

Анализ структуры курса “Автоматизация и стандартизация научных исследований в области машинного обучения и анализа данных” и его результатов*

Стрижов В. В., Мотренко А. П.

ВЦ РАН, Московский физико-технический институт

Разработан курс “Автоматизация и стандартизация научных исследований в области машинного обучения и анализа данных”. Курс состоит из ряда лекций, посвященных организации научных исследований и представления отчетности. В ходе курса каждый студент получает персональное задание, результатом которого является научная статья в формате, принятом в предметной области и программное обеспечение для проведения вычислительного результата. По завершению проекта студент представляет доклад в формате презентации; статья подается в рецензируемый журнал.

Предлагаемый курс читается на кафедре “Интеллектуальные системы” Московского физико-технического института с 2009. За это время в рамках курса было опубликовано более 160 студенческих проектов [1], из них 44 работы — в журнале “Машинное обучение и анализ данных” [2], созданном студентами в 2010 году в рамках выполнения работ по курсу.

Введение

«Численные методы обучения по прецедентам» — практические занятия, посвященные изучению методов научной работы и выполнения исследовательских проектов в области машинного обучения и анализа данных. Одной из задач курса является построить технологический процесс выполнения научно-исследовательской работы таким образом, чтобы он не становился препятствием в работе, но помогал в ее выполнении. Правильно построенный процесс позволяет стандартизировать научную работу, точно установив требования к результатам работы, но оставляя исследователю максимальную творческую свободу.

Также в задачи курса входит упрощение обмена результатами. В связи с огромным количеством доступной информации обмен и поиск релевантной информации затруднены. Поэтому особенно важно, чтобы результаты исследования были представлены в формате, позволяющем выделить их в общем информационном потоке и использовать в дальнейших исследованиях. Это одна из причин, по которой при работе в рамках курса не только не запрещаются, но и поощряются списывание, обмен информацией и результатами, использование чужих материалов с корректной ссылкой на источник; цитаты следует заключать в кавычки, например: «Мера — превыше всего» \cite{Gasparov2004Greece}. Такой подход к заимствованию результатов позволяет снизить вероятность переизобретения существующих методов и дублирования результатов. В результате наработки прошлых лет используются и улучшаются. Кроме того, при улучшении чужих результатов увеличивается личный вклад студента в работу, появляется новизна исследования.

Для достижения поставленных целей используются интернет-ресурс кафедры MachineLearning.ru [3] (обмен задачами и результатами), репозиторий SourceForge.org [4] (обмен алгоритмами), и система контроля версий SVN Tortoise [5] (синхронизация работы).

Курс разделен на три семестровых этапа, каждый из которых имеет свои цель и содержательный результат:

- 3-2 Написание научной статьи. Целью первого семестра является познакомить студентов с технологическим процессом научных исследований, а также научить студентов оформлять результаты своей работы в том виде, в котором они могут быть приняты научным сообществом. Итогом работы первого семестра становится научная статья, пригодная для подачи в рецензируемый журнал.
- 4-1 Создание отчета коммерческого исследовательского проекта. Во второй части курса студенты группой работают над проектом по одной теме, совмещая исследовательскую работу с определенной ролью в процессе написания программного обеспечения для выполнения коммерческого проекта (контроль интерфейсов, юнит-тестирование, системное тестирование и т.д.). Качественным результатом становится программный продукт и отчет, содержащий документацию к продукту с описанием предлагаемых алгоритмов и методов, и позволяющий использовать разработанный продукт не специалистам в области машинного обучения и анализа данных.
- 4-2 Исследование свойств алгоритмов машинного обучения и анализа данных. В рамках последнего семестра курса студентам предлагается набор теоретических и практических задач; для каждой из которых необходимо написать постановку задачи, исследовать литературу и предложить свое решение. Целью курса является расширение области знаний студентов в изучаемой области, углубление понимания уже известных алгоритмов и методов.

Статья посвящена описанию первого этапа — процессу написания статьи студентами третьего курса. Абстрактная цель первого этапа — научиться точно, ясно, красиво излагать свои и чужие идеи; конкретная — написать научную статью, которая была бы принята другими исследователями, работающими в области машинного обучения и анализа данных; сделать доклад.

Результат: научная статья, поданная в рецензируемый журнал из списка ВАК.

Предполагается, что студенты прошли следующие курсы: математический анализ, аналитическая геометрия, дискретный анализ, алгебра логики, комбинаторика, теория графов, основы высшей алгебры, теория формальных систем и алгоритмов, теория вероятностей, случайные процессы, математическая статистика, методы оптимизации, прикладная статистика. Приблизительный план работы студентов включает следующие пункты:

- Получение задания в общем виде.
- Уточнения задания, поиск литературы.
- *Написание статьи*: постановка задачи и алгоритм.
- Написание и отладка кода.
- *Вычислительный эксперимент*.
- *Кросс-рецензирование*.
- Презентация работы — доклад 15 минут.

Ключевые этапы работы над статьей перечислены в хронологическом порядке, однако некоторые пункты, как правило, выполняются параллельно. Например, работа над текстом и выполнение вычислительный эксперимент могут выполняться последовательно, но в несколько итераций. Рассмотрим подробнее выделенные пункты плана.

Написание статьи. Основные разделы, из которых состоит математическая статья, это “Введение”, “Постановка задачи”, “Описание алгоритма”, “Вычислительный эксперимент” и “Литература”. Написание первоначального варианта каждого из разделов сту-

денту отводится не больше недели; в процессе работы содержание и стиль текста дорабатываются.

Вычислительный эксперимент. Как и текст статьи, содержание вычислительного эксперимента зависит от темы исследования. Как правило, проведение вычислительного эксперимента включает в себя

- порождение модельных или загрузку реальных данных;
- визуализацию данных;
- выполнение алгоритма, получение результатов;
- визуализацию результатов;
- исследование свойств алгоритма;

Кросс-рецензирование позволяет подготовить завершённую статью к подаче в журналу и выносу на обсуждение в научное сообщество. Каждой работе назначается два рецензента из группы. Задачи рецензента:

- помочь коллеге донести мысль до читателя,
- проверить правильность полученных результатов,
- проверить стандартность кода,
- составить своё мнение о работе и быть готовым оппонировать.

Критерием качества является готовность использовать результаты рецензируемой работы как свои собственные.

В открытом доступе можно найти значительное количество вспомогательных материалов [6, 7, 8], дающих студентам возможность самостоятельно проводить небольшие исследования и описывать их в формате научных отчетов или статей. Это могут быть основанные на распространенных ошибках сборники советов по написанию научного текста (например, [6]) или подробные руководства по организации исследования [7, 8]. Перечисленные руководства уделяют особое внимание аспектам поиска и обработки литературы: от тонкостей поиска до стратегий чтения материала, составления записей и ранжирования источников на их основе. Также обсуждаются вопросы планирования исследования и структуры статьи, приводятся советы по написанию текста, доработке и подготовке финальной версии статьи к публикации. Тем не менее, результаты пяти лет чтения данного курса свидетельствуют, что проведение исследования, содержащего неоспоримую научную новизну, с последующей публикацией результатов в рецензируемом журнале для студентов третьего курса является чересчур сложной задачей даже под руководством преподавателя. Самостоятельно написать статью, удовлетворяющую требованиям журналов из списка «ВАК», студенту третьего курса практически невозможно: не хватает опыта ставить задачи формально, находить нужную литературу, порождать новые и актуальные идеи и решения задач. Поэтому за каждой из тем, предлагаемых студентам для исследования, закрепляется консультант и эксперт. С одной стороны, студенты недостаточно ориентируются в области, чтобы суметь самостоятельно выделить направление исследований, работая в котором они могли бы внести вклад в выбранную область. Эксперт является поставщиком задачи и владельцем данных и гарантирует новизну и актуальность работы. При составлении задачи эксперт должен иметь виду, что объем работ по семестровому проекту не должен превзойти 34 часов (время, отведенное студентам на самостоятельную работу по курсу, согласно учебному плану). Данные также необходимо подготовить заранее и предоставить их в таком виде, чтобы загрузка заняла не более часа. Для того чтобы студенты могли выбирать из списка предложенных задач, каждому эксперту необходимо предоставить описание задачи в формате [9]. Описание включает в себя на-

звание работы, описание исследуемых данных, цель работы и предполагаемый результат. При необходимости, с помощью преподавателя эксперт обеспечивает каждого студента предлагаемым направлением развития существующих методов в рамках решения данной задачи. С другой стороны, студентам не хватает опыта в организации исследовательского процесса и оформлении его результатов. Анализ результатов прошлых лет показывает, что основными причинами незавершения проекта являлись неясность в постановке задачи или направлении работы, а также трудности чисто технического характера (особенно на начальных этапах). В работе [10] исследуется влияние различных типов мотивации к выполнению задания на успешность его завершения и вероятность использования студентом техник самоорганизации (таких как планирование заданий, учет временных затрат, анализ обстановки). Авторы выделяют три источника мотивации:

- 1) вера в самоэффективность (уверенность в эффективности собственных действий, ожидание успешности результата),
- 2) уверенность в важности выполняемого задания и заинтересованность в нем,
- 3) целевая направленность — результат зависит от того, является ли основной целью студента приобретение навыка (мастерство), получение отметки (внешние цели), либо соревновательный момент, вызванный сравнением студентов между собой (относительное мастерство).

(Подробнее о типах мотивации, целевой направленности и подходов студентов к обучению можно прочитать, например, в обзоре [11].) Исследования, проведенные в [10] показали наличие положительного влияния веры в самоэффективность на академическую успеваемость и организованность студентов: студенты, более уверенные в эффективности своих действий, с большей вероятностью использовали все техники саморегулирования. При исследовании типов целевой направленности была обнаружена положительная корреляция между успеваемостью и определением целевой направленности мастерство или относительное мастерство; внешние цели были единственным источником мотивации, показавшим существенную отрицательную корреляцию с успеваемостью. Эти результаты подтверждают важность заинтересованности студента в выполнении проекта и отдельно — в выполнении каждого этапа работы. Подчеркнуть важность выполнения проекта в целом — задача преподавателя, поддерживать заинтересованность студента в течение выполнения проекта — задача консультанта. Консультант отвечает за то, что студент всегда понимает цель очередного этапа работ; также консультант отвечает на вопросы, с которыми студент не решается обращаться к преподавателю, оценивает выполнение каждого этапа работы, вносит правки и замечания. Консультант помогает студенту в пользовании инструментами, отвечает на вопросы по специальности, консультирует выполнение работ, оперативно реагирует на проблемы, проверяет результаты, ставит оценки. Задачи для курса выбираются и формулируются таким образом, чтобы для каждого консультанта нашлась задача в его области интересов. Предполагается, что консультант в данный момент сам пишет работу по этой теме. Консультантами становятся студенты старших курсов и аспиранты при личном желании и вместе с гарантией, что консультант доведет проект до конца и сможет мотивировать другого человека. Критерии для участия студента или аспиранта в качестве консультанта: две самостоятельно написанные и опубликованные в журналах из списка ВАК статьи. Это дает гарантию, что аспирант самостоятельно может править чужой и писать собственный текст трех типов:

- 1) терминологически точную и ясную передачу идеи работы и ее результатов,
- 2) мотивирующие пояснения в текстах,

- 3) описание хода рассуждений, пояснения к формулам,
- 4) описание постановок и решений задач на формальном языке.

Преподаватель контролирует работу группы в целом, следит за тем чтобы работы в каждой паре студент-консультант выполнялись регулярно, вносит дополнительные замечания по итогам каждого этапа работы.

Описываемый курс включает в себе комбинацию подхода обучение сверстника сверстником [12] (peer tutoring) с использованием ассистентов в учебном процессе. Преимущества участия ассистентов-консультантов в учебном процессе очевидны и перечислены, например, в [13]: студенты получают больше внимания, а у преподавателя появляется возможность посвятить больше времени организации курса в целом. На ранних стадиях (начальные и средние классы школы) обучения привлечение ассистентов к процессу обучения может привести к негативным последствиям [14]. Проведенный авторами [14] опрос ассистентов показывает, что участие ассистентов зачастую сводится к помощи при выполнении определенных заданий, нежели углублению понимания изучаемого материала. Опасность такого формата обучения заключается в его односторонности: подача информации происходит от ассистента к ученику, что делает процесс обучения пассивным и отрицательно сказывается на способности ученика самостоятельно решать задачи. Этих последствий можно избежать, обеспечив необходимую подготовку ассистентов. В описываемом курсе в роли ассистентов выступали консультанты — студенты пятых-шестых курсов и аспиранты. По сравнению со стандартной формой обучения студент-преподаватель, при обучении студента студентом или аспирантом разница в уровне знаний и статусе между “учеником” и “учителем” меньше, и, менее стесненный авторитетом консультанта, студент охотнее выражает свое мнение, задает вопросы, больше экспериментирует [12]. Согласно [12], успешность подобных программ основывается на четко определенных социальных ролях и процедурах взаимодействия между студентами, направленных на передачу определенного материала при непосредственном наблюдении и, при необходимости, вмешательстве преподавателя. В работе [15] проводится сравнение результатов студентов, осваивавших одинаковый материал в разных форматах: для одной группы читались только лекции, для другой количество лекционного материала было уменьшено, а оставшийся материал студенты изучали в группах. Результаты эксперимента показали не только улучшение успеваемости для второй группы, но и большую заинтересованность студентов в изучаемом предмете.

При подготовке курса преподавателю необходимо:

- 1) Заручиться поддержкой экспертов, убедиться что они готовы предоставить необходимые данные в нужный срок.
- 2) Заручиться поддержкой консультантов. Участие консультантов в данном курсе не вознаграждалось и производилось на добровольной основе. Чтобы мотивировать консультантов к участию в проекте, необходимо с каждым из них на основе предоставленного списка задач договориться о желаемой теме исследования.
- 3) Согласовать с экспертами предполагаемые темы задач.

В случае, если предложенных экспертами задач не хватает на всех студентов, выбираются задачи которое уже предлагались студентам прошлых лет, но не были выбраны или были недостаточно исследованы.

Чтобы добиться эффективного взаимодействия студентов и консультантов, преподаватель должен

- предоставить разработанную структуру курса и определить роль консультантов в рамках каждого пункта плана курса,
- убедиться, что консультанты сами понимают задание,
- дать четкие критерии оценивания студенческих работ,
- всячески поощрять обратную связь как со студентами, так и с консультантами.

В ходе курса, помимо наблюдения за отдельными проектами и контроля качества их выполнения, преподавателю необходимо проводить занятия с группой, обеспечивая студентам необходимую базу знаний и задавая общее направление работ. Чтобы познакомить студентов с основными принципами написания научного текста, предлагается [16] на лекциях приводить примеры ошибок при написании с объяснением, как их можно избежать. Советы преподавателю, организующему обучение студентов студентами, приведены, например в [16, 17]. Пример организации курса, обучающего написанию научных статей, с необходимыми материалами можно найти в [18].

Данная работа разбита на разделы, содержащие описание восьми прочитанных лекций с домашними заданиями (полный план работы на семестр оформлен в виде таблицы 1). Каждый раздел включает конспект прочитанной лекции, разбор предыдущего домашнего задания и план следующего. В материал лекций входят мотивация и рекомендации по выполнению каждого этапа исследования с указанием основных трудностей и возможных путей их преодоления. В разбор домашних заданий входит перечисление типичных ошибок, допускаемых студентами, и анализ работы группы в целом. В разделе “Результаты” описаны итоги курса, приводятся различные статистики по курсу, вычисленные на основе опросов студентов и консультантов, а также с помощью инструментов для анализа проектов, выполняемых в системах контроля версий.

Первая лекция (вводная)

Цель вводной лекции — дать студентам мотивацию для выполнения курса и четкого следования его структуре. Преподаватель начинает лекцию с перечисления карьерных возможностей студентов в научной среде и за ее пределами:

- управление/принятие решений,
- владение/анализ,
- наука,

и выделяет общие проблемы, с которыми студентам придется столкнуться независимо от того, по какому пути они пойдут. В частности, что все перечисленные виды деятельности характеризуются наличием потока информации, которым необходимо уметь управлять для успешного решения задач. Преподаватель вместе со студентами формализуют понятие задачи и инструментов передачи информации. Объявляется, что любая задача определяется набором “данные, модель, функционал качества”, среди инструментов перечислены учебник, статья, тезис, доклад, отчет. Для успешной передачи информации, студентам необходимо научиться излагать свои мысли и знания в одном из перечисленных форматов. Здесь преподаватель впервые говорит о целях курса: в рамках курса студент, под руководством преподавателя и старших студентов, проводит математическое исследование с элементами научной новизны и применяет результаты своей работы для решения практической задачи. Результаты исследования оформляются в виде научной статьи. Результатом курса в конце семестра становится подача статьи в научный журнал.

Таблица 1. Приложение: расписание курса на весенний семестр 2014 года.

Дата		ДЗ	Тема лекции	Результат для обсуждения	Код
Февраль	13	1	Вводная лекция.	Задано ДЗ-1	–
	20		Начало, демонстрация интерфейсов. Выбор задачи пробного программирования.	Регистрация в ML и SF. Установлены все необходимые инструменты, прочитаны вводные тексты.	–
Дата		ДЗ	Что делаем	Результат для обсуждения	Код
Март	27	2	Решить пробную задачу, написать код. Выбрать тему исследований.	Пробный код написан и загружен в репозиторий вместе с иллюстрирующими рисунками. Выбранная тема и ссылка на папку студента в SF помещены напротив фамилии нас странице группы.	Test
	6	3	Составить список публикаций по выбранной задаче. Написать аннотацию и введение с обзором выбранной литературы.	Аннотация (600 знаков), введение (1-2 страницы), список литературы в bib-файле.	Annotation, Introduction, Literature
	13	4	Поставить задачу и базовый вычислительный эксперимент. Провести первичный анализ работы алгоритма.	Постановка задачи (0,5-1 страница), код, отчет о работе базового вычислительного эксперимента	Statement, Basic code, Report
	20	5	Поставить вычислительный эксперимент на основе предлагаемого метода с учетом предыдущих результатов.	Код, визуализация полученных результатов, анализ ошибки, анализ качества.	Code, Visualisation
	27	6	Описание алгоритма.	Алгоритмическая часть статьи (второй/третий раздел).	Theory
Апрель	3	7	Доработка теоретической части и описание вычислительного эксперимента. Описание рисунков, выводы, заключение.	Черновой вариант статьи с разделами “Вычислительный эксперимент” и “Заключение”.	Document
	10	8	Завершение вычислительного эксперимента.	Описание эксперимента с анализом ошибок.	Error
	17	9	Показ статьи в целом.	Доработанная статья.	sHeck
	24	10	Доклады и обсуждение.	Статья подана в журнал	Final, Journal.

Затем преподаватель объясняет структуру курса. Преподаватель отвечает а наличие новизны в студенческой работе, а консультанты, имеющие опыт написания научных текстов, помогают студенту ясно и стилистически грамотно изложить свои мысли. Формат курса предполагает равномерную работу в течение семестра: вся работа разбита на небольшие задания с четким критерием оценки, каждое из которых выполняется в течение недели и оценивается студентом-консультантом. Пояним, в чем состоит система оценивания. Каждое задание обозначено буквой (например, Annotation = A, см. поле “код” в таблице 4). Консультант выставляет оценку задания по следующему приципу:

- «A»: результат работы соответствует требованиям,
- «A+»: задание выполнено лучше, чем ожидалось,
- «A++»: работы студента выше всяческих похвал,
- «A-»: работы выполнена неполностью, результат не соответствует требованиям,
- «A--»: Работа выполнена плохо и с задержкой.

Для более качественной обратной связи, рекомендуется выполнение работ за несколько дней до дедлайна, чтобы консультанты успели дать свои комментарии, а студенты — внести соответствующие правки. По возможности, рекомендуется организовать правки текста студента с целью улучшить стиль изложения таким образом, чтобы студент вносил правки самостоятельно. Возможно, при очной встрече или по скайпу.

Далее на лекции обсуждаются примеры прикладных задач, предлагаемых для иллюстрации математических методов, технические инструменты, используемые при работе над статьей и расписание курса. Для синхронизации работ используются следующие основные инструменты:

- Wiki-ресурс machinelearning.ru [3], посвященный алгоритмам и методам машинного обучения и анализа данных, а также деятельности кафедры. На странице группы [19] содержатся данные об успеваемости студентов по курсу, перечислены названия и краткие описания выбранных задач со ссылками на все сопутствующие материалы.
- Код, текст статьи и все основные файлы проекта хранятся в репозитории [4] проектов, выполняемых на кафедре. Синхронизация осуществляется с помощью системы контроля версий SVN Tortoise [5].

Домашнее задание-1, первая часть: инструменты.

- Поставить систему компьютерной алгебры Matlab (класс есть на факультете) или редуцированный заменитель Octave [20]; поставить систему верстки TeX: MikTeX [21] для Windows, TeX Live [22] для Linux и Mac OS; поставить библиографическую систему Jabref [23].
- Зарегистрироваться на сайте-репозитории SourceForge открытого кода SourceForge.net [24]. Прочитать вводные слайды о работе с системами контроля версий: [25]. Скачать программу-оболочку для обмена кодом TortoiseSVN [5] для Windows, RabbitVCS [26] для Linux.
- Прочитать статью [4] про SourceForge, загрузить MLAlgorithm [27], [28].
- Зарегистрироваться на сайте MachineLearning.ru [3].

Посмотреть все, что поставили, понять, как этим пользоваться на уровне интерфейсов.

Домашнее задание-1, вторая часть: LaTeX

- Прочитать статью про LaTeX [29].
- Прочитать основные главы ”Львовский С. М.” [30].
- Загрузить шаблон статьи [31] и скомпилировать.

- Прочитать про BibTeX [32]. Пример библиографической базы [33], пример библиографической записи [34], список баз и поисковых машин [35].

Домашнее задание-1, третья часть: Матлаб

- Прочитать статью “Введение в Матлаб” [36].
- Прочитать документирование функций Matlab, соглашение об именах переменных и создание отчетов о вычислительных экспериментах [37], [38].

Вторая лекция: план работы и основные инструменты

На второй лекции студенты более подробно знакомятся с форматом научной статьи. Обсуждается структура статьи, этапы подачи готовой работы в журнал и некоторые аспекты цикла рецензирования статьи. Для объяснения структуры статьи преподаватель показывает студентам несколько статей опубликованных в различных журналах и объясняет назначение разделов статьи и требования к ним. Студенты учатся выделять в статьях различного стиля единую структуру. Затем студентам предлагается ответить на вопрос: с чего нужно начинать работу? Звучит вариант «с постановки задачи». Но, так как описание задач, предлагаемых третьему курсу, уже содержит минимальную постановку, начинать нужно с поиска литературы. Студентам нужно объяснить, что все задачи уже кто-то решал, в такой же или похожей постановке. Поэтому разумным подходом к решению задачи будет изучить литературу по тематике задачи, выбрать понравившийся метод и попытаться его улучшить. Далее обсуждается порядок подачи и рецензирования статей. Внимание студентов обращается на то, что на первое впечатление рецензента о статье влияют в основном аннотация и вид список литературы. Вопрос о написании аннотации обсуждается в более поздних лекциях, когда студенты выберут задачу. На второй лекции внимание концентрируется на выборе литературы. По отзывам самих студентов, при работе над первой статьей большое затруднение вызывает подбор литературы. Списки литературы в студенческих работах зачастую плохо отражают состояние науки в исследуемой области и содержат либо недостаточно позиций, либо слишком старые или общие работы. Чтобы помочь студентам с поиском литературы, формулируются следующие требования к списку: Список должен содержать порядка 15 позиций, из них 1 — 3 позиции — это работы по узкой тематике, треть работ были опубликованы в течение последних 5 лет, 1 — 3 позиции — это фундаментальные работы обзорного плана. Общие рекомендации по поводу цитирования таковы:

- нужно цитировать того, кто поставил задачу,
- нужно цитировать тех, кто решал задачу по-другому,
- нужно цитировать тех, кто применял на практике похожие методы,

Оставшаяся часть лекции посвящается обзору инструментов: демонстрируются интерфейсы инструментов, с которыми студенты познакомились при выполнении ДЗ-1 и основные приемы работы. Озвучивается ДЗ-2: студентам предлагается список задач на применение простейших алгоритмов машинного обучения и анализа данных. Выполнение задачи может занимать от двух до шести часов. Задачи нужны, чтобы студенты могли почувствовать код, а преподаватель — оценить уровень подготовленности группы.

Домашнее задание-2: пробное программирование.

- Выбрать одну из задач из списка на странице группы [19], написать свою фамилию напротив задачи.
- В папке MLAlgorithms/Group174/Example2014Code создать папку Surmame2014Problem0 (фамилия студента, год, "Problem номер задачи).

- Сохранить туда код и графики (если задача предполагает).
- Графики оформлять в соответствии с рекомендациями [39] (достаточно файла Fig).
- При написании и документировании кода руководствоваться стилевыми рекомендациями (в частности, назвать первый файл main или demoTitle): документирование функций Matlab [?] и [38].

Третья лекция: выбор задачи

О стратегиях распределения задач. В этот день студентам предлагают определиться с темой проекта, над которым они будут работать в течение семестра, и перед преподавателем стоит нелегкая задача: организовать процесс выбора задач, чтобы каждый студент получил задачу по интересующей его тематике. Описываемый курс предлагается студентам третьего курса с 200? года, с небольшими модификациями. В большинстве случаев, темы были представлены в виде списка на странице, посвященной курсу, и запись на выполнение проекта проводилась в режиме он-лайн в течение недели: каждый студент имеет права на редактирование документа и просто вписывает свою фамилию напротив понравившейся ему свободной темы. Таким образом, студент выбирающий тему последним, имеет меньше всего выбора. Однако, такая системы не вызывала неудобств у студентов, так как большинство из них к третьему курсу еще не имели предпочтений и не могли адекватно судить о содержании проекта по названию и аннотации. Студенты третьего только начинают знакомиться с задачами машинного обучения, и большинство предлагаемых для исследования задач являются для них новыми с математической точки зрения. Сменить эту стратегию подтолкнуло изменение в активности студентов в последние годы. Если раньше на заполнение списка уходила практически вся неделя, то теперь список заполняется в течение первого дня. При этом некоторые студенты уже имеют определенные предпочтения и выражают недовольство тем, что все задачи по интересующей их тематике уже выбраны. Возникает желание назначить студентам темы, учитывая предпочтения всех студентов. Поэтому в этом году была реализована следующая схема: список названий проектов с краткой постановкой задачи, описанием данных и предполагаемой новизной был предоставлен студентам за день до лекции/обсуждения. На лекции каждому студенту был выдан список проектов, разделенный на три группы, и предложено в каждой группе задач выбрать по одной (с приоритетом или без). Перед раздачей списков, преподаватель еще раз напоминает студентам, что

- так как курс, кроме всего прочего, математический, не имеет большого значения, на каких данных решать задачу, и не иметь предпочтений по этому поводу — нормально
- в случае, если кто-то все же не останется доволен своей задачей, ее всегда можно обсудить, перепоставить или вообще сменить.

Таким образом удалось добиться покрытия почти всех задач, и каждый студент в итоге получил одну из выбранных им задач. Выяснилось, что наиболее привлекательно для студентов звучат задачи, которые предлагалось решить для анализа «человеческих» данных: социологические, электрокардиограммы, акселерометрические данные о движении человека и т.д.

На третьей лекции речь идет о написании введения и аннотации к статье, а так же о технических аспектах составления списка литературы. Лекция начинается с объявления домашнего задания: к следующему занятию студенты должны собрать литературу по выбранному проекту, написать аннотацию к работе и введение. Особенностям поиска литературы была посвящена предыдущая лекция. Теперь студентам даются рекомендации по обработке литературы и написанию на ее основе введения. Обсуждение начинается с

написания аннотации. Преподаватель обращает внимание студентов на небольшое противоречие: в материалах курса указано, что аннотация пишется в последний момент, когда работа уже завершена. Теперь же студентам предлагается начать написание статьи с написания аннотации. Это объясняется тем, что в аннотации студент пытается сформулировать суть работы, свою цель и возможно, желаемые результаты. Написание аннотации в начале работы помогает в понимании задачи и дальнейших шагов. Вариант плана аннотации:

- 1) чему посвящена работа в целом,
- 2) на чем сконцентрировано исследование,
- 3) в чем особенности исследования,
- 4) в чем новизна,
- 5) какими примерами проиллюстрирована работа?

Как и со структурой статьи в целом, для закрепления материала, студентам демонстрируют несколько вариантов аннотаций, разбирается их структура. Большинство рассмотренных аннотаций строго не следуют предлагаемому плану. Студентам так же предлагается трактовать его как набор советов, а не правил. План нужен только до тех пор, пока написание аннотации (как и любого другого раздела статьи) не станет для студентов естественным процессом. После этого они смогут реализовывать свои творческие идеи без ущерба для структуры и ясности изложения. Следующим шагом является введение. Такая последовательность вызвана скорее порядком их следования в шаблоне статьи, чем необходимостью вначале написать аннотацию и лишь потом введение. При написании первой статьи, студенты часто плохо различают назначение аннотации и введения. В результате, при правках статьи часть аннотации может быть перенесена во введение, и наоборот. Примерный план введения выглядит следующим образом:

- 1) основное сообщение — чему посвящена работа (одна-две фразы),
- 2) обзор литературы — развитие предлагаемой идеи (не более двух абзацев),
- 3) современное состояние области (два-четыре абзаца),
- 4) что предлагается (два абзаца).

При обсуждении введения у студентов возникают вопросы — как именно стоит отражать прочитанную литературу? В каком случае необходимо дать ссылку на работу? Студентам предлагается следующий вариант: при прочтении статьи или книги из списка литературы написать по ней краткий реферат, содержащий основную идею работы, новизну работы и связь работы с проектом студента: это может быть альтернативный метод к предлагаемому или метод, который студент собирается развить. (Студентам может быть неочевидна необходимость цитировать работы, в которых похожая задача решается другими методами. Необходимо пояснить, что подобные ссылки показывают, что автор статьи изучил состояние науки в исследуемой области, и являются в некотором смысле гарантом новизны предлагаемого метода.) Из написанных таким образом рефератов студент может составить части обзор литературы. Посмотреть примеры введения студенты могут не только в работах своих предшественников, но и в статьях, которые они читают при подготовке литературы.

Домашнее задание-А11: выбор задачи, изучение литературы, написание аннотации и введения.

- 1) Выбрать задание, записать название, автора и рецензента (с его согласия) в таблицу [19].
- 2) Найти базовую литературу, занести ее в BibTeX. Проверить правильность заполнения базы BibTeX (порядок написания имен авторов, тома журналов, номера страниц).

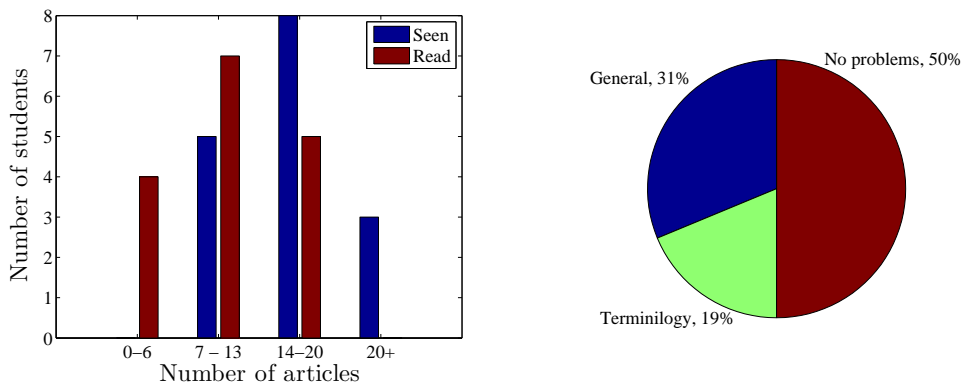


Рис. 1. Результаты поиска литературы.

- 3) Написать аннотацию, поместить ее в шаблон статьи [31].
- 4) Создать папку с названием проекта Surname2014Title в папку группы MAlgorithms/Group174/. Создать там необходимые папки (см. раздел “Работа с репозиторием”), поместить в шаблон-заготовку статьи.
- 5) Просмотреть/прочитать литературу, написать краткие рефераты и вставить их во введение.
- 6) Написать введение.

Таким образом, домашнее задание студентов — понять задачу, прочитать литературу и отразить результаты этой деятельности в аннотации, введении и библиографической базе (списке литературы).

Четвертая лекция: проверка аннотации, введения и списка литературы

С выполнения работ АИЛ начинается взаимодействие студента и консультанта. Студенту необходимо научиться различать вопросы с которыми он может справиться самостоятельно, и вопросы по которым лучше обратиться за помощью, чтобы работа выполнялась без задержек и последовательно развивалась. В процессе первого этапа работы у студентов возникает множество вопросов о координации работы консультантом. Некоторым остается неясным, какие этапы работы фиксировать и предоставлять на проверку консультанту; некоторые не смогли за неделю наладить контакт с консультантом. В некоторых случаях, по договоренности с преподавателем, студент и консультант переносили дедлайн в связи с несовпадением графиков студента и консультанта и т.д.

На четвертой лекции обсуждаются первые результаты совместной работы студентов и их консультантов — список литературы, аннотация к работе и введение. Некоторые результаты опроса студентов об итогах поиска литературы изображены на рисунке 1. Гистограмма на рисунке 1, а, демонстрирует распределение количества статей, просмотренных студентами при подготовке домашнего задания (синим цветом) и распределение количества прочитанных статей. Круговая диаграмма 1, б, отражает количество студентов

- испытывающих затруднения при чтении англоязычной литературы (сектор “General”),
- испытывающих затруднения при чтении статей, посвященных незнакомой тематике (сектор “Terminology”),
- свободно читающих литературу на английском языке (сектор “No problems”).

Преподаватель случайным образом выбирает и зачитывал написанные студентами аннотации и введения, отмечая распространенные ошибки и предлагая варианты их исправления, например:

- не использовать в аннотации и ключевых словах терминов на английском языке (кроме как в том случае, когда общепринятого варианта на русском еще не существует) и аббревиатур. Английский вариант предлагается указать при появлении/определении термина во введении, и затем употреблять только русский вариант. Это же применимо и для аббревиатур. Лучше не вводить в аннотацию понятий, которые не удастся тут же пояснить.
- не стремиться к использованию синонимов в тексте. В математической статье попытка называть одну сущность разными именами может привести к затруднению в понимании. Если одно и то же понятие встречается в литературе под различными названиями, предлагается зафиксировать один вариант при первом появлении термина (с возможным указанием других вариантов и ссылкой на них) и далее в тексте использовать только его.
- избегать общих и непроверяемых утверждений, таких как “один из наиболее эффективных методов ...”, и конструкций со словами “часто, нередко, можно”. Каждое утверждение вида “для решения задачи применяется”, “широко используется” должно быть подкреплено ссылками на соответствующие работы.
- описание используемых методов и обзор литературы желательно построить таким образом, чтобы читателю как можно реже приходилось обращаться к дополнительным источникам в ключевые моменты статьи. В частности, при упоминании во введении ключевого метода, следует кратко описать его.
- избегать чересчур сложных конструкций; по-возможности, стараться избегать пассивного залога. Слишком длинные предложения разбивать на два или более предложений.

Далее преподаватель демонстрирует примеры постановок задачи — с явной постановкой задачи оптимизации функционала качества в виде $\arg \min$ или с описанием оптимизируемой величины. Выделить единый план постановки задачи сложно, так как его форма зависит от решаемой задачи. В той или иной последовательности, постановка задачи может содержать:

- описание выборки (и, возможно, допустимые операции над элементами выборки),
- статистические или иные предположения о характере выборки,
- описание (и, возможно, обоснование) функции ошибки, функции потерь или иной функции качества, согласно которой будет оцениваться качество решения задачи,
- ограничения, накладываемые на решения задачи, способы разбиения выборки,
- возможно, дополнительные функции качества,

Домашнее задание-SBR: постановка задачи, запуск базового алгоритма, отчет о запуске

- 1) Согласовать с консультантом постановку задачи. Посмотреть примеры постановок в предыдущих статьях. Написать постановку (0.5 — 1 страница, см. pdf и tex-файлы со списком рекомендуемых обозначений в архиве [40]).
- 2) Поставить базовый вычислительный эксперимент. Для этого вместе с консультантом найти наиболее простой алгоритм, решающий поставленную задачу (возможно, в частичной постановке).
- 3) Подготовить синтетические данные (или взять несложные реальные, небольшую выборку).

- 4) Запустить алгоритм на синтетических данных, вычислить ошибку. Показать на элементарном графике, что получен адекватный результат.
- 5) Написать мини-отчет о полученных результатах, полстраницы. Он включает:
 - Краткое описание цели эксперимента.
 - Краткое описание способа порождения выборки или краткое описание природы реальных данных (см. например, [41]).
 - Краткое описание графика и комментариев полученных результатов.
- 6) Постановку задачи и отчет поместить в текст статьи. Данные и код — в папки data, code.

Пятая лекция: постановка задачи

Перед началом лекции преподаватель просматривает оценки, выставленные консультантами. Важное значение наличия системы оценивания для преподавателя состоит в том, что по таблице можно понять, какой паре студент-консультант не удалось наладить общение. В предположении, что целью всех студентов, записавшихся на курс, является написание качественной статьи, отсутствие нескольких оценок подряд у студента означает, что студенту непонятна какая-то часть работы, либо у него возникли трудности с выполнением этапа работы, разобраться с которыми самостоятельно студент не смог. Если при этом общение с консультантом не налажено, студент не справляется с очередным этапом работы, что, в свою очередь мешает ему продолжать работать дальше. Вот почему важно еще как можно скорее выявить такие пары и принять меры по решению этой проблемы. В начале пятой лекции студенты получают развернутые анкеты, целью которых является выяснить,

- 1) насколько понятна студенту его задача и дальнейшие действия,
- 2) как построена работа с консультантом,
- 3) какие трудности возникли при выполнении работы,
- 4) сколько часов ушло на этап работы.

Большинство опрошенных (9 из 12) утверждают, что цель задачи и дальнейшие шаги (8 из 12) им понятны. Проблемы, возникшие у одного или более студентов:

- очень много материала для изучения;
- организация работы консультантом: общение происходит только перед делайном;
- нет реальных данных для постановки эксперимента;
- 3 человека не смогли найти реализацию базового алгоритма;
- неясна постановка задачи или какая-то ее часть (2 человека);
- интерпретация результатов.

Затем начинается просмотр студенческих работ. Постараемся выделить основные моменты, требующие внимания:

- первое, что можно отметить, это отсутствие мотивации в начале раздела “Постановка задачи”. Несмотря на то что студенты уже приводили мотивацию своей работы в разделе “Введение” или в аннотации, мотивация необходима во всех разделах и даже чаще. В аннотации и введении приводилась мотивация всей работы в целом; в начале каждого раздела необходимо кратко пояснить назначение этого раздела. Более того, при написании алгоритмической части или при введении новых понятий и обозначений тоже желательно объяснить читателю, о чем он сейчас будет читать, для чего будут введены новые обозначения или рассмотрена некоторая формула.
- Все формулы желательно так же пояснять словами. Большие или имеющие ключевое значение формулы обязательно выключать из текста. Необходимо помнить, что

формулы — это часть текста, и правила пунктуации распространяются на них так же как и на остальной текст. Если предложение заканчивается формулой, в конце нужно ставить точку; не забывать запятые, если они стоят после формул.

- желательно заранее продумать непротиворечивую систему обозначений, чтобы избежать дублирования и двусмысленности. Например, сохранять перечислительный индекс за одной сущностью, которую он перечисляет. (использовать индекс i только для нумерации объектов, а j — только для признаков). Не рекомендуется вводить обозначения, которые далее в работе использоваться не будут.

Помимо постановки задачи, студенты также выполнили небольшой вычислительный эксперимент, отразили результаты на графиках и написали отчет, состоящий из описания эксперимента и полученных рисунков. Общее замечание по поводу отчетов о выполнении вычислительного эксперимента таково: при описании рисунка необходимо указать, что отложено по осям графика, а затем словами объяснить что именно читатель видит на графике, даже если автору это кажется очевидным. Подробнее этот вопрос будет рассмотрен при обсуждении раздела. Также рекомендуется сохранять цветовые обозначения: если синим обозначена кривая для для ошибки предлагаемого метода, а красным — улучшаемого, это обозначение должно соблюдаться на всех рисунках.

У многих студентов на данном этапе возникает вопрос: а что еще осталось сделать? Задача уже поставлена, эксперимент проведен и у многих дает неплохие результаты, работа практически закончена. Ответ прост: далее работа идет по кругу, с улучшением уже полученных результатов и их описания. Следующим домашним заданием становится постановка полного вычислительного эксперимента: проверка предлагаемых методов, сравнение с базовыми и описание результатов. Рекомендации по описанию результатов уже были перечислены при обсуждении задания SBR.

Домашнее задание-CV: планирование и начальная постановка эксперимента, визуализация.

- 1) Подготовить данные, объявленные в проекте. Загрузить их в репозиторий (в случае, если объем файла превосходит 5 МБ, или если файлов много, согласовывать это с консультантами).
- 2) Описать цели вычислительного эксперимента, составить план эксперимента, описать данные.
- 3) Описать дополнительные критерии качества модели и ограничения, если таковые имеются.
- 4) Описать анализ ошибок, перечислить сравниваемые методы.
- 5) Составить список графиков для визуализации результатов, описать графики.
- 6) Выполнить вычислительный эксперимент в начальном варианте (на синтетических, и если возможно, реальных данных) и нарисовать графики.
- 7) Подготовить устный доклад на две минуты. Доклад отрепетировать по секундомеру. План доклада Middle:
 1. Название работы, о чем работа в целом.
 2. В чем особенность работы.
 3. Что сделано.
 4. Какие результаты получены (если есть график, заранее попросить ведущего показать его на экране).
 5. Что предполагается сделать.

Лекция 6: доклады о проведении вычислительного эксперимента

По результатам проведенных экспериментов студенты делают короткие доклады, не дольше двух минут. За это время студент должен успеть кратко сформулировать задачу, объяснить аудитории ее важность и новизну предлагаемых методов, а также описать полученные результаты и сделать выводы о протестированном методе. Это нелегкая задача для студентов третьего курса, но она помогает четче увидеть цель работы, выделить в ней главное и второстепенное, заставляет анализировать и интерпретировать результаты. Кроме, студент получает опыт выступления с докладом о научной работе.

Домашнее задание-Т: теоретическая часть

- 1) Написать решение поставленной задачи:
 - (а) в простом варианте
 - (б) в деталях, насколько это возможно.
- 2) Сравнить написанное решение с кодом, откорректировать код и решение.
- 3) По возможности, заранее (например, в выходные) связаться с консультантами или с преподавателем для проверки постановки задачи и описания ее решения.

Лекция 7: разбор структуры статьи

Согласно плану работы в семестре, к седьмой лекции студентами уже написаны основные разделы статьи: «Введение», «Постановка задачи», «Описание алгоритма» и «Вычислительный эксперимент». Теперь, когда почти весь материал уже изложен, необходимо обратить внимание на структуру статьи. При написании отдельных разделов сложно отслеживать связность изложения всей работы и к данному этапу работы у многих студентов возникает небольшая путаница в статье: некоторые информационные блоки помещены в несоответствующие им разделы. В таких случаях информация подается читателю либо слишком рано, когда он еще не готов ее воспринимать, либо слишком поздно, тогда она менее ценна для читателя. Такие перестановки затрудняют понимание и снижают вероятность, что статья будет прочитана до конца. Чаще всего смешение происходит в разделах «Постановка задачи», «Описание алгоритма» и «Вычислительный эксперимент». Несмотря на то что план этих разделов был детально разобран, нелишне будет напомнить следующие соображения, которыми следует руководствоваться при их написании:

- В постановку должны входить только те обозначения и формулы, которые нужны чтобы сформулировать задачу (записать в явном виде оптимизируемый функционал) и ограничения на ее решения.
- Описание действий, направленных на решение поставленной задачи, приводится в разделе «Описание алгоритма». При этом базовым алгоритмам, следует уделять не более половины страницы. Достаточно описать суть базового алгоритма и дать ссылку на него.
- Более подробная информация о применении предлагаемого метода в работе (количество итераций алгоритма, значения констант, размер данных и т.д.) дается в разделе «Вычислительный эксперимент».

Предлагается такой вариант оценки ясности работы: один из студентов, присутствующих на занятии, должен быстро просмотреть статью своего товарища и рассказать остальным студентам, о чем написано в работе. Как правило, студенты определяют тип решаемой задачи, но затрудняются выделить предлагаемый подход и особенность решаемой задачи. С данным подходом связан студенческий вопрос: «Хорошо ли то, что мы учимся писать

статьи, предполагая что читать их будут невнимательно?” Постараемся ответить на него следующим образом. Статью не будут читать, если она будет написана так, что читать ее трудно и при этом непонятно, нужно ли вообще ее читать. Читатель выбирает статьи из огромного списка работ, представленных научным сообществом. (Студенты наверняка сами помнят, что при выборе литературы им пришлось столкнуться с большим количеством статей на пока еще незнакомую тему.) Наша задача состоит в том чтобы написать статью, открыв которую, читатель сразу мог бы понять, зачем ему читать именно эту статью, и при этом мог бы прочитать и понять ее с минимальными усилиями.

Домашнее задание-D: документ в целом. Подготовить черновик статьи, содержащий все необходимые разделы:

- 1) Название
- 2) Аннотация (пишется в первую очередь и переписываются при завершении работ): изложение краткого содержания и основных результатов (не более 600 знаков)
- 3) Ключевые слова (основные термины; можно использовать те, которые дали хорошие результаты поиска)
- 4) Введение (около страницы); ниже — по абзацам, примерный план)
- 5) Постановка задачи (примерно страница)
- 6) Дано (как устроена выборка)
- 7) Решение: математическая часть (тут название первого раздела)
 - Описание алгоритма
 - Исследуемые свойства алгоритма
- 8) Другие разделы (если есть)
- 9) Решение: алгоритмическая часть (часто уходит в следующий раздел)
- 10) Вычислительный эксперимент
- 11) Заключение (пишется в последнюю очередь): сжатое изложение результатов (1/4 страницы). Желательно вставить ссылку на папку проекта в mlalgorithms для того, чтобы другие исследователи могли проверить результаты или использовать их в дальнейшей работе.
- 12) Литература: опорные статьи за последние 10 лет, максимально покрывающие тематику (не менее 20 статей).

Лекция 8

Предполагается, что к этому времени уже написано заключение, то есть студенты понимают, какие результаты осталось получить или эксперименты провести до окончания работы, и могут описать их, даже если эти результаты еще не были получены (или эксперименты не проведены). К следующему занятию планируется завершение работ над вычислительным экспериментом, поэтому на восьмой лекции обсуждаются способы анализа ошибки и визуализации полученных результатов. Несмотря на то что студенты посещают также курс машинного обучения, анализ данных является для них новой областью знаний, и многие затрудняются интерпретировать полученные результаты. Поэтому восьмая лекция посвящена математическим аспектам описания результатов. Перечислены различные способы введения функции ошибки и продемонстрирована связь между ними. Рассмотрены способы разбиения выборки на контрольную и обучающую, перечислены различные способы анализа качества работы алгоритма (анализ дисперсии ошибки, распределения параметров модели) и варианты дополнительных функционалов качества.

Таблица 2. Учет времени и оценки лучших студенческих работ, 2014 год.

Author	[BMF]LSICUDTPRWS	Score	Grade	Total hours	0	L	S	I	C	UD	TP	R	W
Vasiliy	[BM+F]L+SI+CU-DTPRWS	14.5	10	67		6	8	8		16	16	13	
Evgeniya	[BF]L-S+I+C0DT-0R-0S	9.75	5	25		3				20		2	
Sergey V.	[BM+F]LS-I+CU+ DTP+R-W+S-	14.25	10	53.25		8	4.50	7	8	17	5		3.75
Oleg G.	[BMF]L-SI-C-0DTPRWS	12.25	8	23.75		0.25				9.5	14		
Anna	[M]L0I->>>000C-	2.5		5					0	0	5		
Irina	[BM+F]LS->>>>>00IC-U- D-TP-R-W-S-	11.75	7	16.1	0.1	4	5						7
Vladimir	[B]L-0I-C-0D-00000	3		25	1	7	8					4	5
leksandr I.	[BM+F]LSIC->>>U>DTPR0S-	12.75	8	28	1	3	8						16
Sergey K.	[BF]L-S-I->>>000C- U-DT-P-R-W-S-	9.75	5	33.5	1	2.5						6.5	23.5
Aleksandr K.	[BM+F]L+SI+CU DTPR+W+S	15.25	10	48		3.5	5	4	16	6	6	3.5	4
Aleksandr Kos.	[BF]L+S-IC-	5.75	1	61		16					35	10	
Ekaterina	[BMF]-L-S-IC- UD-000W-S-	8.5	4	12.25	0.25		12						
Aleksandra	[BMF]-L-S-IC- UD-000W-S-	8.5	4	49.17	0.17	3	15		8			20	3
Pavel	[BM+]L-SIC-U- D->PR-W>	9.75	5	32.5		4	4	2.5	3	15			4
Mikhail M.	[BF]L-SIC- UDTPRW+S	12.75	8	145		6	8	5	16		100		10
Andrey M.	[M+F]L+SI-C- UDT-P00S-	10	5	07.5						7.5			
Kirill	[BM+F]LS-I+C U-DTPR-WS-	13.5	9	40						24	16		
Dmitriy	[BM+F]L-SI-CU-DTPRW+S	13.75	9	11.5		0.5	1				3.5	6.5	
Roman	[BF]L+0I->>>C- 0D-00R-W-S-	7.25	3	14.5		6						5.5	3
Aleksey	[BM+F]L+S+I+C+ UDT+P+R+W+S	16.25	10	35.25		8		6	3		16		2.25
Mariya	[BM+F]L-S+I+CU DT+PR+W+S	15.25	10	66		3	3		16	16	18		10
Sergey S.	[BF]L+S+I+CU DT-0R-WS	12.25	8	12		5	3	4					
Oleg U.	[BM+F]L-S-IC> D»R-WS	10.25	6	82.5	5	3	32	7	5.5	9	3	7	11
Rustem	[BM+F]-L-S-IC- U-D-TPRWS-	12.5	8	17	0.5	3	4		4			5.5	
Nikolay	[B]L-S-IC-0D- T-0R-W-S-	7.25	3	20	1.5	1.5		8			9		
Daniil	[M+F]L-S-IC-> UDTPRW-S-	12.75	8	40						10	30		
Task average (median)				30.25	1	3.5	8	6	6.75	12.5	15	6.5	5

Таблица 3. Учет времени консультантов, 2013 год.

Consultants	Total hours	0	L	S	I	C	U+D	T+P	R	W	S	??
Mikhail K.	140	2	8	10	10	16	16	24	6	8	10	30
Nikita I.	160	12	10	10	10	10	30	16	8	12	12	30
Nikita Z.		2	10	10	8	8	10	10	4	6		

Таблица 4. Учет времени и оценки лучших студенческих работ, 2014 год.

Студент	AIL	SBR	CV	T	D	E	H	Total	Оценки по заданиям	Σ
Ирина	20	43	24	5	7	10	7	116	[MF]T+A+I+L+SB ++R+CV+TDE+H	17,25
Андрей И.	8	N/A	10	N/A	8	6	6	38	[MF]TA+I+L+S+B +R+C+VTD>E+HS	17
Михаил Ш.	8	14	23	14	14	23	30	126	[MF]T+AIL+S+B R+CV+T+D+E+H	17
Михаил Шв.	8	N/A	5	N/A	9	N/A	N/A	22	[M+F]T+A+I+L+ S+B+R+CVTD+E	16,25
Римма	8	14	17	10	7	23	N/A	79	[MF]TAI+L+SBR+ CV+T>DEH	16
Мария	8	20	6		8	7	7	56	[MF]T+AI+L+++S B++R+CV+TD+	15,25
Михаил К.	25	10	18	N/A	N/A	N/A	N/A	53	[MF]TAI+L+SB RC+V+TDES	15

A — Annotation, I — Introduction, L — Literature, S — Statement, B — Basic code, R — Report, C — Code, V — Visualization, T — Theory, D — Document, E — Error, H — сCheck
M — Middle report, F — Final report.

Структура презентации, которую студентам предстоит подготовить для финального отчета, практически повторяет структуру статьи. Различия при оформлении вызваны форматом презентации (сравнительно небольшое пространство слайда, невозможность слушателю в любой момент вернуться к произвольному слайду). Желательно при подготовке слайдов соблюдать следующие требования:

- Проследить чтобы на слайде отображался номер текущего слайда и общее количество слайдов в презентации, либо предоставить слушателю другую наглядную оценку объема презентации и расположения текущего слайда в нем. Это особенно важно для долгих докладов, так помогает слушателям сохранять концентрацию.
- Формулы на слайдах не нумеруются. При необходимости обратиться к ранее введенной формуле, лучше ее задублировать или отлистать презентацию до слайда, на котором формула была введена.
- Убрать окружения рисунков и таблиц. В таблице жирно выделить главный результат. Например, если приводится сравнение нескольких методов, выделить предлагаемый.
- Не нужно читать текст на слайде, слушатели могут сделать это самостоятельно. Лучше пересказать содержание другими словами.
- Сообщение слайда должно быть продублировано заголовком. Точки в заголовках не ставятся.

Домашнее задание-Е: анализ ошибки. Дополнить описание вычислительного эксперимента, используя один или несколько предложенных способов анализа ошибки. Подготовить презентацию статьи, следуя плану:

- 1) Аннотация: изложение краткого содержания и основных результатов (не более 600 знаков).
- 2) Введение: раскрытие темы статьи, общая постановка задачи, обзор литературы, описание подхода к решению задачи.
- 3) Постановка задачи: полная формальная постановка, введение обозначений, принятие необходимых гипотез, задание функционалов качества.
- 4) Описание алгоритма (возможно несколько разделов): математическое описание предлагаемого алгоритма, исследование его свойств, доказательство необходимых теорем.
- 5) Вычислительный эксперимент: описание исходных и производных данных, описание технической части алгоритма (если необходимо), описание результатов, сравнение их с результатами других алгоритмов; крайне желательны иллюстрации.
- 6) Заключение: сжатое изложение результатов (1/4 страницы).
- 7) Список литературы: желательно найти опорные статьи за последние 10 лет, максимально покрывающие тематику.

См. рекомендации к написанию отчётов и статей [42].

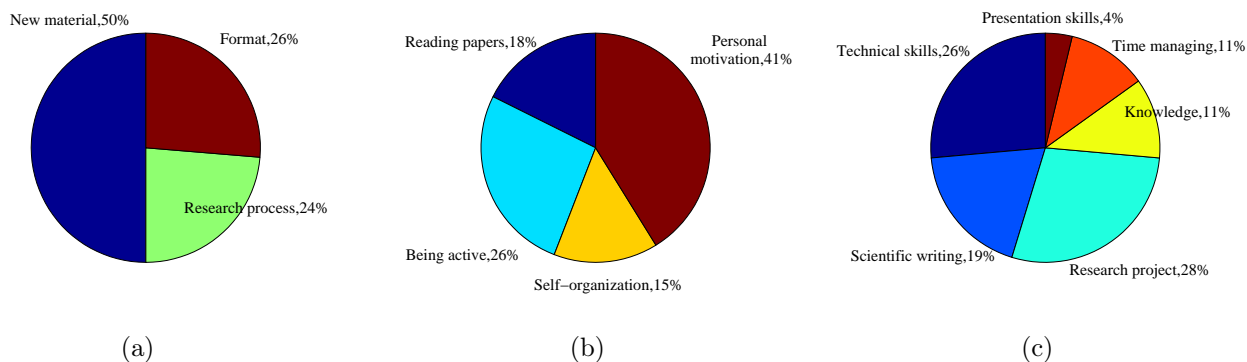


Рис. 2. Диаграммы отражают студенческие отзывы о курсе: (а) что студентам понравилось в курсе, (б) что помогло при выполнении проекта, (в) какие навыки удалось приобрести.

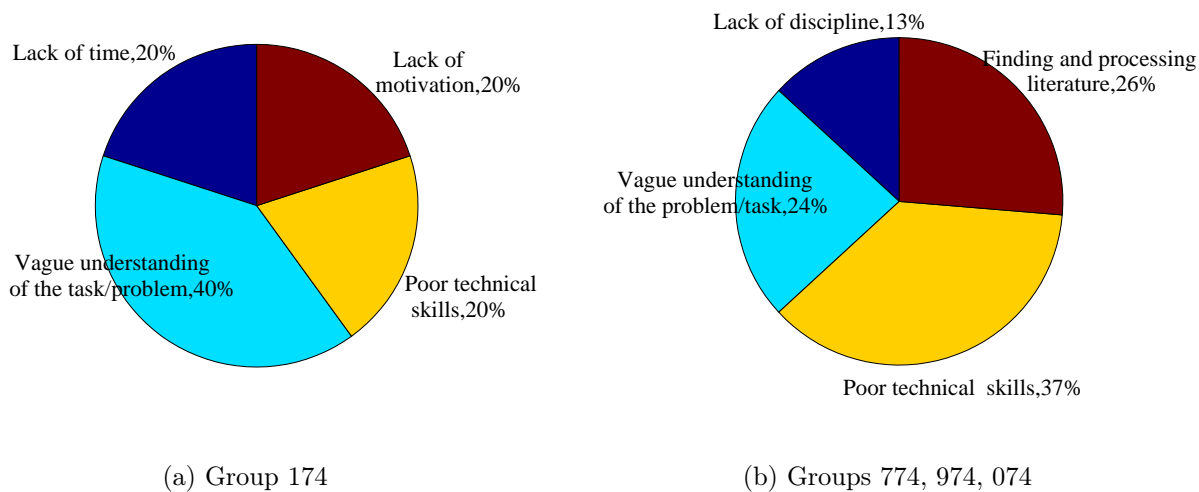


Рис. 3. Основные сложности, отмеченные студентами: (а) третьего курса, (б) старших курсов.

Результаты

В результате работы над проектом 14 из 16 студентов завершили работу, удовлетворяющую критериям положительной оценки, из них 9 статей было подготовлено к публикации в журналы. Статистика по некоторым студенческим работам, заслужившим отметку “отлично” приведена в таблице 4. Работы отранжированы по суммарной оценке выставленной консультантом. Каждая буква, проставленная за очередной этап работы, оценивалась в один бал; каждый плюс и минус добавлял или отнимал соответственно пол бала. Отметку “отлично” выставлялась при наличии суммарного бала выше 14. Также в таблице приведены оценки времени затраченного, на каждое из заданий, полученные из еженедельного анкетирования студентов: на каждой лекции всем присутствующим студентам предлагалось ответить на вопросы о процессе работы над статьей, выразить свое мнение о различных аспектах решаемой задачи и курса в целом. На основе отзывов студентов выделен набор утверждений о курсе. Диаграммы отражают долю опрошенных студентов, согласившихся с данным утверждением. На рисунках 2, 3 приведены диаграммы для ответов на вопросы “Что понравилось в курсе?”, “Приобретенные навыки”, “Что помогало в курсе (советы другим студентам)”, “Основные сложности”. Варианты ответов для первых трех вопросов были примерно одинаковы при опросе студентов третьего курса и студентов старших курсов, прошедших уже два или три семестровых этапа предлагаемого курса, поэтому для этих вопросов приводятся агрегированные результаты. Основные сложности, упоминаемые студентам различались существенно, поэтому мы приводим диаграммы отдельно для третьего курса и агрегированную диаграмму для старших студентов. Далее приводятся пояснения к значениям меток секторов диаграмм.

Что понравилось в курсе (рис. 2, а):

- *Research process*: размышлять над задачей, придумывать новые методы решения. Возможность применить на практике полученные знания, получить опыт работы с реальными данными.
- *Format*: формат курса, сочетание дисциплины и творчества. Частота общения с преподавателям/консультантом.
- *New material*: широкий спектр получаемых знаний и навыков.

Что помогало в работе (рис. 2, б):

- *Self-organization*: работать равномерно, выполнять задания к сроку.
- *Being active*: проявлять активность в общении с преподавателями и консультантами, задавать вопросы с самого начала работы.
- *Reading papers*: изучить как можно больше литературы в области решаемой задачи.

Приобретенные навыки и знания (рис. 2, в):

- *Research project*: опыт работы с реальными данными, постановки и проведения экспериментов; умение работать с литературой.
- *Technical skills*: технические навыки.
- *Scientific writing*: опыт написания научных текстов.
- *Knowledge*: более широкий кругозор в области Machine Learning.
- *Presentation skills*: Опыт выступления с докладами и презентациями.
- *Time management*: умение планировать свое время.

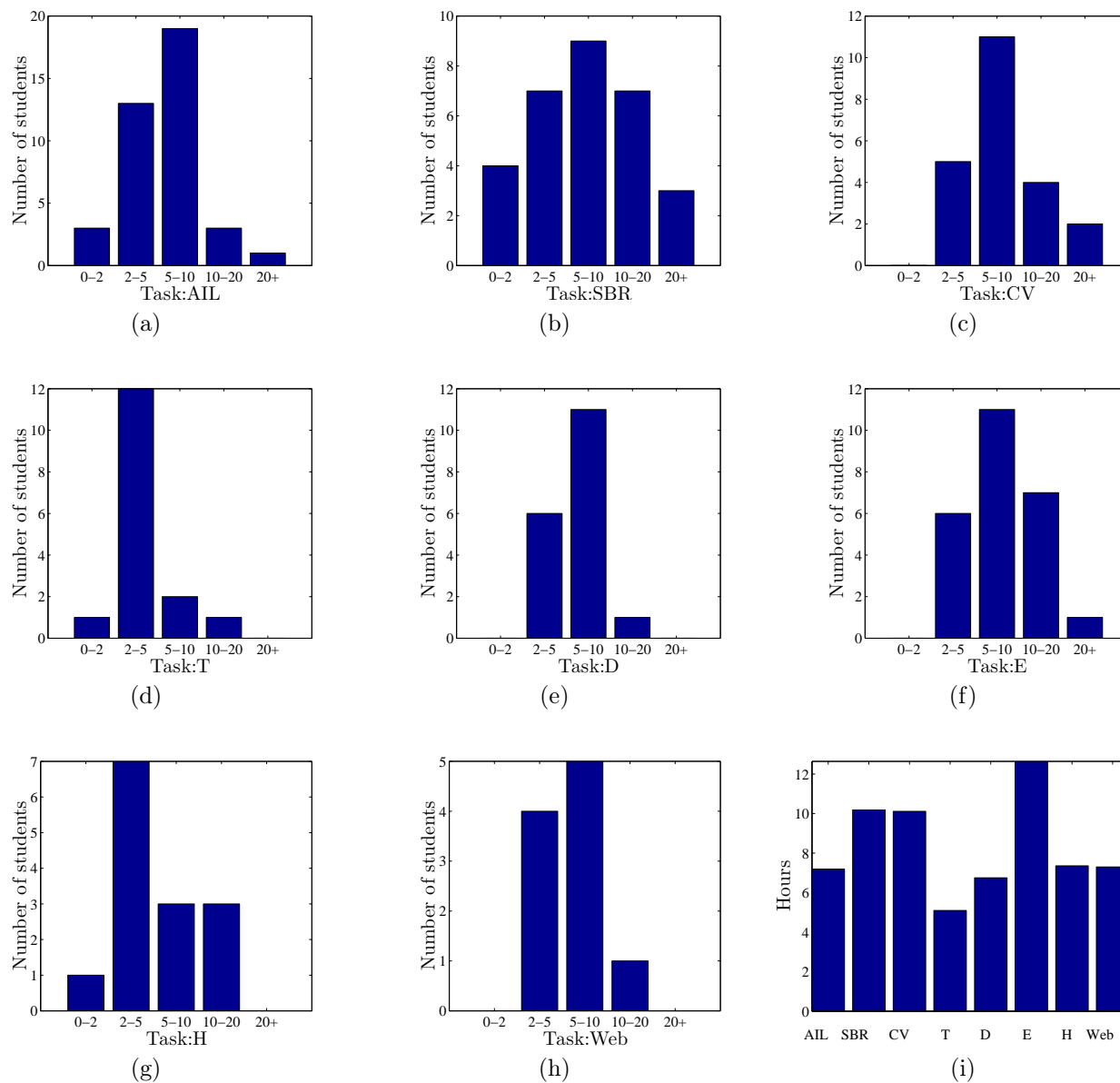


Рис. 4. (a-h): Распределение времени, затраченного студентами на каждое задание (время в часах). Задание “Web” студентами 174 группы не выполнялось. (i): среднее время, затрачиваемое студентами на каждое задание.

Основные сложности, отмеченные студентами третьего курса (рис. 3, а):

- *Poor technical skills*: отсутствие необходимых технических навыков.
- *Lack of motivation*: некоторые студенты отмечали, что мотивация к очередным этапам работы не всегда была достаточна.
- *Vague understanding of the task*: непонимание задачи или задания.
- *Vague understanding of the task*: непонимание задачи или задания.

Основные сложности, отмеченные студентами четвертых-шестых курсов (рис. 3, б).

- *Poor technical skills*: отсутствие необходимых технических навыков.
- *Lack of discipline*: многие студенты отмечали, что задержки заданий и отставание от графика существенно затрудняет выполнение последующих заданий.
- *Vague understanding of the task*: непонимание задачи или задания.
- *Material comprehension*: трудности с освоением материала.
- *Finding or processing literature*: проблемы с поиском необходимой литературы или обработкой имеющейся литературы.

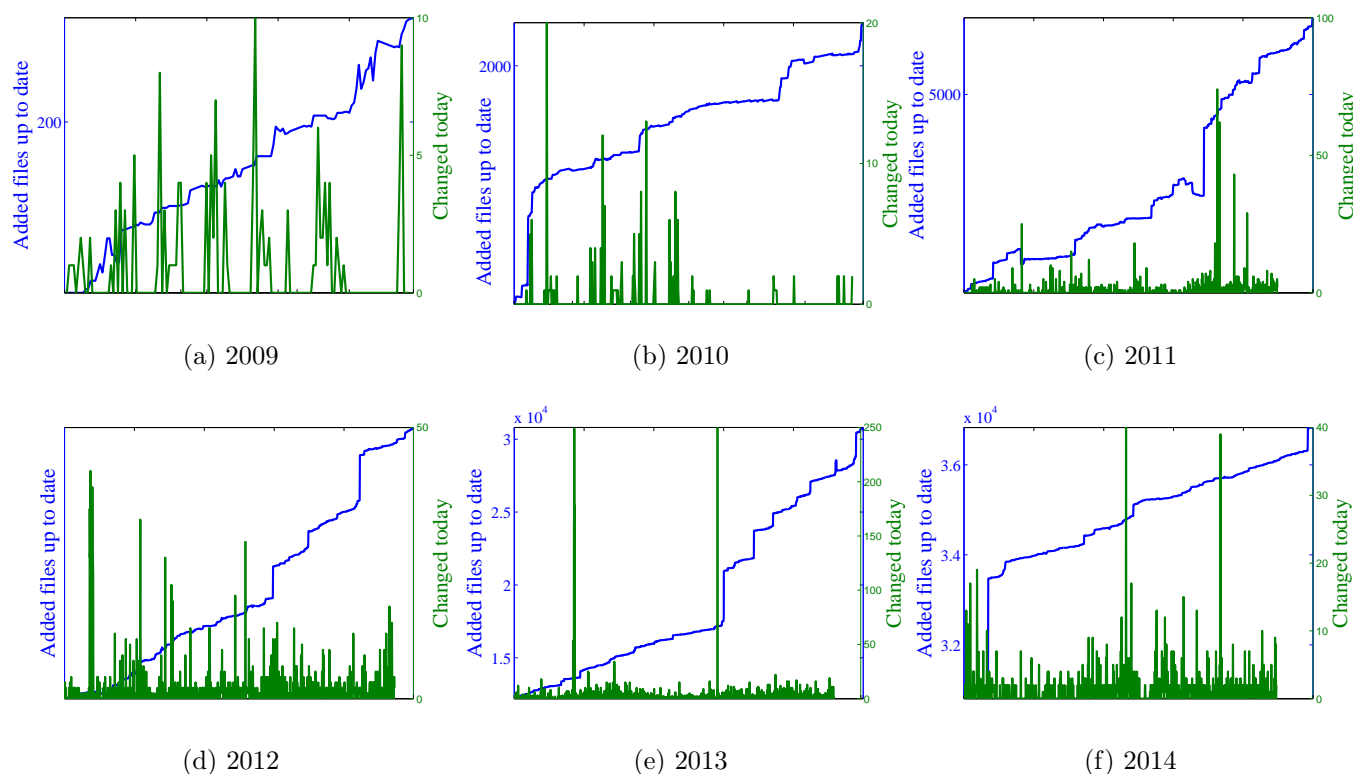


Рис. 5. Динамика добавления и изменения файлов в репозитории за 2009–2014 года.

Отзывы консультантов. Консультанты также были проанкетированы по завершению курса. Целью анкетирования было выяснить, сколько времени ушло на консультирование студентов, как часто проходило общение, кто был его инициатором и в чем фактически заключалась помощь консультанта. Ответы консультантов были примерно одинаковы; ниже приводится список вопросов и типичные ответы консультантов.

Вопрос**Усредненный ответ**

- 1) Сколько времени в целом ушло на консультирование?
 В среднем — по три часа на общение со студентом в начале курса и около получаса — в конце. (Для сравнения, в 2013 году, когда на все группу было выделено 3 консультанта, на проверку задания у каждого уходило от 2 до 30 часов 3.)
- 2) Как часто поступали запросы от студента?
 Активная работа шла практически со всеми студентами до последних двух-трех этапов сдачи. В этот период студенты обращались к консультантам с вопросами три-четыре раза в неделю. В конце семестра частота обращений снижалась до одного раза в неделю (перед дедлайном).
- 3) Какой характер носили запросы?
 Вначале студенты чаще задавали вопросы на понимание задачи и ближайших целей, затем было больше технических вопросов, связанных с реализацией алгоритмов, оформлением рисунков и текста статьи.
- 4) Приходилось ли Вам самим запрашивать у студента материал?
 (Вопрос направлен на оценку среднего по группе характера взаимодействия между студентом и консультантом). Примеры ответов:
 — Ни разу ничего не запрашивала. Переносы были только по уважительным причинам (болезнь).
 — Да, в один момент, когда у студента были проблемы по учебе.
 — Нет. Задания сдавались в срок, либо была договоренность о задержке.
 — Нет, поскольку считаю, что инициатором беседы с консультантом должен выступать студент в силу большей мотивированности.
- 5) В чем, по Вашему мнению, заключается Ваш вклад в работу?
 Типичные варианты ответа:
 — помощь по преодолению технических проблем,
 — правки стиля и оформления,
 — рекомендации по коду, по схеме проведения эксперимента,
 — рекомендации по представлению результатов.
- 6) Хотели бы Вы продолжить консультировать студентов в рамках курса?
 Все консультанты выразили заинтересованность в повторении опыта в следующем году.

Учет затраченного времени. При составлении расписания на очередной семестр были учтены пожелания и замечания студентов, собранные за предыдущие года, а также студенческие оценки времени, затрачиваемого на выполнение заданий. Этим вызвано более плотное наполнение домашних заданий в начале курса — AIL, SBR и CV — по сравнению с более поздними заданиями. В частности, по отзывам студентов, в предыдущие годы им не хватало времени на отладку предложенных методов после получения и анализа первичных результатов. Поэтому в расписании курса на весеннем семестре 2014 года первичный вычислительный эксперимент (задание В) запланирован уже на четвертое ДЗ и входит во второе домашнее задание, посвященное написанию статьи.

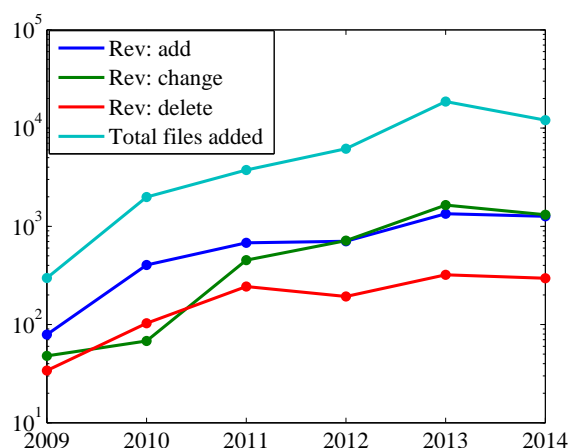


Рис. 6. Развитие репозитория с 2009 по 2014 год.

Гистограммы, отражающие студенческие оценки затраченного времени, изображены на рисунке 4. При построении гистограмм использовались данные, собранные во время лекций в группах 174 и 074 (студенты третьего курса в 2014 и 2013 годах, соответственно). Данные по оценкам затраченного времени приведены в таблицах ?? и 4. Задание “Web”, заключавшееся в подготовке web-интерфейса с примером работы реализованного студентом алгоритма, в 2014 году не было включено в программу, так как в этом году студенческие работы были более насыщены в теоретическом плане. Из рисунка 4, i, видно, что наибольшее количество времени у студентов уходит на написание кода (в рамках заданий SBR и CV) и его отладку/доработку разработанных методов (задание E).

Работа с репозиторием Важным аспектом курса является доступность студенческих результатов как для студентов последующих лет, так и для произвольного исследователя. Для этого итоговые результаты работы, текст статьи с описанием исследования и код, позволяющий повторно получить те же результаты загружаются в репозиторий [28]. Структура репозитория такова: для каждой группы заводится папка с названием GroupNo (например, Group174), в которую студенты загружают результаты своих проектов в течение трех с половиной лет обучения на кафедре. Для нового проекта студент создает папку GroupNo/Surname2014ProjectName, содержащую все файлы и папки проекта:

- папка doc с документацией и со статьей (обязательно хранить там .pdf, .tex, финальные версии рисунков .eps/.pdf; удалять все временные файлы),
- папка code с кодом проекта. В этой папке выделить файл, который нужно запустить, чтобы посмотреть отчет, например demoProjectName.mat,
- папка data с данными по проекту (исходными и производными в .mat),
- (необязательно) папка report с промежуточными рисунками (рисунки и черновики отчетов вместе с кодом хранить нежелательно),
- не следует хранить чужие статьи в репозитории; их удобнее хранить в базе Mendeley [43].

Список студенческих проектов за 2009–2013 года со ссылками на основной документ, аннотациями и библиографической информацией опубликован на странице [1].

На рисунке 5 отражена динамика добавления и изменения файлов в репозитории за 2009–2014 года. Важным для наблюдения показателем проводимых работ является не только рост количества файлов в репозитории, но и частота изменения файлов, так как

она характеризует используемость предыдущих результатов. Можно отметить возрастание количества изменяемых файлов с 2012 года. На рисунке 6 изображено изменение со временем следующих характеристик

- количество правок в год, содержащих добавление файлов,
- количество правок в год, содержащих изменение существующих файлов,
- количество правок в год, содержащих удаление файлов,
- количество добавленных за год файлов.

Наблюдение за процессом добавления кода при работе над проектом позволяет оценить общее качество проекта. Например, на рисунке 7 изображены временные зависимости количества строк кода (tex или matlab), написанных к моменту измерения для студентов, получивших наиболее высокие оценки за курс (Irina, Rimma, Mikhail, Anna) и студентов, получивших более низкие оценки (Anastasia, Andrey). Видно, что первые четыре студента выполняли работу последовательно: хотя бы в одну из папок doc или code вносились изменения более-менее равномерно на протяжении семестра. Для студентов Anastasia и Andrey история загрузок в репозиторий содержит четыре и пять дней соответственно (то есть студенты загружали новый материал только в эти дни). Сделать качественную работу за четыре однодневных подхода сложно, поэтому оценки студентов Anastasia и Andrey оказались ниже.

Разумеется, время и усилия, затраченные студентами на выполнение проекта не полностью описываются студенческой активностью при работе с репозиторием. Например, некоторые студенты намеренно не публикуют промежуточных результатов, если эти результаты не удовлетворяют их ожиданиям. Как правило, при подобном “скрывании” результатов студент не только не публикует полученных результатов, но не оформляет их вовсе, даже для себя. Такое поведение крайне нежелательно по нескольким причинам. Во-первых, при отсутствии визуализации полученных результатов, кода для воспроизведения поставленного эксперимента или текста с его описанием, консультанту сложно разобраться в причине неудачи и помочь студенту. Во-вторых, стремление к получению идеальных результатов на каждом этапе работы может привести к полной ее остановке. Стремясь улучшить получаемые результаты, но не документируя процесс их получения, студент очень скоро забывает, с чего начинал и на каком-то этапе заклинивается. (Именно так произошло со студенткой Anastasia, в результате чего работу пришлось сократить). Таким образом, частота загружаемых обновлений дает хорошую оценку качества выполненной работы.

Заключение

Описан курс, посвященный изучению методов научной работы и выполнения исследовательских проектов в области машинного обучения и анализа данных. Курс рассчитан на студентов технических вузов третьего курса и старше. Целью работы является написание научной статьи с элементами математической новизны с последующей подачей в научный журнал. В процессе работы студенты изучают методы и технику написания научных статей и проведения вычислительных экспериментов, а также сопутствующие этому процессу технологии — язык разметки научных текстов LaTeX, формат представления библиографической записи BibTeX, язык программирования высокого уровня m-code. В рамках курса студентам предлагается четкое разделение проекта на этапы: сбор и анализ литературы, математическая постановка задачи, описание метода решения задачи и исследование его свойств, проведение вычислительного эксперимента. Каждому студенту предлагается персональная тема, по которой он анализирует публикации отечественных и

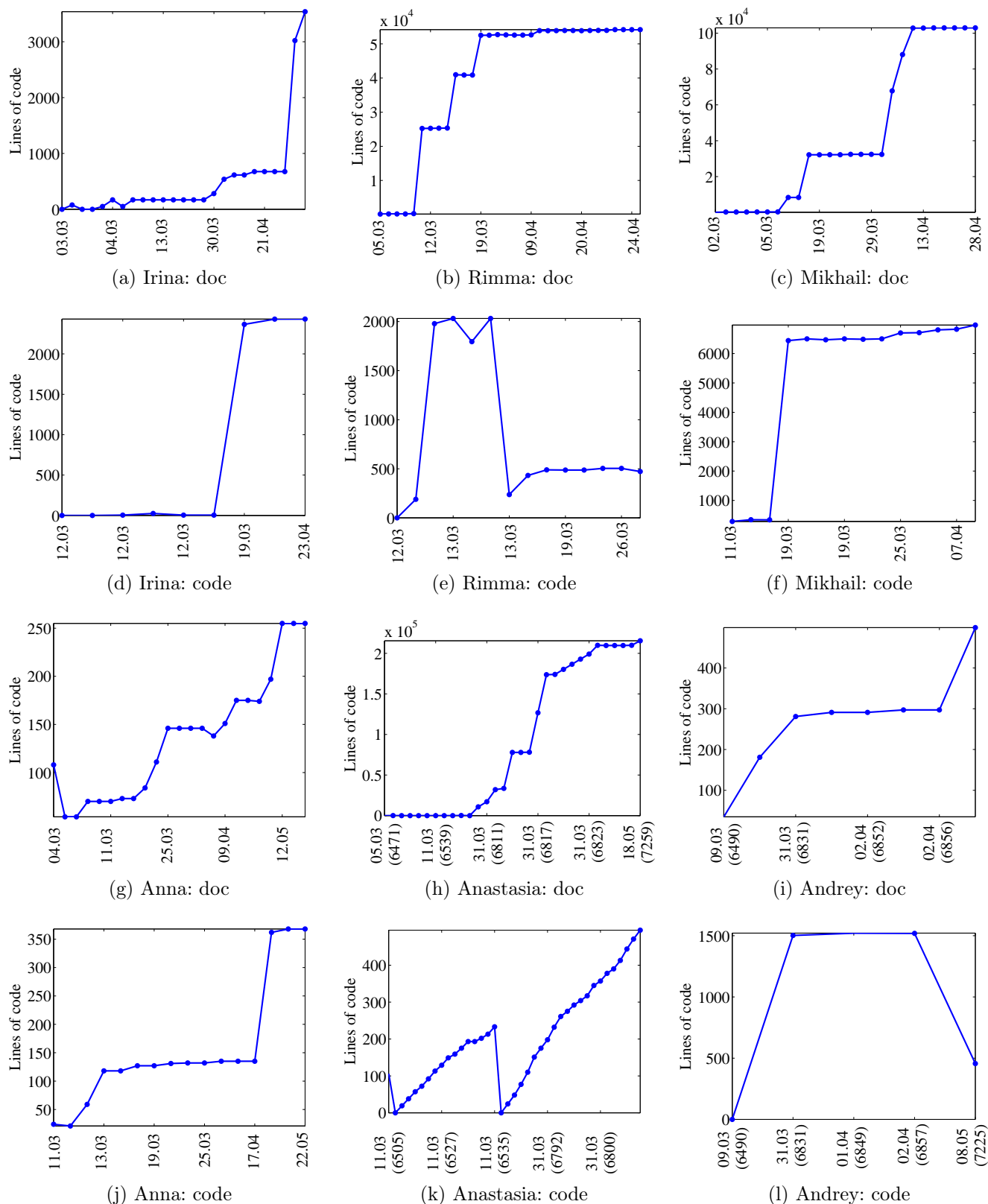


Рис. 7. Динамика количества строк в папках 'doc' и 'code' для студентов 174-ой группы.

зарубежных исследователей за последние десять лет, ставит задачу и делает доклад для группы о полученных результатах. Далее выполняется математическое описание метода,

делается промежуточный доклад о состоянии работ. Последним шагом работы становится вычислительный эксперимент, иллюстрирующий свойства метода и использующий синтетические или реальные данные. Каждая статья рецензируется одногруппниками автора, работы синхронизируются на сайте SourceForge.org, проект “MLAlgorithms”.

Для повышения уровня выполняемых работ в структуру курса были введены роли эксперта — поставщика задач и данных — и консультантов — старших студентов и аспирантов, консультировавших третьекурсников и контролировавших процесс работы над проектом. Эти изменения позволили

- получить достаточное количество актуальных задач, которые можно предложить студентам для исследований,
- обеспечить студентам возможность проконсультироваться с более опытным исследователем на любом этапе выполнения работы.

Участие в курсе позволяет экспертам, имеющим данные и необходимость их исследовать, получить, хотя бы частично, ответы на интересующие их вопросы. Аналогичным образом, чтобы заинтересовать консультантов в участии в проекте, среди предложенных экспертами задач были выбраны задачи, входившие в область интересов хотя бы одного из консультантов. Таким образом, каждый консультант контролировал выполнение исследования, перекликающегося с темой его собственной научной работы. В результате из шестнадцати студентов лишь двое не завершили работу над проектом, при этом девять работ было подготовлено к подаче в рецензируемые журналы. Все консультанты выразили заинтересованность в работе со студентами в следующем году. Также планируется привлечение дополнительных экспертов для расширения базы решаемых задач.

В работе описаны подробности организации курса: развернутый план лекций и домашних заданий, характер взаимодействия между студентом и консультантом, система оценивания, техническая поддержка курса. Приведены основные ошибки студентов, совершаемые на каждом этапе выполнения исследования и написания статьи на его основе, и советы по их устранению.

Особенностью данного курса по сравнению с аналогичными курсами, построенными на выполнении студентами исследовательских проектов в течение семестра, заключается, во-первых, в определении качества студенческих работ. Итогом работы обязательно должна стать научная статья. Несмотря на то что студент может получить отлично, не подавая статью в журнал, высшую оценку получают только студенты, подготовившие статью к подаче в журнал. Таким образом, к защите бакалаврского диплома студент, прошедший три семестра курса, может иметь до трех публикаций.

Кроме того, даже если студент не подал статью в журнал или статья не была принята, написанная статья и все сопутствующие материалы остаются в свободном доступе — в репозитории для хранения и обмена результатами между студентами и аспирантами кафедры. Студенты, не раз обращавшиеся при выполнении исследования к работам прошлых лет, понимают, что представляемые ими результаты будут впоследствии изучены и, вероятно, использованы другими студентами. Подобная открытость результатов положительно сказывается на качестве выполняемых работ: студенты работают не ради оценки, но для того чтобы внести вклад в общую базу знаний. К тому же, непрерывное пополнение репозитория студенческими наработками позволяет постоянно увеличивать сложность предлагаемых студентам задач без потери качества их решения.

Литература

- [1] Список студенческих работ, выполненных с 2009 по 2013 год: http://www.ccas.ru/jmllda/bib_refs_2013.
- [2] Архив журнала машинное обучение и анализ данных http://jmllda.org/?page_id=35.
- [3] Страница регистрации на wiki-ресурсе machinelearning.ru, 2014.
- [4] Статья о репозитории Source Forge на machinelearning.ru.
- [5] TortoiseSVN (a Revision control / version control / source control software for Windows) download page: <http://tortoisesvn.net/downloads.html>, 2014.
- [6] P. R. Halmos. How to write mathematics. *Amer. Math. Soc.*, 1983.
- [7] Robert Gerver. *Writing Math Research Papers: A Guide for Students and Instructors*. Key Curriculum.
- [8] Kate L. Turabia. *A Manual for Writers of Research Papers, Theses, and Dissertations, Eighth Edition: Chicago Style for Students and Researchers*. Wiley, 2013.
- [9] Шаблон описания задачи, предоставляемый экспертом: <http://www.machinelearning.ru>, 2014.
- [10] Paul R. Pintrich. The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research*, 31:459–470, 1999.
- [11] Noel J. Entwistle and Elizabeth R. Peterson. Conceptions of learning and knowledge in higher education: Relationships with study behaviour and influences of learning environments. *International Journal of Educational Research*, 41(6):407–428, 2004.
- [12] W. Damon and E. Phelps. Critical distinctions among three approaches to peer education. *International Journal of Education*, 13:9–19, 1989.
- [13] P. Blatchford, A. Russell, and R. Webster. *Reassessing the impact of teaching assistants: How research challenges practice and policy*. Oxon, UK: Routledge, 2012.
- [14] Julie Radforda, Paula Bosanquet, Rob Webster, Peter Blatchforda, and Christine Rubie-Davies. Fostering learner independence through heuristic scaffolding: A valuable role for teaching assistants. *International Journal of Educational Research*, 63:116–126, 2014.
- [15] A. Fernandez-Santander. Cooperative learning combined with short periods of lecturing: A good alternative in teaching biochemistry. *Biochemistry and Molecular Biology Education*, 36(1):34–38, 2008.
- [16] Edward E. Gordon. *Peer Tutoring: A Teacher's Resource Guide*. ScarecrowEducation, 2005.
- [17] Yale Teaching Center: Teaching Students to Write Good Papers. <http://www.yale.edu/graduateschool/teaching/goodpapers.html>, 2014.
- [18]
- [19] Страница курса “Численные методы обучения по прецедентам (практика, В.В. Стрижов)/Группа 174, весна 2014” на machinelearning.ru, 2014.
- [20] GNU Octave homepage: <http://www.gnu.org/software/octave/>, 2014.
- [21] The MiKTeX project page: <http://miktex.org>, 2014.
- [22] The TeX Live project page: <http://www.tug.org/texlive/>, 2014.
- [23] The JabRef (an open source bibliography reference manager) project page: <http://jabref.sourceforge.net/>.
- [24] Open Source community resource SourceForge.net: <http://sourceforge.net/>.

- [25] Слайды по системам контроля версий:
http://www.machinelearning.ru/wiki/images/2/29/MMP_Praktikum317_2013s_VCS.pdf, 2014.
- [26] The RabbitVCS (version control software for Linux) project page: <http://rabbitvcs.org/>, 2014.
- [27] Страница проекта “Algorithms of Machine Learning” на sourceforge.net:
<https://sourceforge.net/projects/mlalgorithms/>, 2014.
- [28] <https://svn.code.sf.net/p/mlalgorithms/code>, 2014.
- [29] Статья о системе набора в макропакете LaTeX:
<http://ru.wikipedia.org/wiki/LaTeX><http://ru.wikipedia.org/wiki/LaTeX>, 2014.
- [30] С. М. Львовский. *Набор и верстка в системе LaTeX*. МЦНМО, 2006.
- [31] Архив со стилевым файлом и шаблоном статьи:
<http://www.machinelearning.ru/wiki/images/4/4f/Jmla-guides.zip>, 2014.
- [32] Статья, посвященная инструменту для создания форматированных списков библиографии BibTeX: <http://en.wikipedia.org/wiki/Bibtex>, 2014.
- [33] Пример библиографической базы.
- [34] Пример библиографической записи:
<http://liinwww.ira.uka.de/cgi-bin/bibshow?e=Njtd0ECMQ03121/fyqboefe%7d81352582&r=bibtex&mode=intra>.
- [35] Список инструментов для поиска литературы и баз данных, содержащих библиографическую информацию об источниках:
http://en.wikipedia.org/wiki/List_of_academic_databases_and_search_engines, 2014.
- [36] Введение в Матлаб: <http://www.machinelearning.ru/wiki/index.php?title=Matlab>.
- [37] Статья “Документирование функций Matlab, соглашение об именах переменных и создание отчетов о вычислительных экспериментах” на machinelearning.ru, 2014.
- [38] Richard Johnson. *MATLAB Programming Style Guidelines*, version 1.5 edition, October 2002.
- [39] Рекомендации по оформлению графиков:
<http://www.machinelearning.ru/wiki/index.php?title=JMLDA/Fig>.
- [40] Архив со списком рекомендуемых обозначений:
<http://www.machinelearning.ru/wiki/images/3/37/Strijov2013Notation.zip>, 2014.
- [41] С.Р. Bishop. *Pattern recognition and machine learning*. 2006.
- [42] Статья на machinelearning.ru с рекомендациями по написанию отчетов, 2014.
- [43] Mendeley reference manager: <http://www.mendeley.com/>.
- [44] John W. Creswell. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research (3d Edition)*. Upper Saddle River: Pearson., 2008.
- [45] Н Кара. *Research and Evaluation for Busy Practitioners: A Time-Saving Guide*. Bristol: The Policy Press, 2012.
- [46] T.S. Rocco, T. Hatcher, and J.W. Creswell. *The handbook of scholarly writing and publishing*. San Francisco. San Francisco, CA: John Wiley & Sons, 2011.
- [47] William Strunk Jr., E. B. White, and Roger Angell. *The Elements of Style, Fourth Edition*. Bartleby.com, 1999.
- [48] William Zinsser. *On Writing Well: The Classic Guide to Writing Nonfiction*. HarperCollins Publishers, 2006.
- [49] The TexnicCenter (a Latex editor) project page: <http://www.texniccenter.org>, 2014.
- [50] The TeXworks (crossplatform working environment for Tex) project page:
<http://www.tug.org/texworks/>, 2014.