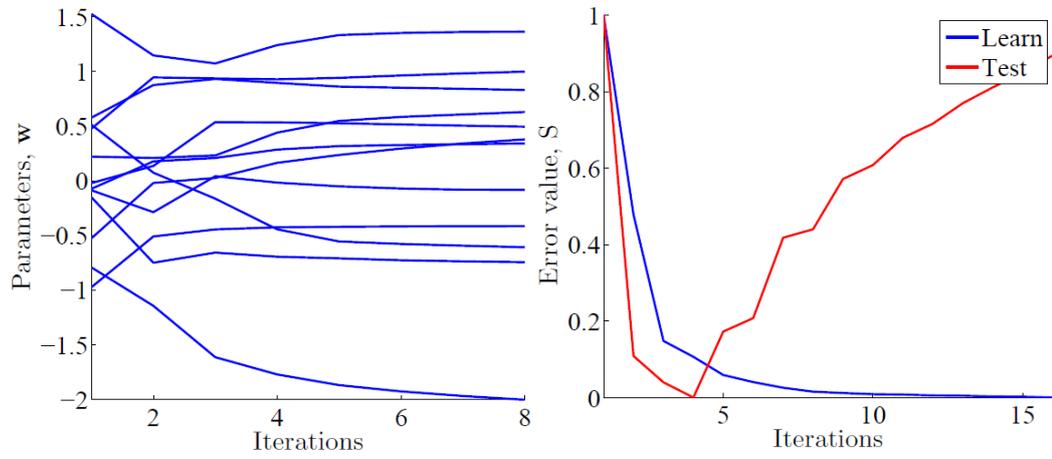
Для оценки качества алгоритмов в этом подразделе рассматривается средняя потеря Хэмминга $L_H(\mathbf{y}, \hat{\mathbf{y}})$, задаваемая формулой

$$L_H(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i|_H, \quad (56)$$

где \mathbf{y} – заданная целевая переменная, а $\hat{\mathbf{y}}$ – его оценка. Отметим, что множество значений вектора \mathbf{y} является упорядоченным множеством меток $Y = \{y_1 \succ \dots \succ y_k\}$, поэтому в качестве нормы $\|\cdot\|_H$ рассматривается функция Хэмминга [79]. Ниже представлены графики сходимости методов и результаты их сравнения.

Алгоритм прямой оценки параметров полиэдральной модели. Для иллюстрации работы алгоритма прямой оценки параметров, рассматриваемого в подразделе 2.3.3, было рассмотрено три подмножества признаков различной мощности. Так, на рис. 16 показан случай малого количества признаков (4), на рис. 17 – среднего (12), на рис. 18 – случай большого количества (все признаки). Синей линией показано значение функции ошибки на обучающей выборке, красной – на контрольной подвыборке, выбранной случайным образом.

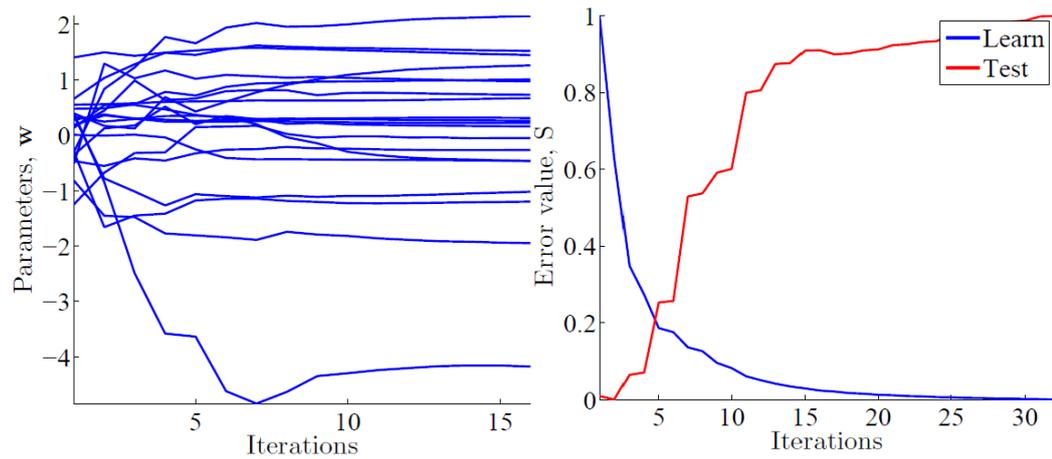
Видно, что в случае малого количества параметров (рис. 16b) значение ошибки на обучающей и контрольной выборке монотонно уменьшается. В случае среднего количества параметров (рис. 17b) значение ошибки на контроле сначала уменьшается, затем начинает расти. На рис. 18b показана сходимость функции ошибки в случае большого количества парамет-



(a) Сходимость параметров

(b) Значение ошибки

Рис. 17. Сходимость параметров полиэдральной модели, случай среднего количества параметров



(a) Сходимость параметров

(b) Значение ошибки

Рис. 18. Сходимость параметров полиэдральной модели, случай большого количества параметров

ров. Алгоритм переобучается с первой же итерации, и ошибка на контроле начинает расти сразу. Приведенные результаты показывают необходимость применения метода снижения размерности в случае большого количества параметров.

Алгоритм криволинейной регрессии. Сходимость параметров алгоритма криволинейной регрессии показана на рис. 19.

На рис. 19а показана сходимость весов регрессии \mathbf{u} , на рис. 19б — сходимость элементов вектора \mathbf{b}_0 . По оси абсцисс отложено количество итераций, по оси ординат — количественное значение каждого признака. Видно, что сходимость наблюдается на десятой итерации.

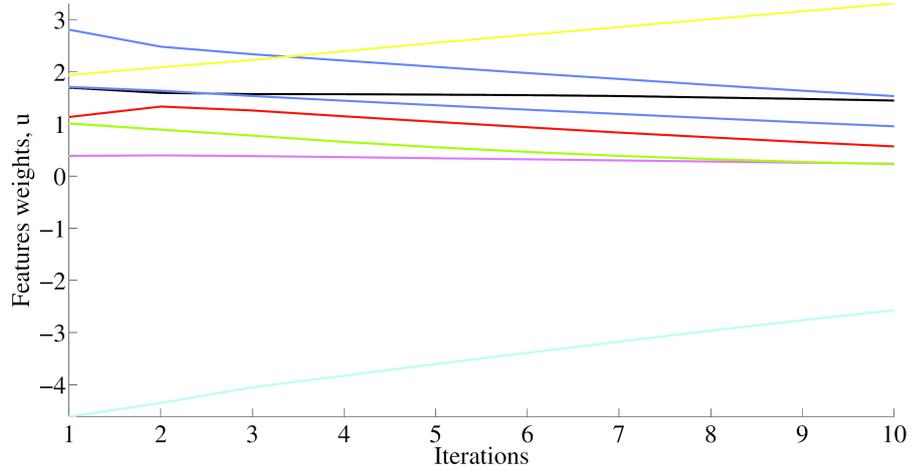
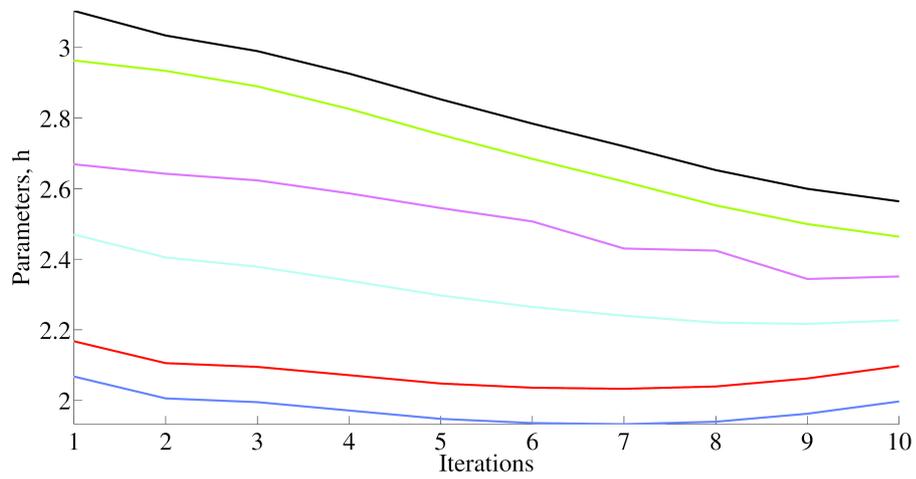
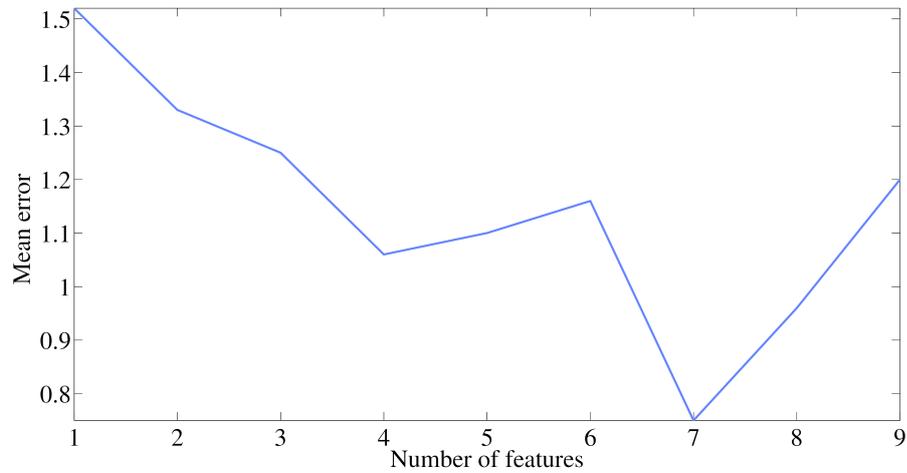
На рис. 19с показана зависимость функции ошибки от количества выбираемых признаков. Видно, что ее минимум достигается при семи признаках, и значение средней ошибки равно $L_H = 0.75$.

Алгоритм на основе копул. Рис. 20 иллюстрирует зависимость функции копулы от количества выбираемых признаков при решении задачи категоризации Красной книги.

На рис. 20а показана зависимость ошибки классификации от количества выбранных признаков. Оптимальное значение достигается при $|\mathcal{A}| = 4$. В исходной таблице данных эти признаки индексированы номерами 22, 24, 23 и 20. На рис. 20б показана зависимость параметра копулы от количества выбранных признаков. Видно, что значение параметра монотонно убывает с ростом количества признаков.

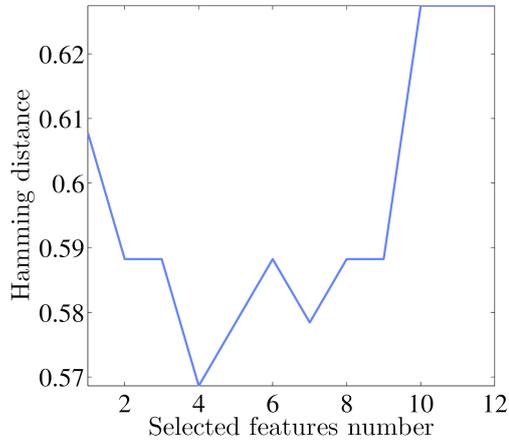
Алгоритм оценивания матрицы частичного порядка. Рис. 21 иллюстрирует первый этап алгоритма из раздела 2.2, оценку матрицы $\hat{\mathbf{Z}}$ по матрицам $\mathbf{Z}_0, \mathbf{Z}_1, \dots, \mathbf{Z}_n$ (20). На рис. 21 показаны исходная \mathbf{Z}_0 и восстановленная $\hat{\mathbf{Z}}$ матрицы для задачи категоризации Красной книги. Для этой иллюстрации выбраны объекты с тремя экспертно заданными категориями (CR — критические, EN — в опасности, VU — в уязвимости), и для них построена бинарная матрица \mathbf{Z}_0 ступенчатого вида, соответствующая целевому отношению предпочтения z_0 , показанная на рис. 21(а). На рис. 21(б) показана восстановленная матрица $\hat{\mathbf{Z}}$. Каждый элемент матрицы $\hat{\mathbf{Z}}(i, k)$ показывает степень доминирования объекта \mathbf{x}_k объектом \mathbf{x}_i .

На рис. 22 показан результат работы второго этапа алгоритма для выборки из Красной книги. Для демонстрации было выбрано три столбца матрицы $\hat{\mathbf{Z}}$, поэтому пространство признаков — трехмерное. Цвет маркера показывает исходную экспертную классификацию, цвет кружка вокруг маркера — классификацию, построенную алгоритмом. На рис. 22, в

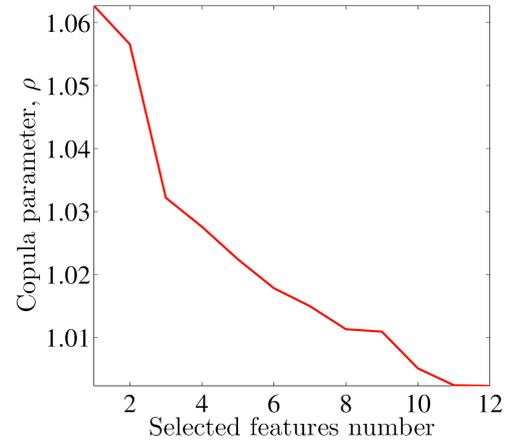
(a) Сходимость весов признаков, \mathbf{u} (b) Сходимость элементов вектора \mathbf{b}_0 

(c) Зависимость функции ошибки от количества выбираемых признаков

Рис. 19. Сходимость параметров метода криволинейной регрессии

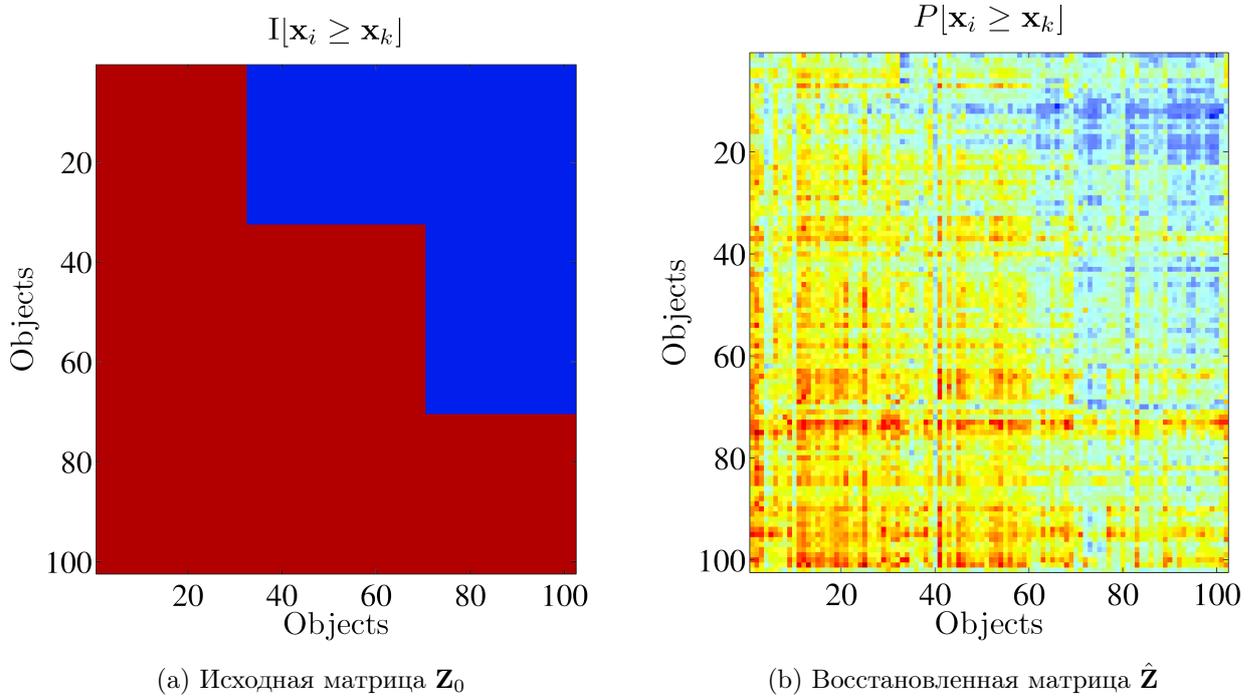


(a) Зависимость ошибки классификации



(b) Зависимость параметра копулы

Рис. 20. Зависимость классификации методом копулы от количества выбранных признаков

Рис. 21. Пример исходной матрицы \mathbf{Z}_0 и восстановленной матрицы $\hat{\mathbf{Z}}$ в задаче категоризации Красной книги

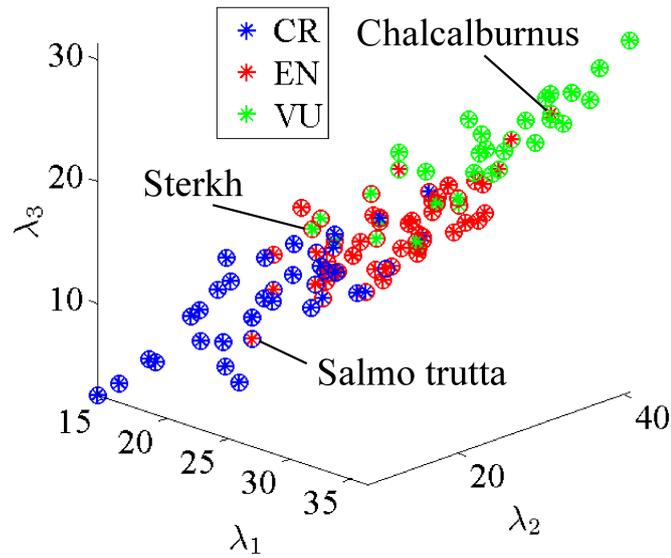


Рис. 22. Иллюстрация классификации в задаче категоризации Красной книги

частности, выделены объекты, отмеченные моделью как «выбросы», классифицированные экспертами глубоко внутри чужого класса.

Сравнение алгоритмов. В табл. 5 показано сравнение перечисленных методов. В качестве функции ошибки рассматривается средняя потеря Хэмминга (56) между метками классов на контрольной выборке. Разбиение на обучение и контроль проводилось методом leave-one-out, в таблице представлена усредненная ошибка на контрольном объекте по 102 запускам каждого алгоритма. Статистически лучший результат показал алгоритм на основе конусов из раздела 2.3.

Таблица 5. Категоризация Красной книги: сравнение алгоритмов

Алгоритм	Средняя потеря Хэмминга
Матрица частичного порядка	0.58
Конусы	0.52
Копулы	0.59
Криволинейная регрессия	0.71
Trees	0.55
SVM	0.66
kNN	0.72

Таблица 6. Описание используемых наборов данных

Набор данных	Признаков	Объектов	Обучение/контроль
Pyrimidines	27	74	50/24
MachineCPU	6	209	150/59
Boston	13	506	300/206
Computer	21	8182	4000/4182
Abalone	8	4177	1000/3177
Cars	6	1728	1000/728
RedBook	101	102	100/1

4.2 Эксперименты на данных UCI

Предложенные методы восстановления отношения предпочтения были использованы для решения задачи порядковой классификации. Для сравнения методов были использованы наборы данных из репозитория UCI. Для того, чтобы протестировать методы на данных порядкового типа, было проведено дополнительное монотонное преобразование признаков для каждого набора данных. Было проведено сравнение методов с двумя стандартными алгоритмами порядковой классификации и методом, учитывающим монотонные ограничения на признаки.

Были использованы следующие наборы данных из репозитория UCI: Pyrimidines, Machine CPU, Housing, Computer Activity, Abalone и Car. Перечисленные наборы данных, кроме последнего, относятся к задаче регрессии, поэтому была проведена дополнительная процедура дискретизации целевой переменной на пять уровней, содержащих равное число объектов, для решения задачи пятиклассовой порядковой классификации.

Схема эксперимента повторяет схему из работ [79, 19]. Каждый набор данных был разбит случайным образом на обучающую и контрольную части, как показано в таблице 6. Это разбиение было проведено независимо 100 раз. Для оценки качества была использована средняя абсолютная потеря L_a ,

$$L_a(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{m} \sum_{i=1}^m [y_i \neq \hat{y}_i],$$

и средняя потеря Хэмминга L_H , определяемая формулой (56).

Было проведено сравнение алгоритма на основе взвешивания матрицы частичного порядка (OW) из раздела 2.2, а также с методом классификации на основе Парето-фронт (POF), предложенных авторами в [79].

Таблица 7. Результаты на данных UCI: линейные признаки

Набор данных	Средняя абсолютная потеря (± 0.01)					Средняя потеря Хэмминга (± 0.01)				
	SVM	POF	Trees	OW	KNN	SVM	POF	Trees	OW	KNN
Pyrimidines	0.50	0.62	0.61	0.54	0.55	0.64	0.90	0.84	0.75	0.75
MachineCPU	0.44	0.44	0.47	0.42	0.51	0.53	0.53	0.53	0.49	0.61
Boston	0.38	0.48	0.41	0.39	0.47	0.46	0.65	0.47	0.46	0.62
Computer	0.32	0.71	0.38	0.34	0.60	0.35	1.36	0.41	0.39	0.90
Abalone	0.53	0.59	0.57	0.56	0.60	0.78	0.92	0.77	0.81	0.88

Таблица 8. Результаты на данных UCI: порядковые признаки

Набор данных	Средняя абсолютная потеря (± 0.01)					Средняя потеря Хэмминга (± 0.01)				
	SVM	POF	Trees	OW	KNN	SVM	POF	Trees	OW	KNN
Pyrimidines	0.57	0.58	0.60	0.62	0.49	0.71	0.77	0.79	0.79	0.76
MachineCPU	0.51	0.39	0.47	0.40	0.43	0.65	0.45	0.56	0.47	0.51
Boston	0.40	0.48	0.40	0.43	0.41	0.49	0.68	0.46	0.50	0.51
Computer	0.44	0.69	0.41	0.37	0.45	0.53	1.38	0.45	0.44	0.55
Abalone	0.78	0.59	0.57	0.58	0.59	1.78	0.92	0.76	0.85	0.89
Cars	0.19	0.19	0.08	0.16	0.06	0.24	0.26	0.08	0.19	0.07
RedBook	0.56	0.61	0.50	0.49	0.59	0.66	0.74	0.55	0.59	0.72

Для дополнительного сравнения качества было использовано два алгоритма порядковой классификации, метод решающих деревьев (Trees) и метод опорных векторов (SVM), которые были скомбинированы со схемой порядковой классификации, описанной в [38]. Кроме того, был использован метод классификации на основе ближайших соседей с монотонными ограничениями (KNN), описанный в [27].

Результаты для пяти искомых наборов данных (всех перечисленных, кроме «Cars») представлены в таблице 7. Жирным выделены случаи, когда соответствующий алгоритм показал статистически лучший результат по сравнению с остальными для 100 разбиений данных. Видно, что хотя метод SVM превосходит остальные методы из-за линейной природы признаков, тем не менее, предложенный метод OW срабатывает лучше на наборах данных MachineCPU и Boston.

Для сравнения методов на данных порядковой природы была произведена порядковая трансформация признаков исходных наборов данных. Также, как и в случае с целевой переменной, была произведена дискретизация каждого признака на пять уровней, содержащих одинаковое количество объектов.

Результаты для преобразованных наборов данных, а также для набора «Cars», признаки которого изначально имели порядковую природу, показаны в таблице 8. Метод SVM для данных такого типа сработал значительно хуже, а методы Trees и KNN показали лучший

Таблица 9. Ранжирование заповедников: сравнение алгоритмов

	\hat{y}_{OLS}	\hat{y}_γ	\hat{y}_{Cheb}	\hat{y}_{Cones}
Pearson, r	0.69	0.55	0.6	0.66
Kendall, τ	0.47	0.47	0.38	0.51

результат. Рассматриваемый метод OW продемонстрировал лучший результат в терминах средней абсолютной потери.

Результаты тестирования для задачи категоризации Красной книги показаны в последней строке таблицы 8. В отличие от остальных наборов данных, для этой задачи была использована схема разбиения leave-one-out, как и в подразделе 4.1. Предложенный метод продемонстрировал наилучший результат наряду с методом Trees.

4.3 Ранжирование заповедников для оценки эффективности их управления

В этом разделе приведем результаты работы алгоритмов согласования экспертных оценок, описанных в главе 3. Результаты работы проиллюстрированы на примере задачи ранжирования заповедников из табл. 1. Для решения задачи по оценке эффективности управления заповедниками были собраны ежегодные отчеты заповедников за 1995-2000 гг., включающие экспертные оценки интегральных индикаторов заповедников и экспертные оценки весов показателей ежегодных отчетов. Подробное описание данных приведено в работе [69].

Анализ точности предлагаемых алгоритмов. Результаты предложенных алгоритмов продемонстрированы в табл. 9. За критерий качества предлагается принять коэффициент корреляции между экспертной оценкой объектов y_0 и вычисленной оценкой \hat{y} . Предлагается использовать коэффициент корреляции Пирсона, $r(y_0, \hat{y})$, для измерения качества в линейной шкале, и коэффициент корреляции Кендалла, $\tau(y_0, \hat{y})$, для измерения качества в ранговой шкале. Отметим, что оценка \hat{y} для всех алгоритмов строилась методом Leave-One-Out, что позволяет оценить обобщающую способность предлагаемых алгоритмов.

Для иллюстрации результатов предложено взять оценки объектов \hat{y} , построенные четырьмя алгоритмами.

1. \hat{y}_{OLS} — оценка, полученная экспертно-статистическим методом (48),
2. \hat{y}_γ — оценка, полученная методом γ -согласования (52),

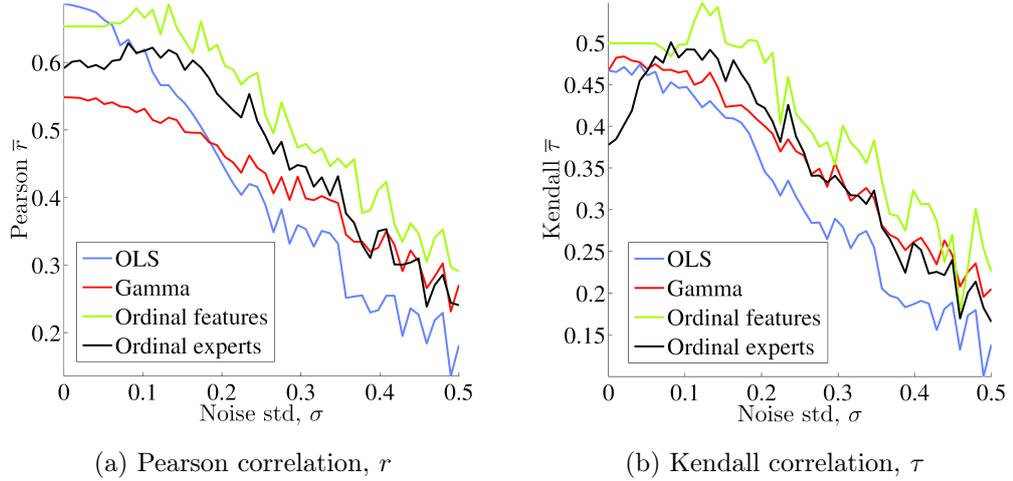


Рис. 23. Анализ устойчивости критериев качества при возмущении матрицы \mathbf{X}

3. $\hat{\mathbf{y}}_{\text{Cheb}}$ — оценка, полученная методом поиска устойчивой точки конуса (55),
4. $\hat{\mathbf{y}}_{\text{Cones}}$ — оценка, полученная для непараметрического порядкового признакового описания (38).

Результаты в табл. 9 показывают, что алгоритм стандартной линейной регрессии (OLS) демонстрирует наилучший результат в смысле линейной корреляции Пирсона. Однако подход на основе конусов, разработанный для учета порядковой информации, является лучшим в смысле порядковой τ -корреляции.

Анализ устойчивости предлагаемых алгоритмов. В качестве анализа устойчивости исследуемых алгоритмов предлагается возмутить элементы матрицы \mathbf{X} . Для этого рассмотрим матрицу $\mathbf{X}_{\Delta} = \mathbf{X} + \Delta$, где $\Delta(i, j) \sim \mathcal{N}(0, \sigma)$. Будем изменять дисперсию возмущения σ от минимального значения, $\sigma = 0$, до максимального значения σ_{\max} и следить, как изменяется при этом критерий качества для оценок $\hat{\mathbf{y}} = (\hat{\mathbf{w}}, \mathbf{X}_{\Delta})$. Результаты для различных методов показаны на рис. (23).

Заключение

Основные результаты диссертационной работы заключаются в следующем.

В главе 1 рассмотрены различные задачи обучения по предпочтениям. Для описания отношения предпочтения предложено использовать граф, соответствующий частично упорядоченному множеству, и конус предпочтений. Рассмотрено два представления конуса предпочтений: полиэдральное и порождающее. Установлено, что столбцы матрицы предпочтений являются порождающими элементами конуса предпочтений. Кроме того, с использованием матрицы предпочтений получены обобщенные формулы для известных порядковых метрик, таких как τ -корреляция и площадь под кривой.

Глава 2 посвящена задаче восстановления отношения предпочтения на множестве объектов, заданных порядковым описанием. Для решения этой задачи предложено несколько подходов. Во-первых, предложен подход на основе восстановления матрицы предпочтений, основанный на максимизации обобщенной τ -корреляции. Во-вторых, разработан метод на основе построения суммы Минковского конусов, соответствующих порядковым признакам. Предложен ряд методов оценки параметров построенной суперпозиции. Установлено соответствие использования центральных точек конусов с задачей восстановления матрицы предпочтений. Кроме того, предложено два альтернативных метода: метод криволинейной регрессии и метод на основе функции копулы.

В главе 3 разработаны методы согласования экспертных оценок. Введено понятие согласованности оценок. Рассмотрено два типа оценок: линейные и порядковые. В линейном случае предложена процедура согласования оценок с использованием структурного параметра. В порядковом случае предложен метод согласования на основе минимизации расстояния между векторами в конусах, соответствующих порядковым экспертным оценкам.

Глава 4 посвящена анализу прикладных задач. Приведено описание программного комплекса методов построения модели предпочтений. Приведено решение задачи категоризации редких видов из Красной книги РФ, а также задачи ранжирования заповедников. Кроме того, проведено сравнение разработанных методов для решения задачи порядковой классификации на стандартных наборах данных из репозитория UCI. Разработанные методы показывают улучшение качества в сравнении с известными подходами.

Список иллюстраций

1	Разложение точки по порождающим элементам конуса	26
2	Центральная точка конуса	40
3	Проекция точки \mathbf{y}_0 на конус \mathcal{X}	45
4	Коцула Клейтона	55
5	Иллюстрация метода главных компонент	61
6	Сравнение метода взвешенной суммы с оценкой $\hat{\mathbf{y}}_{PCA}$, полученной методом главных компонент	62
7	Иллюстрация метода α -согласования	65
8	Иллюстрация метода γ -согласования	66
9	Иллюстрация α - и γ -согласования	67
10	Сравнение экспертно-статистического метода и метода γ -согласования	67
11	Иллюстрация конусов, соответствующих экспертным оценкам	68
12	Нахождение ближайших векторов в конусах	70
13	Иллюстрация вектора \mathbf{y}_p как наиболее устойчивого решения	73
14	Сравнение оценок, полученных методом с порядковыми экспертными данными и методом с порядковыми признаками	74
15	Признаки в задаче категоризации таксонов	77
16	Сходимость параметров полиэдральной модели, случай малого количества па- раметров	81
17	Сходимость параметров полиэдральной модели, случай среднего количества параметров	82
18	Сходимость параметров полиэдральной модели, случай большого количества параметров	82
19	Сходимость параметров метода криволинейной регрессии	84
20	Зависимость классификации методом коцулы от количества выбранных при- знаков	85

21	Пример исходной матрицы \mathbf{Z}_0 и восстановленной матрицы $\hat{\mathbf{Z}}$ в задаче категоризации Красной книги	85
22	Иллюстрация классификации в задаче категоризации Красной книги	86
23	Анализ устойчивости критериев качества при возмущении матрицы \mathbf{X}	90

Список таблиц

1	Задача ранжирования заповедников, используемые данные	59
2	Экспертное признаковое описание объектов Красной книги РФ	78
3	Доминирование признаков: биологические	79
4	Доминирование признаков: воздействие и угрозы	80
5	Категоризация Красной книги: сравнение алгоритмов	86
6	Описание используемых наборов данных	87
7	Результаты на данных UCI: линейные признаки	88
8	Результаты на данных UCI: порядковые признаки	88
9	Ранжирование заповедников: сравнение алгоритмов	89

Список основных обозначений

x — исследуемый объект

\mathbb{X} — множество значений объекта, $\mathbf{x} \in \mathbb{X}$

X — множество объектов размера m , $X = \{x_1, \dots, x_m\}$

z_1, \dots, z_n — набор отношений предпочтения, заданных на X

z_0 — целевое отношение предпочтения, заданное на X

Y — множество значений целевой переменной

\mathbf{X} — матрица плана, $\mathbf{X} = [\mathbf{x}_1^\top; \dots; \mathbf{x}_m^\top]$

\mathbf{y} — вектор значений целевой переменной, $\mathbf{y} = [y_1; \dots; y_m]$

f — модель, отображение множества объектов X во множество \mathbb{R} или Y

\mathbf{w} — вектор параметров модели $f(\mathbf{w}, \mathbf{x})$

$S(f, X, z_0)$ — функция потерь модели f на объектах из X

$\hat{\mathbf{w}}$ — оптимальные параметры, минимизирующие функцию потерь на X

\mathbf{w}_0 — значение вектора параметров, заданное экспертом

\mathcal{W} — допустимое множество значений параметров, $\mathbf{w} \in \mathcal{W}$

\mathbf{Z} — матрица инцидентности графа, соответствующего графу конечного частично упорядоченного множества

\mathcal{X} — конус, соответствующий конечному частично упорядоченному множеству

ζ_k — k -й порождающий элемент конуса \mathcal{X}

$\boldsymbol{\lambda}$ — неотрицательный вектор параметров разложения точки конуса

$\hat{\mathbf{Z}}$ — матрица отношения доминирования объектов

$C : [0, 1]^n \rightarrow [0, 1]$ — функция копулы

Литература

- [1] Adomavicius Gediminas, Tuzhilin Alexander. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions // Knowledge and Data Engineering, IEEE Transactions on. — 2005. — V. 17, no. 6. — P. 734–749.
- [2] Ailon Nir, Charikar Moses, Newman Alantha. Aggregating inconsistent information: ranking and clustering // Journal of the ACM (JACM). — 2008. — V. 55, no. 5. — P. 23.
- [3] Воронцов Константин Вячеславович. Локальные базисы в алгебраическом подходе к проблеме распознавания: дис. . . . канд. физ.-мат. наук / Константин Вячеславович Воронцов. — диссертация на соискание ученой степени к. ф.-м. н., М.: ВЦ РАН. — 1999.
- [4] Arrow Kenneth J. 1963. social choice and individual values // Wiley, New York. — 1951.
- [5] Arrow Kenneth J. Social choice and individual values. — Yale university press, 2012. — V. 12.
- [6] Asuncion Arthur, Newman David. Uci machine learning repository. — 2007.
- [7] Baeza-Yates Ricardo, Ribeiro-Neto Berthier et al. Modern information retrieval. — ACM press New York, 1999. — V. 463.
- [8] Bartholdi III John, Tovey Craig A, Trick Michael A. Voting schemes for which it can be difficult to tell who won the election // Social Choice and welfare. — 1989. — V. 6, no. 2. — P. 157–165.
- [9] Boyd S., Vandenberghe L. Convex Optimization. — Cambridge University Press, 2006.
- [10] Learning to rank using gradient descent / Chris Burges, Tal Shaked, Erin Renshaw et al. // Proceedings of the 22nd international conference on Machine learning / ACM. — 2005. — P. 89–96.
- [11] Top-k selection based on adaptive sampling of noisy preferences / R. Busa-Fekete, B. Szoreny, P. Weng et al. // Proc. ICML-13, 30th International Conference on Machine Learning (JMLR). — 2013. — V. 28, no. 3. — P. 1130–1138.

- [12] Adapting ranking svm to document retrieval / Yunbo Cao, Jun Xu, Tie-Yan Liu et al. // Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval / ACM. — 2006. — P. 186–193.
- [13] Learning to rank: from pairwise approach to listwise approach / Zhe Cao, Tao Qin, Tie-Yan Liu et al. // Proceedings of the 24th international conference on Machine learning / ACM. — 2007. — P. 129–136.
- [14] Cardoso Jaime S, Da Costa Joaquim F Pinto. Learning to classify ordinal data: The data replication method. // Journal of Machine Learning Research. — 2007. — V. 8, no. 1393–1429. — P. 6.
- [15] Charnes A, Cooper WW. The strong minkowski farkas-weyl theorem for vector spaces over ordered fields // Proceedings of the National Academy of Sciences of the United States of America. — 1958. — V. 44, no. 9. — P. 914.
- [16] Chen Donghui, Plemmons Robert J. Nonnegativity constraints in numerical analysis // Symposium on the Birth of Numerical Analysis. — 2009. — P. 109–140.
- [17] Ranking measures and loss functions in learning to rank / Wei Chen, Tie-Yan Liu, Yanyan Lan et al. // Advances in Neural Information Processing Systems. — 2009. — P. 315–323.
- [18] Predicting partial orders: Ranking with abstention / Weiwei Cheng, Michael Rademaker, Bernard De Baets, Eyke Huellermeier // Machine Learning and Knowledge Discovery in Databases. — 2010. — V. 6321. — P. 215–230.
- [19] Chu Wei, Keerthi S Sathiya. Support vector ordinal regression // Neural computation. — 2007. — V. 19, no. 3. — P. 792–815.
- [20] Cohen William W., Schapire Robert E., Singer Yoram. Learning to order things // J. Artif. Int. Res. — 1999. — V. 10, no. 1. — P. 243–270.
- [21] Conitzer Vincent, Davenport Andrew, Kalagnanam Jayant. Improved bounds for computing kemeny rankings // AAAI. — V. 6. — 2006. — P. 620–626.
- [22] Robust ordinal regression in preference learning and ranking / Salvatore Corrente, Salvatore Greco, Milosz Kadzinski, Roman Slowinski // Machine Learning. — 2013. — V. 93. — P. 381–422.

- [23] Cossock David, Zhang Tong. Subset ranking using regression // Learning theory. — Springer, 2006. — P. 605–619.
- [24] Danilov V. I., Sotskov A. I. Social Choice Mechanisms. — Springer-Verlang, 2002.
- [25] Davenport Andrew, Kalagnanam Jayant. A computational study of the kemeny rule for preference aggregation // AAAI. — V. 4. — 2004. — P. 697–702.
- [26] Dembczyński Krzysztof, Kotłowski Wojciech, Słowiński Roman. Ordinal classification with decision rules // Mining Complex Data. — Springer, 2008. — P. 169–181.
- [27] Duivesteijn Wouter, Feelders Ad. Nearest neighbour classification with monotonicity constraints // Machine Learning and Knowledge Discovery in Databases / Ed. by Walter Daelemans, Bart Goethals, Katharina Morik. — Springer Berlin Heidelberg, 2008. — V. 5211 of Lecture Notes in Computer Science. — P. 301–316.
- [28] Литвак Б. Г. Экспертная информация: методы получения и анализа. — Радио и связь, 1982.
- [29] Ehrgott Matthias. Multicriteria optimization. — Springer, 2005. — V. 2.
- [30] Кузнецов М. П., Стрижов В. В., Медведникова М.М. Алгоритм многоклассовой классификации объектов, описанных в ранговых шкалах // Научно-технический вестник СПб ГПУ. Информатика. Телекоммуникации. Управление. — 2012. — Т. 5. — С. 92–95.
- [31] Кузнецов М.П. Уточнение ранговых экспертных оценок с использованием монотонной интерполяции // Машинное обучение и анализ данных. — 2011. — Т. 1, № 2. — С. 154–162.
- [32] Кузнецов М.П. Построение интегрального индикатора в ранговых шкалах с использованием копул для анализа совместного распределения критериев // Машинное обучение и анализ данных. — 2012. — Т. 1. — С. 411–419.
- [33] Кузнецов М.П. Комбинирование отношений порядка для восстановления предпочтения на наборе объектов // Математические методы распознавания образов ММРО-15. Тезисы докладов 17-й Всероссийской конференции с международным участием. г. Светлогорск, Калининградская обл. М.: Торус пресс. — 2015. — С. 18–19.
- [34] Кузнецов М.П., Ивкин Н.П. Алгоритм классификации временных рядов акселерометра по комбинированному признаковому описанию // Машинное обучение и анализ данных. — 2015. — Т. 1, № 11. — С. 1471–1483.

- [35] Кузнецов М.П., Стрижов В.В. Уточнение ранговых экспертных оценок с использованием монотонной интерполяции // Математические методы распознавания образов ММРО-15. Доклады 15-й Всероссийской конференции. М.: МАКС Пресс. — 2011. — С. 162–165.
- [36] Кузнецов М.П., Стрижов В.В. Построение интегрального индикатора с использованием ранговой матрицы описаний // Интеллектуализация обработки информации. Доклады 9-й международной конференции. М.: Торус пресс. — 2012. — С. 130–132.
- [37] Fogel Efi, Halperin Dan. Exact and efficient construction of minkowski sums of convex polyhedra with applications // Computer-Aided Design. — 2007. — V. 39. — P. 929–940.
- [38] Frank Eibe, Hall Mark. A simple approach to ordinal classification. — Springer, 2001.
- [39] Frees Edward W., Valdez Emiliano A. Understanding relationships using copulas // North american actuarial journal. — V. 2. — P. 104–141.
- [40] An efficient boosting algorithm for combining preferences / Yoav Freund, Raj Iyer, Robert E Schapire, Yoram Singer // The Journal of machine learning research. — 2003. — V. 4. — P. 933–969.
- [41] Fuernkranz Johannes, Huellermeier Eyke. Preference learning. — Springer, 2011.
- [42] Golub Gene H, Van Loan Charles F. Matrix computations. — JHU Press, 2012. — V. 3.
- [43] Hang LI. A short introduction to learning to rank // IEICE TRANSACTIONS on Information and Systems. — 2011. — V. 94, no. 10. — P. 1854–1862.
- [44] Herbrich Ralf, Graepel Thore, Obermayer Klaus. Large margin rank boundaries for ordinal regression // Advances in neural information processing systems. — 1999. — P. 115–132.
- [45] Hoeffling Holger. A path algorithm for the fused lasso signal approximator // Journal of Computational and Graphical Statistics. — 2010. — V. 19. — P. 984–1006.
- [46] Медведникова М. М., Стрижов В. В., Кузнецов М. П. Алгоритм многоклассовой монотонной парето-классификации с выбором признаков // Известия Тульского государственного университета, Естественные науки. — 2012. — Т. 3. — С. 132–141.
- [47] Подиновский В. В. Введение в теорию важности критериев в многокритериальных задачах принятия решений // М.: Физматлит. — 2007. — Т. 7.
- [48] Орлов А.И. Экспертные оценки // Учебное пособие. — 2002. — Т. 31.

- [49] Jolliffe I. T. *Principal Component Analysis*. — Springer, 2002.
- [50] Jullien-Ramasso S., Mauris G., L. Valet Ph. Bolon. A decision support system for animated film selection based on a multi-criteria aggregation of referees' ordinal preferences // *Expert Systems with Applications*. — 2012. — V. 39. — P. 4250–4257.
- [51] Kaci Souhila. *Working with Preferences: Less Is More*. — Springer Berlin Heidelberg, 2011.
- [52] Metric concentration search procedure using reduced matrix of pairwise distances / A. M. Katrutsa, M. P. Kuznetsov, V. V. Strijov, K. V. Rudakov // *Intelligent Data Analysis*. — 2015. — V. 19, no. 5. — P. 1091–1108.
- [53] Kemeny John G. *Mathematics without numbers* // *Daedalus*. — 1959. — V. 88, no. 4. — P. 577–591.
- [54] Kendall Maurice G. A new measure of rank correlation // *Biometrika*. — 1938. — P. 81–93.
- [55] Khurshid Anwer, Sahai Hardeo. Scales of measurements: An introduction and a selected bibliography // *Quality and Quantity*. — 1993. — V. 27, no. 2684. — P. 303–324.
- [56] Kotlowski Wojciech, Slowinski Roman. On nonparametric ordinal classification with monotonicity constraints // *IEEE Transactions on Knowledge and Data Engineering*. — 2013. — V. 25, no. 11. — P. 2576–2589.
- [57] Supervised topic classification for modeling a hierarchical conference structure / Mikhail Kuznetsov, Marianne Clausel, Massih-Reza Amini et al. // *International Conference on Neural Information Processing* / Springer. — 2015. — P. 90–97.
- [58] Kuznetsov M.P., Strijov V.V. Methods of expert estimations concordance for integral quality estimation // *Expert Systems with Applications*. — 2014. — March. — V. 41, no. 4. — P. 1988–1996.
- [59] Lawson Charles L, Hanson Richard J. *Solving least squares problems*. — SIAM, 1974. — V. 161.
- [60] Levin Jonathan, Nalebuff Barry. An introduction to vote-counting schemes // *The Journal of Economic Perspectives*. — 1995. — P. 3–26.
- [61] Liu Tie-Yan. Learning to rank for information retrieval // *Foundations and Trends in Information Retrieval*. — 2009. — V. 3, no. 3. — P. 225–331.

- [62] Supervised rank aggregation / Yu-Ting Liu, Tie-Yan Liu, Tao Qin et al. // Proceedings of the 16th international conference on World Wide Web / ACM. — 2007. — P. 481–490.
- [63] A model for multi-label classification and ranking of learning objects / Vivian F. Lopez, Fernando de la Prieta, Mitsunori Ogihara, Ding Ding Wong // Expert Systems with Applications. — 2012. — V. 39. — P. 8878–8884.
- [64] McCullagh Peter. Regression models for ordinal data // Journal of the royal statistical society. Series B (Methodological). — 1980. — P. 109–142.
- [65] Moshkovich Helen M., Mechitov Alexander I., Olson David L. Rule induction in data mining: effect of ordinal scales // Expert Systems with Applications. — 2002. — V. 22. — P. 303–311.
- [66] Nelsen Roger B. An Introduction to Copulas. — Springer, 1998.
- [67] Стрижов В. В. Уточнение экспертных оценок, выставленных в ранговых шкалах, с помощью измеряемых данных // Заводская лаборатория. Диагностика материалов. — 2011. — Т. 77, № 7. — С. 72–78.
- [68] Стрижов В. В., Кузнецов М. П., Рудаков К. В. Метрическая кластеризация последовательностей аминокислотных остатков в ранговых шкалах // Математическая биология и биоинформатика. — 2012. — Т. 7(1). — С. 345–359.
- [69] Стрижов Вадим Викторович. Согласование экспертных оценок при построении интегральных индикаторов: дис. ... канд. физ.-мат. наук / Вадим Викторович Стрижов. — диссертация на соискание ученой степени к. ф.-м. н., М.: ВЦ РАН. — 2002.
- [70] Parkes David C, Xia Lirong. A complexity-of-strategic-behavior comparison between schulze's rule and ranked pairs. // AAAI. — 2012.
- [71] Saari Donald G. Mathematics and voting // Notices of the American Mathematical Society. — 4/2008.
- [72] Saari Donald G, Merlin Vincent R. A geometric examination of kemeny's rule // Social Choice and Welfare. — 2000. — V. 17, no. 3. — P. 403–438.
- [73] Collaborative filtering recommender systems / J. Ben Schafer, Dan Frankowski, Jon Herlocker, Shilad Sen // Lecture Notes in Computer Science. — 2007. — V. 4321. — P. 291–324.

- [74] Schalekamp Frans, van Zuylen Anke. Rank aggregation: Together we're strong. // *ALENEX / SIAM*. — 2009. — P. 38–51.
- [75] Shashua Amnon, Levin Anat. Ranking with large margin principle: Two approaches // *Advances in neural information processing systems*. — 2002. — P. 937–944.
- [76] Slawski Martin, Hein Matthias et al. Non-negative least squares for high-dimensional linear models: Consistency and sparse recovery without regularization // *Electronic Journal of Statistics*. — 2013. — V. 7. — P. 3004–3056.
- [77] Spirin Nikita, Vorontsov Konstantin. [Learning to rank with nonlinear monotonic ensemble](#) // *Multiple Classifier Systems* / Ed. by Carlo Sansone, Josef Kittler, Fabio Roli. — Springer Berlin Heidelberg, 2011. — V. 6713 of *Lecture Notes in Computer Science*. — P. 16–25.
- [78] Relevance-aware filtering of tuples sorted by an attribute value via direct optimization of search quality metrics / Nikita V Spirin, Mikhail Kuznetsov, Julia Kiseleva et al. // *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* / ACM. — 2015. — P. 979–982.
- [79] Stenina M.M., Kuznetsov M.P, Strijov V.V. Ordinal classification using pareto fronts // *Expert Systems with Applications*. — 2015. — V. 42, no. 14. — P. 5947–5953.
- [80] Strijov Vadim, Granic Goran, et al. Integral indicator of ecological impact of the croatian thermal power plants // *Energy*. — 2011. — V. 36, no. 7. — P. 4144–4149.
- [81] Tibshirani Ryan Joseph. *The solution path of the generalized lasso*. — Stanford University, 2011.
- [82] Tideman T Nicolaus. Independence of clones as a criterion for voting rules // *Social Choice and Welfare*. — 1987. — V. 4, no. 3. — P. 185–206.
- [83] Trotman Andrew. *Learning to rank* // *Information Retrieval*. — 2005. — V. 8. — P. 381.
- [84] Uhanov M.V. Polygons sum construction algorithm // *Bulletin of South Ural State University*. — 2001. — V. 1, no. 7. — P. 39–44.
- [85] Vapnik Vladimir N. *Statistical Learning Theory*. — Wiley-Interscience, 1998.
- [86] Listwise approach to learning to rank: theory and algorithm / Fen Xia, Tie-Yan Liu, Jue Wang et al. // *Proceedings of the 25th international conference on Machine learning* / ACM. — 2008. — P. 1192–1199.

- [87] Xia Lirong, Conitzer Vincent. Determining possible and necessary winners under common voting rules given partial orders. // *AAAI*. — V. 8. — 2008. — P. 196–201.
- [88] Xia Lirong, Conitzer Vincent. A maximum likelihood approach towards aggregating partial orders // *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*. — V. 22. — 2011. — P. 446.
- [89] Young H Peyton. Condorcet's theory of voting // *American Political Science Review*. — 1988. — V. 82, no. 04. — P. 1231–1244.