УДК 004.832.38 DOI: 10.17587/prin.8.177-185

А. Н. Шмарин, аспирант, e-mail: tim-shr@mail.ru,
Воронежский государственный университет

Кластеризация множества слоев в задаче нечеткого LP-вывода

LP-структуры представляют алгебраическую модель логических систем продукционного типа, широко распространенных в информатике. Представлен алгоритм кластеризации слоев в LP-структуре. Слой — это одноэлементная выборка из классов эквивалентности фактор-множества, полученного разбиением LP-структуры по уникальным правым частям пар бинарного отношения. Исследованы вопросы практической реализации данного алгоритма. Результаты могут быть использованы для разработки интеллектуальных систем продукционного типа, работающих с неполными знаниями, а также минимизирующих число трудоемких запросов о значениях признаков классифицируемых объектов предметной области.

Ключевые слова: *LP-вывод, бинарное отношение, фактор-множество, кластеризация, стохастический алгоритм, машинное обучение*

Введение

С увеличением объемов и сложности обрабатываемой информации все более актуальным становится использование систем искусственного интеллекта. При этом одной из наиболее распространенных моделей представления знаний является продукционная модель [1]. Алгоритмы поиска решений в системах искусственного интеллекта основаны на механизме логического вывода. На практике существует большое число задач принятия решений в условиях неопределенности, для которых получение информации о значениях признаков классифицируемых объектов некоторой предметной области является ресурсоемкой операцией. Такие задачи возникают, например, когда робот исследует поверхность другой планеты. Если цель робота состоит в том, чтобы добраться до определенной скалы, но будущий маршрут является наблюдаемым лишь отчасти, то робот должен попытаться выработать оптимальное решение, касающееся того, как достичь цели с учетом ограниченности ресурсов, доступных для дополнительной разведки местности. Другим примером, в котором стоимость получения новой информации может быть высока, является коммерческая медицина. Для минимизации издержек необходимо находить приемлемый способ лечения с помощью минимального количества анализов, выполненных за минимальное время. Задержка в предоставлении лечения, вызванная проведением дополнительных анализов, приводит к увеличению затрат. Более того, состояние пациента, не дождавшегося помощи, может существенно ухудшиться. Стоимость дополнительных исследований является существенной и в сфере разведки полезных ископаемых. Может оказаться, что более выгодное решение состоит в том, чтобы приступить к бурению скважины, если уверенность в успехе составляет 95 %, чем потратить значительные средства, чтобы добиться 98 %-ной уверенности.

С фактором ресурсоемкости операций получения данных для принятия решения связана задача минимизации числа запросов о признаках классифицируемого объекта (из некоторой предметной области) в ходе логического вывода. Данная задача является NP-трудной [2]. Для достижения глобальной минимизации числа внешних запросов предложен общий метод LP-вывода [3]. К сожалению, он обладает экспоненциальной вычислительной сложностью относительно числа атомарных фактов в базе знаний. Идея кластерно-релевантного LP-вывода [4] основана на вычислении специфических оценок (показателей релевантности) для ограниченного подмножества продукций и на обобщении полученных оценок на все множество продукций. Исследования и эксперименты показывают [5], что использование метода LP-вывода по сравнению с использованием обычного обратного вывода снижает число внешних запросов в среднем на 15...20 %.

Ранее [6] обсуждались вопросы реализации вычисления приближенной оценки числа слоев без циклов в задаче нечеткого LP-вывода. В настоящей работе рассмотрена математическая формулировка задачи нечеткого LP-вывода, исследованы свойства монотонности функции оценки числа слоев без циклов, а также предложен алгоритм кластеризации множества слоев — выборок по одному элементу из классов эквивалентности фактор-множества, полученного разбиением заданного на алгебраической решетке бинарного отношения по уникальным правым частям пар. Данный алгоритм предназначен для итеративного анализа таких подмножеств продукций, которые наиболее существенно влияют на показатель релевантности.

О методе нечеткого LP-вывода

Обозначим через X множество объектов (ситуаций, прецедентов) некоторой предметной области. Например, в задачах машинного обучения, встречающихся в медицине, объектами могут являться пациенты, в сфере кредитования — заемщики, в задаче фильтрации спама — отдельные сообщения.

Признак — результат измерения некоторой характеристики объекта, т. е. отображение *feature*: $X \to D_{feature}$, где $D_{feature}$ — множество допустимых значений признака. Значениями признаков в прикладных задачах могут быть числовые последовательности, изображения, тексты, функции, графы, результаты запросов к базе данных и т. д.

Пусть имеется набор признаков $feature_1, ..., feature_n$, которые соответствуют характеристикам объектов $x \in X$ некоторой предметной области. Вектор $(feature_1(x), ..., feature_n(x)) \in D_{feature_1} \times ... \times D_{feature_n}$ называется признаковым описанием объекта $x \in X$. Признаковое описание можно отождествлять с самими объектами, т. е. $X = D_{feature_1} \times ... \times D_{feature_n}$.

Предположим, что объекты X разделены некоторым образом на классы, которым соответствует конечное множество их номеров (имен, меток) Y. Пусть также задана продукционная база знаний, отражающая целевую зависимость — отображение $y^* : X \to Y$.

Задача нечеткого LP-вывода состоит в том, чтобы как можно достовернее классифицировать произвольный объект $x \in X$, т. е. найти для него класс $y^*(x) \in Y$, используя для этого как можно меньшую выборку значений из признакового описания этого объекта.

Далее рассмотрена более формализованная постановка задачи в терминах алгебраических решеток и бинарных отношений.

Пусть задано конечное множество $F = \{ f \}$. На его основе определим атомно-порожденную ограниченную алгебраическую решетку $L = \lambda(F)$, где λ — функция-булеан [7]. На решетке зададим дополнительное бинарное отношение $R = \{r = (\tau, f): \tau \in L, f \in F\}$, являющееся каноническим [5].

Применительно к продукционным логическим системам такое отношение R соответствует множеству продукций, F (множество атомов решетки) соответствует элементарным фактам продукционной системы, а сама решетка L соответствует множеству предпосылок и заключений продукций.

Атом x решетки L называется начальным при отношении R, если в R нет ни одной пары вида (τ, x).

Обозначим множество всех начальных атомов решетки как F_{init}: $F_{init} = \{f \in F : \nexists (\tau, f) \in R, \tau \in L\}$. Также обозначим множество всех атомов решетки, не являющихся начальными: $F_{notinit} = \{f \in F : \exists (\tau, f) \in R, \tau \in L\}$.

Упорядоченная пара $\{\alpha, \beta\} \in L$, $\alpha \in L$, $\beta \in L$ называется дистрибутивно связанной отношением R, если существует такое множество $\{(\alpha_1, \beta_1), ..., (\alpha_p, \beta_p)\}$, что $\forall i = \overline{1, p}$ выполняется следующее условие: $\alpha_i \subseteq \alpha \wedge \beta \subseteq \beta_i \wedge ((\alpha_i = \beta_i \in L) \vee ((\alpha_i, \beta_i) \in R))$.

Упорядоченная пара (α, β), $\alpha \in L$, $\beta \in L$ называется логически связанной отношением R, если она дистрибутивно связана отношением R, либо существует упорядоченный набор $(\alpha, \gamma_1, ..., \gamma_l, \beta)$, $\gamma_i \in L$, $i = \overline{1, l}$ такой, что каждая пара в последовательности (α, γ_1), (γ_1, γ_2), ..., (γ_{l-1}, γ_l), (γ_l, β) дистрибутивно связана отношением R. Если пара (α, β) логически связана отношением R, то β будем называть образом α при отношении R, а α — прообразом β при отношении R. Прообраз в атомно-порожденной решетке называется начальным, если все его атомы являются начальными (при отношении R).

Введем отображение $M_R : R \to [0, 1]$, которое каждому элементу отношения R ставит в соответствие степень истинности: $M_R(r) = \mu$, $r \in R$, $\mu \in [0, 1]$. Обозначим также $M_{init} : F_{init} \to [0, 1]$ — отображение, ставящее в соответствие некоторую степень истинности каждому начальному атому решетки.

Применение отображения M_{init} соответствует выполнению запроса о неизвестном значении некоторого признака классифицируемого объекта и может оказаться ресурсоемкой операцией.

Применительно к продукционным системам отображение M_R вычисляет степени истинности продукций, а M_{init} — степени истинности значений признаков. Степень истинности задается числом $\mu \in [0, 1]$ и означает коэффициент уверенности, с которым соответствующее утверждение истинно. Пусть, например, продукции вида "если $x_1 \wedge ... \wedge x_n$, то x_{n+1}" соответствует некоторый элемент $r \in R$. Тогда значение $M_R(r) \in [0, 1]$ будет означать степень, с которой истинно утверждение данной продукции. Аналогично, если для объекта предметной области $x \in X$ значению некоторого его признака $feature(x) \in D_{feature}$ соответствует атом $f \in F_{init}$, то значение $M_{init}(f) \in [0, 1]$ будет представлять степень, с которой истинно утверждение "признак *feature* объекта x равен значению $feature(x)$".

Во введенных обозначениях задача нечеткого LP-вывода заключается в том, чтобы в некотором подмножестве $F_C \subseteq F_{notinit}$ неначальных атомов решетки найти элемент, обладающий наибольшей степенью истинности $M_R(X)$, применив при этом отображение M_{init} как можно меньшее число раз. Данный элемент соответствует решению, имеющему наибольшую степень истинности. Множество F_C соответствует некоторому множеству классов, на которые разделены объекты предметной области.

Процесс поиска обладающего наибольшей степенью истинности $f_c \in F_C$ состоит из нескольких шагов. В первую очередь, для каждого атома решетки, принадлежащего множеству $F_C \subseteq F_{notinit}$, выполняется поиск начальных прообразов при отношении R (т. е. поиск для атома логически связанных отношением R подмножеств начальных атомов решетки). Затем на основе результатов поиска вычисляются степени истинности соответствующих атомов решетки. В заключение выбирается атом с наибольшей степенью истинности.

Предположим, что заданы некоторые бинарные операции T и T_{con} такие, что T является операцией t-нормы, а T_{con} — соответствующей ей операцией t-конормы [8]. Вычисление степеней истинности выполняется для всевозможных атомов $f_c \in F_C$ и соответствующих им начальных прообразов $\alpha \in L$. Для вычисления используют отображения M_R, M_{init}, операции t-нормы и t-конормы, а также обобщенное правило вывода *modus ponens*.

По каждому атому $f_c \in F_C$ формируется либо восстанавливается из кэш-памяти подмножество $R_1 \subseteq R$, которое логически связывает выбранные атом x и начальный прообраз α. Нахождение степени истинности некоторого неначального атома f, участвующего в логической связи α и x, начинается с выбора подмножества элементов, содержащих атом f в своей правой части: $R_f = \{(\tau, f) \in R_1, \tau \in L\}$. Обозначим мощность данного подмножества символом $n = |R_f|$. Тогда $R_f = \{r_1, ..., r_n\} = \{(\tau_1, f), ..., (\tau_n, f)\}$, где $\tau_i \in L$, $1 \leqslant i \leqslant n$. По определению, элементы τ_i решетки L состоят из ее атомов. То есть $\tau_i = \{g_{i,1}, ..., g_{i,m_i}\}$, где $g_{i,j} \in F$ — атомы решетки, m_i — число таких атомов в τ_i, $1 \leqslant j \leqslant m_i$, $1 \leqslant i \leqslant n$.

Используя введенные обозначения, можно выразить формулы для вычисления $\mu(f)$ — степени истинности некоторого атома f. Если $f \in F_{notinit}$, то атом является неначальным и его степень истинности задается рекурсивно:

$$\mu(f) = T_{con}\Big(T\Big(T\big(\mu(g_{1,1}), ..., \mu(g_{1,m_1})\big), M_R(r_1)\Big), ...,$$
$$T\Big(T\big(\mu(g_{n,1}), ..., \mu(g_{n,m_n})\big), M_R(r_n)\Big)\Big),$$

где $r_i = (\tau_i, f) \in R_f$, $\tau_i = \{g_{i,1}, ..., g_{i,m_i}\}$, $g_{i,j} \in F$, $1 \leqslant j \leqslant m_i$, $1 \leqslant i \leqslant n$.
Если же $f \in F_{init}$, т. е. атом является начальным, то $\mu(f) = M_{init}(f)$. Применение отображения M_{init} к некоторому начальному атому означает нахождение для этого атома соответствующей степени истинности. Поскольку атом является начальным, его степень истинности невозможно вывести на основе отношения R или степеней истинности других атомов решетки.

Поиск такой степени истинности эквивалентен выполнению внешнего запроса о значении неизвестного признака классифицируемого объекта. Минимизация числа подобных запросов достигается специальным порядком применения отображения M_{init} к начальным атомам решетки. В первую очередь, это отображение следует применять к тем начальным атомам, которые входят в наибольшее число найденных прообразов. Кроме того, соответствующие прообразы должны содержать как можно меньшее число элементов.

Порядок применения отображения M_{init} к начальным атомам решетки определяется на основе так называемых показателей релевантности [9]. Нахождение их точных значений является вычислительно трудной задачей [2]. Таким образом, для практического использования целесообразно вычислять показатели релевантности в ограниченных "кластерах" LP-структуры.

Обозначим символом A фактор-множество, являющееся разбиением бинарного отношения R на классы эквивалентности относительно уникальных правых частей его элементов.

То есть, если $r, t \in R$, $r = (\tau_r, f_r)$, $t = (\tau_t, f_t)$, то $r \sim\Leftrightarrow f_r = f_t$;

$$A = R/\backsim \Leftrightarrow A = \left\{\alpha_f = \left\{r = (\tau, f) : r \in R\right\} : f \in F\right\}.$$

Данное фактор-множество представляет основу структурного расслоения [3] исходного отношения R на частичные отношения. Следует отметить, что в силу своего построения число классов эквивалентности фактор-множества A равно числу всех начальных атомов решетки, т. е. $|A| = |F_{notinit}|$. *Слоем* называется выборка по одному элементу из каждого класса эквивалентности $\alpha \in A$: $l = \left\{r_1 \in \alpha_1, ..., r_{|F_{notinit}|} \in \alpha_{|F_{notinit}|}\right\}$.

Нахождение показателей релевантности основано на поиске прообразов в таких слоях. Слой либо содержит прообраз, либо его элементы образуют цикл и не могут содержать прообраз [9].

Обозначим $S(A)$: $A \to \lambda(R)$ отображение, переводящее фактор-множество $A = \left\{\alpha_1, ..., \alpha_{|F_{notinit}|}\right\}$ в множество всех слоев, которые A описывает.

$$S(A) = \left\{\left\{r_1, ..., r_{|A|}\right\} | r_i \in \alpha_i, \ i = \overline{1, |A|}\right\} =$$
$$= \bigcup_{(r_1, ..., r_{|A|}) \in \alpha_1 \times ... \times \alpha_{|A|}} \left\{r_1, ..., r_{|A|}\right\}.$$

Если фактор-множество A состоит из классов $\alpha_1, ..., \alpha_n$, то число слоев на основе A равно числу всевозможных выборок из каждого класса эквивалентности по одному элементу, т. е. $|S(A)| = |\alpha_1| \cdot ... \cdot |\alpha_n|$.

Для кластеризации множества слоев могут быть использованы свойства логической связанности пар бинарного отношения, ассоциированные с этими парами значения степеней истинности, а также оценки числа слоев без циклов, включающих в себя заданный атом решетки.

Оценка числа слоев без циклов, включающих в себя заданный атом решетки

Ранее [10] были рассмотрены вопросы оценки числа слоев без циклов, включающих в себя заданный атом решетки. Далее приведем основные определения и результаты из работы [10], а затем рассмотрим свойства монотонности для функции оценки числа слоев без циклов, необходимые для нахождения ее приближенных значений.

Отображения $in(\alpha, f) = \left\{(\tau, g) \in \alpha : f \in \tau\right\}$ и $\overline{in}(\alpha, f) = \left\{(\tau, g) \in \alpha : f \notin \tau\right\}$ переводят класс эквивалентности α в множество пар, которые соответственно содержат либо не содержат в своих левых частях атом f. Следует заметить, что $in(\alpha, f) = \alpha \backslash \overline{in}(\alpha, f)$ и $\overline{in}(\alpha, f) = \alpha \backslash in(\alpha, f)$. Функции in и $\overline{in}$ можно обобщить на все фактор-множество A:

$$in(A, f)=\left\{\bigcup_{a\in A}\{in(\alpha, f)\}\right\},\ \overline{in}(A, f)=\left\{\bigcup_{a\in A}\{\overline{in}(\alpha, f)\}\right\},$$

а также на некоторое подмножество атомов $\upsilon \subseteq F$:

$$in(\alpha,\ \upsilon)=\{(\tau,\ g)\in\alpha : \upsilon\subseteq\tau\}\ \text{ и}$$

$$\overline{in}(\alpha,\ \upsilon)=\{(\tau,\ g)\in\alpha : \upsilon\not\subseteq\tau\};$$

$$in(A,\ \upsilon)=\left\{\bigcup_{a\in A}\{in(\alpha,\ \upsilon)\}\right\},$$

$$\overline{in}(A, \upsilon)=\left\{\bigcup_{a\in A}\{\overline{in}(\alpha,\ \upsilon)\}\right\}.$$

Отображение $\lambda(X, k, i)$: $X \to \lambda(X)$, $k = \overline{1, |X|}$, $i = \overline{1, \binom{|X|}{k}}$ переводит множество X в i-й элемент множества всех его k-элементных подмножеств.

Отображение $A_f = \{\alpha \in A | \exists(\tau, g) \in \alpha: f \in \tau\} = \{\alpha \in A: \mathrm{in}(\alpha, f) \neq \varnothing\}$ переводит фактор-множество A в такое подмножество его классов эквивалентности, что каждый элемент подмножества включает в себя хотя бы одну пару, содержащую f в левой части.

На основе бинарного отношения R построим граф $G = \{\nu\}$:

$$\nu\in G,\ \nu=(f,\ g),\ f,\ g\in F \Leftrightarrow \exists(\tau,\ g)\in R : f\in\tau.$$

По графу G, используя некоторый из известных математических методов, построим фундаментальную систему циклов $H = \{h : h \subseteq G\}$. На основе отношения R и системы циклов H определим E — *фундаментальную систему циклов канонического бинарного отношения*: на основе дуг $\nu = (f, g), f, g \in F$, образующих элементы фундаментальной системы циклов $H = \{h\}$, построим множество $E = \{e_1, ..., e_{|S|}\}$, состоящее из подмножеств отношения R таких, что $\forall h = \{\nu = (f, g) \in \mathrm{G}\}$: $e_i = \{(\tau, g) \in R : f \in \tau\}$.

Пусть $\varepsilon \subseteq E$, $\varepsilon = \{e \in E\}$ — подмножество фундаментальных циклов канонического бинарного отношения R, $\dot{\varepsilon}(\varepsilon) = \bigcup_{e\in\varepsilon}\bigcup_{r\in e} r$ — объединение всех пар бинарного отношения, формирующих элементы данного подмножества. Сужение фактор-множества A, описывающее всевозможные слои, которые содержат только фундаментальные циклы из множества ε, определяется следующим отображением:

$$\theta(\alpha,\ \varepsilon)=\begin{cases}\alpha\cap\dot{\varepsilon}(\varepsilon), & \text{если } \alpha\cap\dot{\varepsilon}(\varepsilon)\neq\varnothing\\ \alpha, & \text{иначе}\end{cases},\ \varepsilon\subseteq E,$$

$$\Theta(A,\ \varepsilon)=\left\{\theta(\alpha_1\varepsilon),\ ...,\ \theta\left(\alpha_{|A|},\ \varepsilon\right)\right\},\ \varepsilon\subseteq E.$$

Определим специальные отображения $S_{all}(A, \Theta, \varepsilon)$ и $S_{subsets}(A, \Theta, \varepsilon)$.

- $S_{all}(A, \Theta, \varepsilon)$ — множество всевозможных слоев, включающих в себя все циклы из множества $\varepsilon \subseteq E$.

$$S_{all}(A,\ \Theta,\ \varepsilon)=S\left(\Theta\left(A,\ \varepsilon\right)\right),\ \ \varepsilon\subseteq E;$$

$$\left|S_{all}\left(A,\ \Theta,\ \varepsilon\right)\right|=\left|\theta\left(\alpha_1,\varepsilon\right)\right|\cdot\ldots\cdot\left|\theta\left(\alpha_{|A|},\varepsilon\right)\right|,$$

$$A=\left\{\alpha_1,\ldots,\alpha_{|A|}\right\}.$$

Отображение S_{all} можно обобщить на подмножество атомов $\upsilon \subseteq F$:

$$\vartheta(\alpha,\ \upsilon)=\left\{(\tau,\ g)\in\alpha : \upsilon\cap\tau\neq\varnothing\right\},\ \alpha\in A,\ \upsilon\subseteq F;$$

$$\theta(\alpha,\ \upsilon)=\begin{cases}\vartheta(\alpha,\ \upsilon), & \text{если } \vartheta(\alpha,\ \upsilon)\neq\varnothing\\ \alpha, & \text{иначе}\end{cases},\ \upsilon\subseteq F;$$

$$\Theta(A,\ \upsilon)=\left\{\theta\left(\alpha_1\upsilon\right),\ ...,\ \theta\left(\alpha_{|A|},\ \upsilon\right)\right\},\ \ \upsilon\subseteq F;$$

$$S_{all}(A,\ \upsilon)=S\left(\Theta\left(A,\ \upsilon\right)\right),\ \upsilon\subseteq F;$$

$$\left|S_{all}\left(A,\ \upsilon\right)\right|=\left|\theta(\alpha_1,\ \upsilon)\right|\cdot\ldots\cdot\left|\theta\left(\alpha_{|A|},\ \upsilon\right)\right|.$$

- $S_{subsets}(A, \Theta, \varepsilon)$ — множество всевозможных слоев, включающих в себя всевозможные подмножества циклов из множества $\varepsilon \subseteq E$.

В работе [10] показано, что для фундаментальной системы циклов $E = \{x_1, ..., x_l\}$ канонического бинарного отношения $S_{subsets}$ может быть выражено в следующем виде:

$$\left|S_{subsets}\left(A,\ \Theta,\ E\right)\right|=\sum_{x_1\in E}\left|S_{all}\left(A,\ \Theta,\ \{x_1\}\right)\right|-$$

$$-\sum_{\substack{\{x_1,x_2\}\subset E,\\ x_1\neq x_2}}\left|S_{all}\left(A,\ \Theta,\ \{x_1,\ x_2\}\right)\right|+$$

$$+\ldots+(-1)^{l+1}\left|S_{all}\left(A,\ \Theta,\ \{x_1,\ldots,x_l\}\right)\right|=$$

$$=\sum_{k=1}^{l}(-1)^{k+1}\sum_{i=1}^{\binom{l}{k}}\left|S_{all}\left(A,\ \Theta,\ \{\lambda(E,\ k,\ i)\}\right)\right|.$$

Определим вспомогательные отображения:

$$F(A_1,\ \Theta,\ \varepsilon)=\left\{e\in\varepsilon|\forall\alpha_e\in e/\sim\exists\alpha_{A_1}\in A_1/\sim:\alpha_e\subseteq\alpha_{A_1}\right\},$$

$$\varepsilon\subseteq E;$$

$$\tilde{A}(i,\ s,\ f)=in\left(\lambda\left(A_f,\ i,\ s\right),\ f\right)\cup\overline{in}\left(A\backslash\lambda\left(A_f,\ i,\ s\right),\ f\right),$$

$$f\in F.$$

Общее число слоев для атома f вычисляется как сумма [10]:

$$W(A,\ f)=\sum_{i=1}^{|A_f|}\sum_{s=1}^{\binom{|A_f|}{i}}\prod_{\beta\in\tilde{A}(i,s,f)}|\beta|.$$

Число слоев без циклов, включающих в себя заданный атом F, вычисляется следующим образом:

$$W_{noloops}(A, f) = \sum_{i=1}^{|A_f|} \sum_{s=1}^{\binom{|A_f|}{i}} \left(\prod_{\beta \in \tilde{A}(i,s,f)} |\beta| - \left| S_{subsets}\left(\tilde{A}(i, s, f), \Theta, F\left(\tilde{A}(i, s, f), \Theta, \varepsilon\right)\right) \right| \right).$$

Следует отметить, что число элементов в приведенных суммах растет экспоненциально относительно значения $|A_f|$.

Свойства монотонности оценок числа слоев без циклов

Рассмотрим взаимосвязь значений числа слоев для различных выборок.

Утверждение 1. Между элементами фактор-множеств $\tilde{A}(i, s, f)$ и A имеет место соотношение:

$$\forall i, s, f : \left| S\left(\tilde{A}(i, s, f)\right) \right| \leqslant |S(A)|.$$

Доказательство. Из определения отображений in и $\overline{in}$ следует, что $|in(\alpha, f)| \leqslant |\alpha| \wedge |(\alpha, f)| \leqslant |\alpha|$. Следовательно

$$\left| S\left(\tilde{A}(i, s, f)\right) \right| = \prod_{\alpha \in \lambda(A_f, i, s)} |in(\alpha, f)| \prod_{\alpha \in A \setminus \lambda(A_f, i, s)} \overline{in}|\alpha, f| \leqslant \prod_{\alpha \in A} |\alpha| = |S(A)|.$$

Утверждение 2. $\forall l, m, s_l, s_m : l \in \{1, ..., |A_f|\} \wedge m \in \{1, ..., |A_f|\} \wedge s_l \in \left\{1, ..., \binom{|A_f|}{l}\right\} \wedge s_m \in \left\{1, ..., \binom{|A_f|}{m}\right\}$, выполняется

$$\left| S\left(\tilde{A}(i, s_l, f)\right) \right| = \frac{C(A_f, l, m, s_l, s_m)}{C(A_f, m, l, s_m, s_l)} \cdot \left| S\left(\tilde{A}(m, s_m, f)\right) \right|,$$

где

$$C(X, l, m, i, j) = \begin{cases} \prod_{\alpha \in \lambda(X, l, i) \setminus \lambda(X, m, j)} \dfrac{|in(\alpha, f)|}{|\overline{in}(\alpha, f)|}, & \text{если} \quad \lambda(A_f, l, i) \setminus \lambda(A_f, m, j) \neq \varnothing \\ 1, & \text{иначе} \end{cases}.$$

Данное утверждение позволяет выразить число слоев для левой выборки через число слоев правой выборки, скорректированное соответственно изменениям в выборках классов эквивалентности.

Доказательство. Зафиксируем некоторые $\{\alpha_1, \alpha_2\} \subseteq A_f$, $i < |A_f|$, $j < |A_f|$, $s_1 \in \left\{1, ..., \binom{|A_f|}{i}\right\}$, $s_2 \in \left\{1, ..., \binom{|A_f|}{j}\right\}$ такие, что $\lambda(A_f, i, s_1) \setminus \lambda(A_f, j, s_2) = \{\alpha_1\} \wedge \lambda(A_f, j, s_2) \setminus \lambda(A_f, i, s_1) = \{\alpha_2\}$. Тогда:

$$\left| S\left(in\left(\lambda\left(A_f, j, s_2\right), f\right)\right) \right| = \frac{|in(\alpha_2, f)|}{|in(\alpha_1, f)|} \cdot \left| S\left(in\left(\lambda\left(A_f, i, s_1\right), f\right)\right) \right|,$$

$$\left| S\left(\overline{in}\left(A \setminus \lambda\left(A_f, j, s_2\right), f\right)\right) \right| = \frac{|\overline{in}(\alpha_1, f)|}{|\overline{in}(\alpha_2, f)|} \cdot \left| S\left(\overline{in}\left(A \setminus \lambda\left(A_f, i, s_1\right), f\right)\right) \right|.$$

Следовательно

$$\left| S\left(\tilde{A}(j, s_2, f)\right) \right| = \frac{|in(\alpha_2, f)|}{|\overline{in}(\alpha_2, f)|} \cdot \frac{|\overline{in}(\alpha_1, f)|}{|in(\alpha_1, f)|} \cdot \left| S\left(\tilde{A}(i, s_1, f)\right) \right|.$$

Рассмотрим теперь другую ситуацию. Пусть заданы $X \subseteq A_f$, $Y \subseteq A_f$, $i < |A_f|$, $j < |A_f|$, $s_1 \in \left\{1, ..., \binom{|A_f|}{i}\right\}$, $s_2 \in \left\{1, ..., \binom{|A_f|}{j}\right\}$ такие, что $\lambda(A_f, i, s_1) \setminus \lambda(A_f, j, s_2) = X \wedge \lambda(A_f, j, s_2) \setminus \lambda(A_f, i, s_1) = Y$. Предположим также, что выполняется равенство

$$\left|S\left(\tilde{A}(j, s_2, f)\right)\right| = \prod_{\alpha\in Y}\frac{|in(\alpha, f)|}{|\overline{in}(\alpha, f)|}\cdot\prod_{\alpha\in X}\frac{|\overline{in}(\alpha, f)|}{|in(\alpha, f)|}\cdot\left|S\left(\tilde{A}(i, s_1, f)\right)\right|.$$

Зафиксируем некоторый класс эквивалентности $\beta\in A_f$, $\beta\neq\varnothing$ такой, что: $\beta\notin X\wedge\beta\notin Y$. В силу определения отображений S и $\tilde{A}$, имеет место факт

$$\left|S\left(\tilde{A}(j, s_2, f)\right)\right| = \left|in\left(\lambda\left(A_f, j, s_2\right), f\right)\right|\cdot\left|\overline{in}\left(A\backslash\lambda\left(A_f, j, s_2\right), f\right)\right|.$$

Кроме того,

$$\left|S\left(in\left(\lambda\left(A_f, j, s_2\right)\cup\{\beta\}, f\right)\right)\right|\cdot\left|S\left(\overline{in}\left(A\backslash\left(\lambda\left(A_f, j, s_2\right)\cup\{\beta\}\right), f\right)\right)\right| =$$

$$= \frac{|in(\beta, f)|}{|\overline{in}(\beta, f)|}\cdot\left|S\left(in\left(\lambda\left(A_f, j, s_2\right), f\right)\right)\right|\cdot\left|S\left(\overline{in}\left(A\backslash\lambda\left(A_f, j, s_2\right), f\right)\right)\right| =$$

$$= \frac{|in(\beta, f)|}{|\overline{in}(\beta, f)|}\cdot\left|S\left(\tilde{A}(j, s_2, f)\right)\right| = \frac{|in(\beta, f)|}{|\overline{in}(\beta, f)|}\cdot\prod_{\alpha\in Y}\frac{|in(\alpha, f)|}{|\overline{in}(\alpha, f)|}\cdot\prod_{\alpha\in X}\frac{|\overline{in}(\alpha, f)|}{|in(\alpha, f)|}\cdot\left|S\left(\tilde{A}(i, s_1, f)\right)\right|.$$

Очевидно, для $\lambda\left(A_f, j, s_2\right)\cup\{\beta\}$ есть соответствующая выборка классов эквивалентности, т. е. $\exists k,\ s_3$: $\lambda\left(A_f, j, s_3\right) = \lambda\left(A_f, j, s_2\right)\cup\{\beta\}$. Таким образом, индуктивное предположение доказано в силу произвольности X, Y и β.

Рассмотрим, как зависят возрастание или убывание общего числа слоев от числа элементов в выборках, т. е. как $W(A, f, i)$ зависит от изменения i.

По определению $W(A, f, i) = \sum_{s=1}^{\binom{|A_f|}{i}}\left|S\left(\tilde{A}(i, s, f)\right)\right|$.

Биномиальный коэффициент $\binom{|A_f|}{i}$ возрастает при $i\in\left[1, \left[\frac{|A_f|}{2}\right]\right]$ и убывает при $i\in\left[\left[\frac{|A_f|+1}{2}\right], |A_f|\right]$. Если $|A_f|$ — четное, то $\left[\frac{|A_f|}{2}\right] = \left[\left(|A_f|+1\right)/2\right]$. Если же $|A_f|$ — нечетное, то равны соответствующие биномиальные коэффициенты. То есть $\forall i = \overline{\left[|A_f|/2\right]+1, |A_f|}: \binom{|A_f|}{i}\leqslant\binom{|A_f|}{i-1}$.

В силу утверждения 2, для произвольных допустимых i, k, s выполняется

$$\left|S\left(\tilde{A}(i, k, f)\right)\right| = \frac{C\left(A_f, i, i-1, k, s\right)}{C\left(A_f, i-1, i, s, k\right)}\left|S\left(\tilde{A}(i-1, s, f)\right)\right|.$$

Если $\lambda(A_f, i-1, s)\subset\lambda(A_f, i, k)$, то $C(A_f, i-1, i, s, k) = 1$, а $C\left(A_f, i, i-1, k, s\right) = \frac{|in(\alpha, f)|}{|\overline{in}(\alpha, f)|}$, где $\{\alpha\} = \lambda\left(A_f, i, k\right)\backslash\lambda\left(A_f, i-1, s\right)$. Если при этом число элементов класса эквивалентности α, содержащих в своей левой части атом f, меньше, чем число элементов, не содержащих заданный атом, то будет выполняться соотношение $\left|S\left(\tilde{A}(i, k, f)\right)\right| < \left|S\left(\tilde{A}(i-1, s, f)\right)\right|$. Такая ситуация может встречаться довольно часто при больших значениях $|R|$ и больших мощностях классов эквивалентности $|\alpha|\in A = R/\sim$. В интеллектуальных системах такая ситуация соответствует базам знаний с большим числом правил.

Исходя из полученных соотношений, можно сделать следующие эвристические предположения:

1) при $\left[|A_f|/2\right]+1\leqslant i\leqslant|A_f|$ часто будет выполняться соотношение $W(A, f, i)\leqslant W(A, f, i-1)$;

2) при $2\leqslant i\leqslant|A_f|$ часто будут выполняться соотношения

$$\binom{|A_f|}{i-1}\max_{s=\binom{|A_f|}{i}}\left|S\left(\tilde{A}(i,\,s,\,f)\right)\right| \leqslant W(A,\,f,\,i-1);$$

$$W(A,\,f,\,i) \leqslant \binom{|A_f|}{i}\min_{s=\binom{|A_f|}{i-1}}\left|S\left(\tilde{A}(i-1,\,s,\,f)\right)\right|.$$

Далее рассмотрим зависимость числа элементов в множестве $\theta(\alpha,\varepsilon\cup\{e\})$ от структуры класса эквивалентности α, подмножества фундаментальных циклов $\varepsilon\subseteq E$ и некоторого фундаментального цикла $e\in E$, $e\notin\varepsilon$. Имеем

$$\begin{cases}\alpha\cap\dot{\varepsilon}(\{e\})=\varnothing\\ \alpha\cap\dot{\varepsilon}(\varepsilon)=\varnothing\end{cases}\to\begin{cases}\theta(\alpha,\,\{e\})=|\alpha|\\ \theta(\alpha,\,\varepsilon)=|\alpha|\end{cases}\to|\theta(\alpha,\varepsilon)|=|\theta(\alpha,\varepsilon\cup\{e\})|;$$

$$\begin{cases}\alpha\cap\dot{\varepsilon}(\{e\})=\varnothing\\ \alpha\cap\dot{\varepsilon}(\varepsilon)\neq\varnothing\end{cases}\to\begin{cases}\theta(\alpha,\,\{e\})=|\alpha|\\ \theta(\alpha,\,\varepsilon)=|\alpha\cap\dot{\varepsilon}(\varepsilon)|\end{cases}\to|\theta(\alpha,\,\{e\}\cup\varepsilon)|=|\theta(\alpha,\,\varepsilon)|;$$

$$\begin{cases}\alpha\cap\dot{\varepsilon}(\{e\})\neq\varnothing\\ \alpha\cap\dot{\varepsilon}(\varepsilon)=\varnothing\end{cases}\to\begin{cases}\theta(\alpha,\,\{e\})=|\alpha\cap\dot{\varepsilon}(\{e\})|\\ \theta(\alpha,\,\varepsilon)=|\alpha|\end{cases}\to|\theta(\alpha,\,\{e\}\cup\varepsilon)|<|\theta(\alpha,\,\varepsilon)|;$$

$$\begin{cases}\alpha\cap\dot{\varepsilon}(\{e\})\neq\varnothing\\ \alpha\cap\dot{\varepsilon}(\varepsilon)\neq\varnothing\\ \alpha\cap\dot{\varepsilon}(\{e\})\subseteq\alpha\cap\dot{\varepsilon}(\varepsilon)\end{cases}\to\begin{cases}\theta(\alpha,\,\{e\})=|\alpha\cap\dot{\varepsilon}(\{e\})|\\ \theta(\alpha,\,\varepsilon)=|\alpha\cap\dot{\varepsilon}(\varepsilon)|\end{cases}\to|\theta(\alpha,\,\{e\}\cup\varepsilon)|=|\theta(\alpha,\,\varepsilon)|;$$

$$\begin{cases}\alpha\cap\dot{\varepsilon}(\{e\})\neq\varnothing\\ \alpha\cap\dot{\varepsilon}(\varepsilon)\neq\varnothing\\ \alpha\cap\dot{\varepsilon}(\{e\})\nsubseteq\alpha\cap\dot{\varepsilon}(\varepsilon)\end{cases}\to\begin{cases}\theta(\alpha,\,\{e\})=|\alpha\cap\dot{\varepsilon}(\{e\})|\\ \theta(\alpha,\,\varepsilon)=|\alpha\cap\dot{\varepsilon}(\varepsilon)|\end{cases}\to|\theta(\alpha,\,\{e\}\cup\varepsilon)|>|\theta(\alpha,\,\varepsilon)|.$$

Таким образом, при добавлении нового цикла $e\in E$ к подмножеству $\varepsilon\subseteq E$ значение $|\,S_{subsets}\left(\tilde{A}(i,\,s,\,f),\,\Theta,\,F\left(\tilde{A}(i,\,s,\,f),\,\Theta,\,\varepsilon\right)\right)$ — число слоев, которые можно построить по сужению $\Theta(A,\,\varepsilon)$, будет изменяться следующим образом:

- **убывать**, если $\forall\alpha\in A:\alpha\cap\dot{\varepsilon}(\varepsilon)=\varnothing$ и $\exists\alpha\in A:\alpha\cap\dot{\varepsilon}(\{e\})\neq\varnothing$;
- **не возрастать**, если $\forall\alpha\in A:\alpha\cap\dot{\varepsilon}(\{e\})=\varnothing\vee\alpha\cap\dot{\varepsilon}(\varepsilon)=\varnothing\vee\alpha\cap\dot{\varepsilon}(\{e\})\subseteq\alpha\cap\dot{\varepsilon}(\varepsilon)$;
- **не убывать**, если $\forall\alpha\in A:\alpha\cap\dot{\varepsilon}(\{e\})=\varnothing\vee\alpha\cap\dot{\varepsilon}(\varepsilon)\neq\varnothing$.

Исходя из полученных соотношений можно сделать эвристическое предположение, что наибольшие значения $|S_{all}(A,\,\Theta,\,\varepsilon)|$ достигаются при $|\varepsilon|\in\{1,|E|\}$. Таким образом, для приближенного подсчета $|S_{subsets}(A,\,\Theta,\,\varepsilon)|$ можно просуммировать некоторое число начальных и конечных компонент формулы, а приближенное значение суммы внутренних компонент найти, например, путем интегрирования методом Монте-Карло [11]. Выборку для этого метода можно строить по подпоследовательностям циклов, порождающих неубывающую последовательность числа слоев.

Кластеризация множества слоев

Ниже предложен способ эвристической кластеризации виртуального расслоения бинарного отношения, на основании которого будет построен алгоритм стохастического поиска прообразов.

Для бинарного отношения R построим граф $G=\{\nu\}$:

$$\nu\in G,\ \nu=(f,\,g),\ f,\,g\in F\Leftrightarrow\exists(\tau,\,g)\in R:f\in\tau.$$

Далее, используя один из известных математических методов кластеризации графов [12], разобьем вершины графа G на m кластеров: $G_i=\{f\in F\}$, $i=\overline{1,m}$.

Для каждого атома решетки вычислим приближенную оценку отношения числа слоев без циклов к общему числу слоев:

$$\omega(f)\approx\frac{W_{noloops}(A,\,f)}{W(A,\,f)},\ f\in F,$$

а также медиану соответствующих степеней истинности:

$$\mathrm{med}_\mu(f)=$$

$$=\begin{cases}\mathrm{median}\left(\{M_R(r)|r=(\tau,\,f)\in R\}\right), & \text{если } f\in F_{notinit}\\ \mathrm{median}\left(M_{init}(f)\right), & \text{иначе}\end{cases}.$$

Таким образом, для каждого атома $f \in F$ можно построить вектор из трех числовых значений: $\nu = \{i, \omega(f), \text{med}_{\mu}(f)\}$, где i — номер кластера графа G. Обозначим множество таких векторов как $\mathbf{V} = \{\nu_1, ..., \nu_{|F|}\}$.

Используя один из методов кластеризации [13, 14], выполним кластеризацию вектора **V** и обозначим полученные кластеры символами V_j, $j = \overline{1, n}$, где n — число кластеров. Обозначим $f_{i,k}$ атом, соответствующий k-му элементу кластера V_j. Для каждого кластера определим значение вероятности:

$$l_j = \sum_{k=1}^{|V_j|} \sqrt{\omega(f_{j,k}), \text{med}_{\mu}(f_k)}, \quad p_j = \frac{l_j}{\sum_{k=1}^{n} l_k}.$$

На основе введенных обозначений можно построить алгоритм кластеризации и стохастического поиска прообразов в виртуальном расслоении бинарного отношения. Он принимает на вход четыре ограничения: C_t — максимальное время работы; C_i — максимальное число итераций; C_{ρ} — максимальное число найденных уникальных прообразов; C_W — максимальное число слоев в кластере для анализа. Возвращаемое значение алгоритма — множество найденных прообразов.

В ходе работы алгоритма подразумевается использование нескольких внешних функций:

- rand(X) = $x \in X$ — генератор случайного элемента x множества $X = \{x_1, ..., x_m\}$ по равномерному распределению вероятностей;
- rand(X, P) = $x \in X$ — генератор случайного элемента x множества $X = \{x_1, ..., x_m\}$ по распределению вероятностей $P = \{p_1, ..., p_m\}$;
- minPreImages(v) —функция, выполняющая поиск прообразов на множестве слоев $S_{all}(A, v)$. Спецификация ее работы аналогична одноименной функции в методе LP-вывода [4].

Далее приведен псевдокод алгоритма стохастического поиска прообразов.

input: C_t, C_i, C_{ρ}, C_W
output: ρ
$i \leftarrow 0$, $t \leftarrow 0$, $\rho \leftarrow \varnothing$
while $t < C_t \wedge i < C_i \wedge |\rho| < C_{\rho}$ **do**
  $v \leftarrow \varnothing$
  while $v = \varnothing \vee |S_{all}(A, v)| > C_W$ **do**
    $j \leftarrow \text{rand}(\{1, ... n\}, \{p_1, ... p_n\})$
    $\gamma \leftarrow \{f_{j,1}, ..., f_{j,|V_j|}\}$
    while $|\gamma| > 0 \vee |S_{all}(A, v)| > C_W$ **do**
      $f \leftarrow \text{rand}(\gamma)$
      $\gamma \leftarrow \gamma \setminus \{f\}$
      $v \leftarrow v \cup \{f\}$
    end
  end
  $\rho \leftarrow \rho \cup \text{minPreImages}(v)$
  $t \leftarrow \Delta t$
  $i \leftarrow i + 1$
end.

Заключение

Рассмотрена математическая формулировка задачи нечеткого LP-вывода. Исследованы свойства монотонности функции оценки числа слоев без циклов. Построен алгоритм кластеризации и стохастического поиска прообразов в виртуальном расслоении бинарного отношения. Полученные результаты могут быть применены для ускорения обратного вывода в интеллектуальных системах продукционного типа. Они могут быть распространены и на более сложные логические системы в информатике [15].

Список литературы

1. **Джарратано Д., Райли Г.** Экспертные системы: принципы разработки и программирование: пер. с англ. М.: Вильямс, 2007. 1152 с.
2. **Махортов С. Д., Шмарин А. Н.** Оптимизация метода LP-вывода// Нейрокомпьютеры. Разработка, применение. 2013. № 9. С. 59—63.
3. **Махортов С. Д.** Основанный на решетках подход к исследованию и оптимизации множества правил условной системы переписывания термов // Интеллектуальные системы. Теория и приложения. 2009. Т. 13. С. 51—68.
4. **Махортов С. Д.** Интегрированная среда логического программирования LPExpert // Информационные технологии. 2009. № 12. С. 65—66.
5. **Болотова С. Ю., Махортов С. Д.** Алгоритмы релевантного обратного вывода, основанные на решении продукционно-логических уравнений // Искусственный интеллект и принятие решений. 2011. № 2. С. 40—50.
6. **Шмарин А. Н., Махортов С. Д.** О приближенных оценках количества слоев без циклов в задаче нечеткого LP-вывода // Нейрокомпьютеры. Разработка, применение. 2015. № 12. С. 44—51.
7. **Салий В. Н., Богомолов В. А.** Алгебраические основы теории дискретных систем. М.: Физматлит, 1997. 368 с.
8. **Пегат А.** Нечеткое моделирование и управление. М.: Бином. Лаборатория знаний, 2013. 798 с.
9. **Махортов С. Д.** LP-структуры для обоснования и автоматизации рефакторинга в объектно-ориентированном программировании // Программная инженерия. 2010. № 2. С. 15—21.
10. **Шмарин А. Н.** О реализации приближения числа слоев без циклов в задаче нечеткого LP-вывода // Программная инженерия. 2016. № 7. С. 330—336.
11. **Ермаков С. М.** Метод Монте-Карло в вычислительной математике. СПб.: Невский Диалект, 2009. 192 с.
12. **Schaeffer S. E.** Graph clustering // Computer Science Review. 2007. Vol. I. P. 27—64.
13. **Kanunfo T., Mount D. M., Netanyahu N. S.** An Efficient k-Means Clustering Algorithm: Analysis and Implementation // IEEE Transactions on pattern analysis and machine intelligence. 2002. Vol. 24, No. 7. P. 881—892.
14. **Patwary M. M., Palsetia D. A.** New Scalable Parallel DBSCAN Algorithm Using the Disjoint-Set Data Structure // IEEE. 2012. Vol. 12. P. 10—16.
15. **Чечкин А. В.** Нейрокомпьютерная парадигма информатики // Нейрокомпьютеры: разработка, применение. 2011. № 7. С. 3—9.

Clustering the Set of Layers in the Fuzzy LP-Inference Problem

A. N. Shmarin, e-mail: tim-shr@mail.ru, Voronezh State University, Voronezh, 394006, Russian Federation

Corresponding author:

Shmarin Artem N., Postgraduate Student, Voronezh State University, Voronezh, 394006, Russian Federation, E-mail: tim-shr@mail.ru

Received on December 25, 2016
Accepted on January 11, 2017

The increase in the volumes and complexity of processed information makes the use of artificial intelligence systems more and more actual. In such systems, one of the most widespread knowledge representation models is the production model, and search algorithms are based on the inference engine. In practice, the values measurement of the features of the classified objects from some data domain is a time-consuming operation for many problems. These tasks appear, for example, when the robot explores the surface of another planet. If the goal of the robot is to get to some rock, but the future route is observed only partially, then the robot should try to make the best decision on how to achieve the goal, taking into account limitation of the resources available to additional exploration of terrain. Another example, in which the cost of obtaining new information may be high, is the commercial medicine. To minimize costs, it is necessary to find the acceptable treatment method using the minimum number of analyses performed in minimal time. The delay in providing treatment, caused by the additional analyses, leads to increased costs. Moreover, the condition of a patient who was not provided timely assistance, can significantly deteriorate. Also, the cost of additional research is essential in the field of mineral exploration. It can turn out that a more cost-effective solution is to begin drilling, if confidence of the success is 95 %, than to spend the considerable resources to achieve the 98 % confidence.

The complexity of the operations of data acquisition necessary for decision-making, leads to the problem of minimizing the number of requests for information about values of the features of the classified object during inference. This problem is NP-hard. To achieve global minimization there is the general method of LP-inference with exponential computational complexity relative to the number of atomic facts in the knowledge base. However, some of its heuristic modifications have polynomial complexity. The goal of the provided research is development of the approximating method to better minimize the number of feature values requests performed in the inference.

Earlier, the questions of implementation of computing approximate estimates of the number of layers without cycles in the fuzzy LP-inference task were considered. This paper presents the mathematical statement of the fuzzy LP-inference task, researches the monotonicity properties of the function evaluating the number of layers without cycles, and presents the clustering algorithm of the set of layers — singleton samples from equivalence classes of the quotient set, that obtained by partitioning of the binary relation by the unique right parts of pairs. This algorithm is designed for the iterative analysis of such subsets of productions, which most significantly influence to the relevancy. The presented approach can be used to accelerate the reverse inference in production type systems of artificial intelligence.

Keywords: *LP-inference, binary relation, quotient set, clustering, stochastic algorithm, machine learning, algorithms*

For citation:

Shmarin A. N. Clustering the Set of Layers in the Fuzzy LP-Inference Problem, *Programmnaya Ingeneria,* 2017, vol. 8, no. 4, pp. 177—185.

DOI: 10.17587/prin.8.177-185

References

1. **Dzharratano D., Rajli G.** *Jekspertnye sistemy: principy razrabotki i programmirovanie* (Expert systems: design principles and programming), Moscow, Wil'ams, 2007, 1152 p. (in Russian).
2. **Makhortov S. D., Shmarin A. N.** Optimizacija metoda LP-vyvoda (Optimizing LP-inference method), *Nejrokomp'jutery. Razrabotka, primenenie*, 2013, no 9, pp. 59—63 (in Russian).
3. **Makhortov S. D.** Osnovannyj na reshetkah podhod k issledovaniju i optimizacii mnozhestva pravil uslovnoj sistemy perepisyvanija termov (Research and optimization of a set of rules in a conditional term rewriting system using approach based on the lattices), *Intellektual'nye sistemy. Teorija i prilozhenija*, 2009, Vol. 13, pp. 51—68 (in Russian).
4. **Makhortov S. D.** Integrirovannaja sreda logicheskogo programmirovanija LPExpert (LPExpert: Integrated development environment for logical programming), *Informacionnye tehnologii*, 2009, no. 12, pp. 65—66 (in Russian).
5. **Bolotova S. Ju., Makhortov S. D.** Algoritmy relevantnogo obratnogo vyvoda, osnovannye na reshenii produkcionno-logicheskih uravnenij (Algorithms of the relevant backward inference on production-logic equations solving), *Iskusstvennyj intellekt i prinjatie reshenij*, 2011, no. 2, pp. 40—50 (in Russian).
6. **Makhortov S. D., Shmarin A. N.** O priblizhennyh ocenkah kolichestva sloev bez ciklov v zadache nechetkogo LP-vyvoda (About approximate estimates of the number of layers without cycles in the fuzzy LP-inference), *Nejrokomp'jutery. Razrabotka, primenenie*, 2015, no. 12, pp. 44—51 (in Russian).
7. **Salij V. N., Bogomolov V. A.** *Algebraicheskie osnovy teorii diskretnyh sistem* (Algebraic theory of discrete systems), Moscow, Fizmatlit, 1997, 368 p. (in Russian).
8. **Pegat A.** *Nechetkoe modelirovanie i upravlenie* (Fuzzy Modeling and Controling), Moscow, Binom, Laboratorija znanij, 2013, 798 p. (in Russian).
9. **Makhortov S. D.** LP-struktury dlja obosnovanija i avtomatizacii refaktoringa v ob#ektno-orientirovannom programmirovanii (LP-structure for validation and automation of the refactoring in object-oriented programming), *Programmnaya Ingeneria*, 2010, no 2, pp. 15—21 (in Russian).
10. **Shmarin A. N.** O realizacii priblizhenija chisla sloev bez ciklov v zadache nechetkogo LP-vyvoda (About Implementation of Approximation of Layers without Cycles in the Fuzzy LP-Inference Problem), *Programmnaya Ingeneria*, 2016, vol. 7, no. 7, pp. 330—336 (in Russian).
11. **Ermakov S. M.** *Metod Monte-Karlo v vychislitel'noj matematike* (Monte Carlo method in computational mathematics), St. Petersburg, Nevskij Dialekt, 2009, 192 p. (in Russian).
12. **Schaeffer S. E.** Graph clustering, *Computer Science Review,* 2007, vol. I, pp. 27—64.
13. **Kanunfo T., Mount D. M., Netanyahu N. S.** An Efficient k-Means Clustering Algorithm: Analysis and Implementation*, IEEE Transactions on pattern analysis and machine intelligence,* 2002, vol. 24, no. 7, pp. 881—892.
14. **Patwary M. M., Palsetia D. A.** New Scalable Parallel DBSCAN Algorithm Using the Disjoint-Set Data Structure, *IEEE,* 2012, vol. 12, pp. 10—16.
15. **Chechkin A. V.** Nejrokomp'juternaja paradigma informatiki (Neurocomputing paradigm of informatics), *Nejrokomp'jutery: Razrabotka, Primenenie*, 2011, no. 7, pp. 3—9 (in Russian).