Байесовский подход к решению задачи прогнозирования на основе информации экспертов и таблицы данных*

Лбов Г. С., Неделько В. М.
lbov@math.nsc.ru, nedelko@math.nsc.ru
Новосибирск, Институт математики СО РАН

Предлагается постановка и метод решения некоторой достаточно общей задачи прогнозирования на основе анализа эмпирической информации, которая может быть представлена либо набором высказываний экспертов, либо таблицей экспериментальных данных (выборкой), либо экспертными высказываниями и таблицей данных одновременно. При этом допускается, что информация различных экспертов несогласована. Для решения рассматриваемой задачи предлагается реализация байесовского подхода, использующая слабые ограничения на класс распределений. Эффективность метода иллюстрируется на примерах.

В работе предлагается метод решения задачи прогнозирования на основе анализа эмпирической информации, которая может быть представлена либо набором высказываний экспертов, либо таблицей экспериментальных данных (выборкой), либо экспертными высказываниями и таблицей данных одновременно.

Чтобы распространить байесовский подход на случай, когда эмпирическая информация представлена экспертными суждениями, предлагается статистическая интерпретация экспертной информации. При этом предполагается, что экспертная информация представлена высказываниями нескольких экспертов. Допускается, что высказывания различных экспертов могут быть несогласованы и, в-частности, быть противоречащими друг другу в той или иной степени, совпадающими, частично совпадающими, дополнительными.

Одним из достоинств предлагаемой в работе реализации байесовского подхода к решению поставленной задачи является использование слабых гипотез относительно распределений.

Постановка задачи

Рассмотрим следующую задачу прогнозирования.

Пусть имеется генеральная совокупность объектов Γ, для которой определена произвольная вероятностная мера $\mathcal{P}(\Gamma)$. Каждый объект $a \in \Gamma$ может быть охарактеризован значениями переменных $X_1, ..., X_n$, а также значениями так называемых целевых (прогнозируемых) переменных $Y_1, ..., Y_m$, то есть каждому $a \in \Gamma$ путем проведения измерений могут быть сопоставлены значения $x_1, ..., x_n; y_1, ..., y_m$ переменных $X_1, ..., X_n; Y_1, ..., Y_m$. Данные переменные могут быть произвольных типов (вещественные, целые, порядковые, номинальные, бинарные).

Рассматриваемая задача состоит в том, чтобы для произвольного объекта a из Γ по известным значениям переменных $X_1, ..., X_n$ предсказать значения переменных $Y_1, ..., Y_m$ на основе анализа имеющейся эмпирической информации. Заметим, что задачи построения решающей функции распознавания и регрессионной функции являются частным случаем рассматриваемой задачи.

Обозначим через D_{X_j} – множество допустимых значений переменной X_j, D_{Y_j} – множество допустимых значений переменной Y_j, $D_X \stackrel{def}{=} \prod_{j=1}^{n} D_{X_j}$, $D_Y \stackrel{def}{=} \prod_{j=1}^{m} D_{Y_j}$, $D \stackrel{def}{=} D_X \times D_Y$. Тогда $x = (x_1, ..., x_n)$ может рассматриваться как точка в пространстве D_X, $y = (y_1, ..., y_m)$ – точка в пространстве D_Y, $z = (x_1, ..., x_n; y_1, ..., y_m)$ будет обозначать точку в пространстве D. Заметим, что пространство D в общем случае является разнотипным и может быть разложено в прямое произведение непрерывного D_c и дискретного D_d подпространств.

Поскольку значения всех переменных могут быть измерены для любого $a \in \Gamma$, то существует отображение из Γ в D, которым (учитывая существование меры $\mathcal{P}(\Gamma)$) в пространстве D определяется вероятностная мера $\mathcal{P}(D)$.

Введем в пространстве D меру μ следующим образом. Поскольку любая область E дискретно-непрерывного пространства D может быть представлена как $E = \bigcup_{j=1}^{|E_d|} E_c^j \times \{z_d^j\}$, где E_d – проекция E на D_d, z_d^j – точка из E_d, E_c^j – соответствующая z_d^j область в D_c, то меру произвольной подобласти E естественно положить равной $\mu(E) \stackrel{def}{=} \sum_{j=1}^{|E_d|} \mu(E_c^j)$, где $\mu(E_c^j)$ – лебегова мера множества E_c^j.

Предположим, что отображение $\Gamma \to D$ таково, что существует $\rho(z)$ – плотность меры $\mathcal{P}(D)$ относительно меры μ, т. е. для любого измеримого подмножества E пространства D выполняется $P(E) = \int_E \rho(z)\, d\mu \stackrel{def}{=} \int_E \rho(z)\, dz$.

Введем обозначение

$$g(z) \equiv g(x, y) \stackrel{def}{=} \rho(y/x) = \frac{\rho(z)}{\int_{D_X} \rho(z)\, dx}.$$

*Работа выполнена при финансовой поддержке РФФИ, проект № 95-01-00930-а.

Таким образом, $\rho(^y/_x)$ представляет собой условную плотность распределения в пространстве D_Y при условии, что значения переменных $X_1, ..., X_n$ равны $x = (x_1, ..., x_n)$.

Под задачей прогнозирования будем понимать восстановление условной плотности $\rho(^y/_x)$ на основе эмпирической информации, то есть построение некоторой оценки $\overline{g}(z)$ функции $g(z)$.

Для удобства дальнейших обозначений введем множество $\mathcal{C}$, элементы которого будем называть стратегияммиприроды и обозначать c. При этом каждой из всевозможных плотностей $\rho(^y/_x)$, то есть каждой функции $g(z)$ поставим во взаимнооднозначное соответствие некоторую $c \in \mathcal{C}$. В дальнейшем вместо $g(z)$ будем записывать $g_c(z)$.

В отличие от известных подходов к решению задач восстановления распределений на основе выборки предлагаемый в данной работе метод ориентирован на использование эмпирической информации как в виде таблицы экспериментальных данных, так и в виде набора высказываний, полученных от многих экспертов.

Формально высказывание эксперта может быть представлено в виде следующей тройки:

$$B_i = \langle E^i, \{(V_k^i, \beta_k^i) | k = \overline{1, l_i}\}, \gamma_i \rangle,$$

где $i = \overline{1, N}$ (N – число высказываний, полученных от экспертов). При этом $E^i \subseteq D_X$, $V_k^i \subset D_Y$, при $k \neq l$ $V_k^i \bigcap V_l^i = \oslash$; $\beta_k^i \in [0,1]$, $\sum\limits_{k=1}^{l_i} \beta_k^i \leqslant 1$, $\gamma_i \in [0,1]$.

Данное высказывание может быть проинтерпретировано следующим образом: "Если $x(a)$ – некоторая (сответствующая объекту a) точка из D_X, то вероятность того, что $y(a)$ будет находиться в области V_1^i равна β_1^i, вероятность попадания $y(a)$ в V_2^i равна β_2^i, . . . ". Число γ_i отражает уверенность эксперта в истинности сделанного утверждения (подробнее интерпретацию уверенности см. [2]). Такого вида высказывание несет в себе информацию о функции $g(z)$. Действительно, $\forall x \in E^i$, $\beta_k^i \simeq \int\limits_{V_k^i} g(x,y)\, dy$. Знак $\simeq$ означает, что β_k^i — это лишь экспертная оценка соответствующей вероятности, а не ее истинное значение. Заметим, однако, что смысл высказывания имеет для нас значение лишь с точки зрения его влияния на зависимость вероятности появления данного высказывания от стратегии c.

Предположим, что для каждой стратегии c нам известна $P(^{B_i}/_c)$ – условная вероятность появления высказывания B_i. Тогда, поскольку эксперты делают высказывания независимо, то $P(^{\mathcal{B}}/_c) = \prod\limits_{i=1}^{N} P(^{B_i}/_c)$. Здесь $\mathcal{B} = \{B_i | i = \overline{1, N}\}$ – множество всех высказываний, представленных экспертами.

Если помимо экспертной информации мы имеем и обычную таблицу экспериментальных данных, т. е. $\mathcal{B} = \{B_i | i = \overline{1, N}\} \bigcup \{v_k | k = \overline{1, M}\}$, где $v_k = z(a_k)$ – вектор значений переменных, измеренных для объекта a_k, случайным образом выбранного из Γ, то $P(^{\mathcal{B}}/_c) = \prod\limits_{i=1}^{N} P(^{B_i}/_c) \cdot \prod\limits_{k=1}^{M} P(^{v_k}/_c)$. При этом $P(^{v_k}/_c) = g_c(z(a_k))\ f(z(a_k))$, где $f(z)$ – некоторая функция, не зависящая от c (в дальнейшем при подстановке в формулу Байеса $f(z(a_k))$ сократится).

Байесовский подход

Чтобы воспользоваться байесовским подходом к восстановлению $g(z)$, нужно задать априорную вероятностную меру $\mathcal{P}(\mathcal{C})$. После этого

$$\overline{g}^*(z) = \frac{\int\limits_{\mathcal{C}} g_c(z) P(^{\mathcal{B}}/_c)\ d\mathcal{P}(\mathcal{C})}{\int\limits_{\mathcal{C}} P(^{\mathcal{B}}/_c)\ d\mathcal{P}(\mathcal{C})}. \quad (1)$$

Данное выражение, полученное из формулы Байеса, дает оптимальную оценку для функции $g(z)$ в том смысле, что $\overline{g}^*(z)$ есть условное (при заданном $\mathcal{B}$) математическое ожидание функции $g(z)$.

Алгоритм, который на основе имеющейся информации $\mathcal{B}$ строит оценку функции $g(z)$ в соответствии с формулой (1), будем называть байесовским алгоритмом решения рассматриваемой задачи прогнозирования.

Обычно в теории статистических решений при задании априорной меры $\mathcal{P}(\mathcal{C})$ вводятся сильные ограничения на класс стратегий природы, которые во многих случаях не оправданы. В данной работе используются слабые ограничения на класс $\mathcal{C}$. Для этого вводится весьма широкий подкласс стратегий $\overline{\mathcal{C}} \subset \mathcal{C}$.

Стратегия $\overline{c}$ принадлежит классу $\overline{\mathcal{C}}$ тогда и только тогда, когда $g_{\overline{c}}(z)$ – кусочно-постоянна, причем области постоянства заданы в виде декартова произведения $(\prod\limits_{j=1}^{n} E_{X_j}) \times (\prod\limits_{j=1}^{m} E_{Y_j})$, $E_{X_j} \subseteq D_{X_j}$, $E_{Y_j} \subseteq D_{Y_j}$, где E_{X_j} – интервал, если X_j – переменная с упорядоченным множеством значений, и E_{X_j} – произвольное подмножество из D_{X_j}, если X_j – номинальная переменная, то есть переменная с конечным неупорядоченным множеством значений. Условия на E_{Y_j} аналогичны. Число областей постоянства функции $g_{\overline{c}}(z)$ будем называть сложностью стратегии $\overline{c}$.

Класс $\overline{\mathcal{C}}$ является универсальным, то есть $\forall c \in \mathcal{C}$, $\forall \varepsilon > 0$, $\exists \overline{c} \in \overline{\mathcal{C}}$: $\Delta(c, \overline{c}) < \varepsilon$, где $\Delta(c, \overline{c}) \stackrel{def}{=} \int\limits_{D} (g_c(z) - g_{\overline{c}}(z))^2\, dz$. Доказательство утверждения аналогично доказательству теоремы об универсальности класса логических решающих функций, изложенной в [4].

При задании вероятностной меры на $\overline{\mathcal{C}}$ более простым $\overline{c}$ приписывается бóльшая вероятность, нежели более сложным, то есть мера подкласса стратегий сложности l_1 меньше меры подкласса стратегий сложности l_2, если $l_1 > l_2$.

Благодаря универсальности класса $\overline{\mathcal{C}}$, вероятностную меру на нем оказалось возможным задать так, что

$$\forall c \in \mathcal{C}, \forall \varepsilon > 0, \int\limits_{\overline{c}:\Delta(c,\overline{c})<\varepsilon} d\mathcal{P}(\overline{\mathcal{C}}) > 0.$$

Ввиду указанного свойства введенной меры $\mathcal{P}(\overline{\mathcal{C}})$ ограничение, связанное с тем, что мера вводится не на $\mathcal{C}$, а на $\overline{\mathcal{C}} \subset \mathcal{C}$, оказывается практически несущественным.

Статистическая интерпретация экспертной информации

В отличие от известных методов логического анализа экспертных высказываний в данной работе предлагается статистическая интерпретация экспертной информации. При этом полагается, что вероятность появления каждого высказывания зависит от того, какая стратегия природы имеет место.

Эксперта s будем представлять как оператор, сопоставляющий каждой стратегии c условное распределение вероятностей $P_s({}^{u}/_{c})$ на множестве $\mathcal{U}$, где $\mathcal{U}$ – множество подмножеств всех допустимых высказываний. Заметим, что $\mathcal{U}$ – конечно ввиду конечности алфавита, в котором записываются высказывания, и естественных ограничений на длину и количество высказываний. Если эксперты, делающие высказывания, случайно выбираются из множества S в соответствии с некоторой вероятностной мерой $\mathcal{P}(S)$, то $P({}^{u}/_{c}) = \underset{s}{\in} S \to \sum P_s({}^{u}/_{c})P(s)$.

Заметим, что функция $P({}^{u}/_{c})$ является "в природе"вполне определенной, а потому ее конкретный вид может быть установлен непосредственным экспериментальным исследованием.

Примеры

В прилагаемых примерах иллюстрируется случай, когда имеется только одна бинарная целевая переменная Y с множеством значений $\{0, 1\}$. В этом случае вместо $g_c(z)$ достаточно рассматривать $g_c(x, 0)$ – условную вероятность $P({}^{y(a)=0}/_{x(a)=x})$, поскольку $g_c(x, 1) = 1 - g_c(x, 0)$.

Высказывания экспертов в этом случае можно привести к виду: $B_i = \langle E_i, (\{0\}, \beta^i), \gamma_i \rangle$. Здесь $V^i = \{0\}$. Если экспертом использовалась область $V^i = \{1\}$, то ее можно заменить на $\{0\}$, заменив также β^i на $1 - \beta^i$. Высказывание теперь можно записать в более простом виде $\langle E_i, \beta^i, \gamma_i \rangle$, где $E_i \subseteq D_X$, β^i – предполагаемое экспертом значение функции $g(x, 0)$ при $x \in E_i$, γ_i – уверенность в правильности оценки.

Получаемую в результате применения (1) оценку $\overline{g}^*(x, 0)$ будем записывать также в форме набора высказываний, которые будут иметь вид: $\langle \tilde{E}_l, \tilde{\beta}_l \rangle$, где $\tilde{E}_l \subseteq D_X$, $\overline{g}^*(x, 0) = \tilde{\beta}_l$, при $x \in \tilde{E}_l$; $l = \overline{1, \tilde{N}}$, где $\tilde{N}$ – число областей постоянства функции $\overline{g}^*(x, 0)$ в пространстве D_X.

Пример 1. Пусть исходная информация представлена единственным высказыванием:

1) $\langle\, E,\ 0.9, 0.8 \,\rangle$.

Применяя соотношение (1), получаем:

1) $\langle\, E,\ 0.86 \,\rangle$.

Видим, что $\tilde{\beta}$ оказалась не равной β, а сдвинутой к 0.5 (т. е. в сторону меньшей определенности). Это вполне оправдано, поскольку уверенность эксперта (как и доверие ему) не абсолютна.

Пример 2. Пусть даны два одинаковых высказывания двух экспертов:

1) $\langle\, E,\ 0.9, 0.8 \,\rangle$;
2) $\langle\, E,\ 0.9, 0.8 \,\rangle$.

В результате получим:

1) $\langle\, E,\ 0.89 \,\rangle$.

Оценка $\tilde{\beta}$ уже гораздо ближе к β_1 (или к β_2), чем в первом случае. Это означает, что степень доверия двум экспертам, независимо подтверждающим мнения друг друга, выше, чем доверие высказыванию одного эксперта.

Пример 3. Дана информация:

1) $\langle\, \{a, b\},\ 0.9, 0.8 \,\rangle$;
2) $\langle\, \{b, c\},\ 0.1, 0.8 \,\rangle$;
3) $\langle\, \{c, d\},\ 0.9, 0.8 \,\rangle$.

В данном примере пространство D_X состоит из значений $\{a, b, c, d\}$ единственной номинальной переменной X.

Результат согласования:

1) $\langle\, (x = a),\ 0.8 \,\rangle$; 2) $\langle\, (x = b),\ 0.72 \,\rangle$;
3) $\langle\, (x = c),\ 0.72 \,\rangle$; 4) $\langle\, (x = d),\ 0.8 \,\rangle$.

Результат представляется интуитивно вполне оправданным. Действительно, поскольку высказывания противоречат друг другу, то оценки для $g(x, 0)$ значительно сдвинуты к 0.5. Далее, ввиду того, что второй эксперт противоречит одновременно как первому так и третьему, то доверие его утверждению относительно ниже. Этим, в частности, объясняется, например, то, что для точки b решение ближе к оценкам первого эксперта, чем второго.

Заметим, однако, что непосредственная реализация байесовского алгоритма приводит к большому объему вычислений. В связи с этим был разработан некоторый эвристический алгоритм принятия решения на основе несогласованной информации экспертов [3]. Данный алгоритм, имея вполне приемлемую трудоемкость, позволяет получить результаты, с хорошей точностью приближающие ответ, получаемый при непосредственном применении соотношения (1).

Литература

[1] *Лбов Г. С., Неделько В. М.* Распознавание образов на основе вероятностных логических высказываний

экспертов. // Тезисы докладов Всероссийской конференции "Математические методы распознавания образов (ММРО - 6) 22–26 ноября 1993, с. 40–41.

[2] *Лбов Г. С., Неделько В. М.* Анализ и прогноз экологической ситуации на основе информации нескольких различных экспертов. // Математические проблемы экологии. Сборник статей. Институт математики СО РАН, 1994 г., с. 118–125.

[3] *Lbov G. S., Nedelko V. M.* The Consrtucting of Decision Rule, Using the Probabilistic Statements of an Experts (Построение решающей функции распознавания на основе вероятностных высказываний экспертов). // Int. J. of Pattern Recognition And Image Analysis, 1995, vol. 5, No 2, pp. 165-–171.

[4] *Лбов Г. С., Старцева Н. Г.* Сложность распределений в задачах классификации. // Доклады РАН, Том 338, ь5, 1994.