

О сведении нечеткого информационного поиска к информационному поиску большей размерности

А. А. Харина

В статье предлагается некоторое обобщение понятия задачи информационного поиска, получаемое заменой отношения поиска на нечеткое множество; исследуется возможность решения задач с нечеткой логикой сведением к задачам информационного поиска большей размерности. Также приводится пример решения задачи с нечеткой логикой алгоритмом для задачи информационного поиска.

1. Введение

В статье предлагается некоторое обобщение задачи информационного поиска (ЗИП), суть которого состоит в замене отношения поиска на нечеткое множество. Полученные таким образом задачи названы задачами нечеткого поиска (ЗНП), для них формализована модель поиска решения с использованием нечетких информационных графов (НИГ), аналогичная информационно-графовой модели [1] для ЗИП.

Приводятся результаты исследования, в каких случаях решение ЗНП может быть сведено к решению ЗИП заменой характеристических функций нечетких множеств на предикаты. Как и следовало ожидать, такая замена возможна только для очень узкого класса задач.

Понятие ЗНП является обобщением понятия ЗИП, но, в то же время, показано, что ЗНП в пространстве размерности N является частным случаем ЗИП в пространстве размерности $N + 1$.

В статье приводится пример решения одной ЗНП сведением к задаче информационного поиска большей размерности.

Автор выражает благодарность Э. Э. Гасанову за постановку задачи и помощь в работе.

2. Основные понятия

Задачей информационного поиска (ЗИП) назовем тройку $I = \langle X, V, \rho \rangle$. Здесь X — множество запросов, V — некоторое конечное подмножество множества записей (объектов поиска) Y , ρ — бинарное отношение на $X \times Y$, называемое *отношением поиска*. Содержательно ЗИП $I = \langle X, V, \rho \rangle$ состоит в перечислении для произвольно взятого запроса $x \in X$ всех тех и только тех записей $y \in V$, для которых $x\rho y$.

Рассмотрим некоторое обобщение понятия ЗИП.

Пусть X — множество запросов, Y — множество записей. Пусть задано отображение $\eta(x, y) : X \times Y \rightarrow [0, 1]$, которое будем называть *отношением нечеткого поиска*. Тройку $I = \langle X, V, \eta \rangle$, где V — некоторое конечное подмножество множества Y , будем называть *задачей нечеткого поиска* (ЗНП). Будем считать, что содержательно ЗНП $I = \langle X, V, \eta \rangle$ состоит в том, чтобы для произвольного числа $c \in (0, 1]$ и произвольного запроса $x \in X$ перечислить все те и только те записи $y \in V$, такие что $\eta(x, y) \geq c$.

Особо отметим, что понятие ЗИП, введенное выше, является частным случаем понятия ЗНП. Действительно, ЗИП $I = \langle X, Y, \rho \rangle$ можно рассматривать как ЗНП $I = \langle X, Y, \rho(x, y) \rangle$ с отношением нечеткого поиска

$$\rho(x, y) = \begin{cases} 1, & x\rho y, \\ 0, & x\not\rho y. \end{cases}$$

Для описания алгоритмов решения определенных таким образом ЗИП и ЗНП удобно использовать информационные графы, предложенные в [1].

Пусть $\mathcal{F} = \{f | f : X \rightarrow [0, 1]\}$ — базовое множество нечетких функций на X .

Нечеткий информационный граф (НИГ) над базовым множеством \mathcal{F} определим следующим образом. Возьмем конечную многополосную ориентированную сеть. В ней выберем некоторый полюс, который будем называть *корнем*. Остальные полюса будем называть *листьями*, поставим им в соответствие записи из Y , причем разным листьям могут соответствовать одинаковые записи. Ребрам поставим в соответствие функции из множества \mathcal{F} . Нагруженную таким образом многополосную ориентированную сеть будем называть НИГ над базовым множеством \mathcal{F} .

Понятие *информационного графа* (ИГ) определяется аналогичным образом, только вместо базового множества нечетких функций \mathcal{F} используется *базовое множество предикатов* F .

Функционирование НИГ определяется следующим образом. Проводимостью ребра НИГ назовем функцию, равную функции, приписанной этому ребру; проводимостью ориентированной цепочки ребер НИГ назовем функцию, равную минимуму проводимостей ребер цепочки; функцией фильтра вершины β НИГ (обозначается $\varphi_\beta(x)$), назовем функцию, равную максимуму функций проводимости ориентированных цепочек, ведущих из корня НИГ в вершину β ; скажем, что запись y , приписанная листу α , попадает в ответ НИГ на запрос $x \in X$ для числа $c \in (0, 1]$, если $\varphi_\alpha(x) \geq c$. *Ответом* НИГ U на запрос x для числа $c \in (0, 1]$ назовем множество записей, попавших в ответ НИГ на запрос x для числа $c \in (0, 1]$, и обозначим его $\mathcal{J}_U(x, c)$. Эту функцию $\mathcal{J}_U(x, c)$ будем считать результатом функционирования НИГ U .

Каждому НИГ U сопоставляется некий алгоритм A_U , входными данными которого является запрос, а выходными — множество записей. Будем считать, что на каждом шаге алгоритм выделяет в сети U подмножество вершин, которое назовем рабочим множеством, причем рабочие множества разных шагов не будут пересекаться. Рабочее множество первого шага состоит из одной единственной вершины — корня.

Пусть на вход алгоритма A_U поступил запрос x . На каждом шаге, начиная с первого, алгоритм просматривает вершины из рабочего множества и для каждой вершины вычисляет значения всех функций, соответствующих ребрам, исходящим из вершины.

Тем самым вычисляются проводимости всех ребер, исходящих из рабочих вершин. Теперь каждая вершина, в которую ведет ребро, проводимость которого оказалась не меньше c , включается в рабочее множество следующего шага, если только она не являлась рабочей на предыдущих шагах. Через конечное число шагов мы просмотрим все вершины, функции фильтра которых не меньше c на запросе x , то есть эти вершины на каком-то из шагов побывают в рабочем множестве алгоритма. Если среди этих вершин есть листья, то записи, соответствующие этим листьям, включаются в выходные данные алгоритма.

Отметим, что множество записей, получаемое на выходе алгоритма A_U , совпадает с множеством $\mathcal{J}_U(x, c)$, определяемым функцией ответа НИГ U .

Функционирование ИГ определяется аналогичным образом, при этом параметр c предполагается равным единице и опускается.

Пусть на множестве X задано вероятностное пространство $\langle X, \sigma, \mathbf{P} \rangle$. Пусть на отрезке $[0, 1]$ также определено вероятностное пространство: $\langle [0, 1], \sigma', \mathbf{P}' \rangle$, где σ' — алгебра подмножеств отрезка $[0, 1]$, \mathbf{P}' — вероятностная мера на σ' .

Вероятностные пространства, заданные на X и на $[0, 1]$, порождают вероятностное пространство $\langle X \times [0, 1], \sigma'', \mathbf{P}'' \rangle$ на $X \times [0, 1]$, где σ'' можно определить как $\sigma \times \sigma'$, а в качестве \mathbf{P}'' взять $\mathbf{P} \times \mathbf{P}'$, то есть для любого множества $A \times B \subset X \times [0, 1]$, где $A \subset X$, $B \subset [0, 1]$, $\mathbf{P}''(A \times B) = \mathbf{P}(A)\mathbf{P}'(B)$.

Назовем сложностью НИГ U на запросе x с параметром c величину $T(U, x, c)$, равную числу функций, вычисленных в процессе функционирования НИГ U на запросе x с параметром c , то есть

$$T(U, x, c) = \sum_{\varphi_\beta(x) \geq c} \psi_\beta.$$

Здесь сумма берется по всем вершинам β НИГ U , для которых $\varphi_\beta(x) \geq c$, ψ_β — полустепень исхода вершины β .

Будем предполагать, что величина $T(U, x, c)$ σ'' -измерима.

Определим сложность НИГ U как среднюю сложность на запросах $x' = (x, c)$:

$$T(U) = M_{x,c}T(U, x, c).$$

Скажем, что НИГ U *допустим* для ЗНП $I = \langle X, V, \eta \rangle$, если для любого запроса $x \in X$ и любого числа $c \in (0, 1]$ выполняется $\mathcal{J}_U(x, c) = \{y \in V : \eta(x, y) \geq c\}$.

Множество НИГ над базовым множеством \mathcal{F} , допустимых для ЗНП I , обозначим $\mathcal{U}(I, \mathcal{F})$.

Назовем сложностью ЗНП I при базовом множестве нечетких функций \mathcal{F} величину

$$T(I, \mathcal{F}) = \inf\{T(U) : U \in \mathcal{U}(I, \mathcal{F})\}.$$

Аналогичным образом определяется сложность $T(U, x)$ для ИГ U на запросе x , сложность $T(U)$ ИГ U , сложность $T(I, F)$ ЗИП I при базовом множестве предикатов F , но вместо базового множества нечетких функций рассматривается базовое множество предикатов F , а параметр c предполагается равным единице и опускается.

Если $y \in Y$, то множество $O(y, \eta) = \{x \in X : \eta(x, y) > 0\}$ назовем *тенью записи* y .

Пусть U — некоторый НИГ, y — запись из Y . Через $L_U(y)$ обозначим множество листьев НИГ U , которым соответствует запись y .

Пусть f — некоторая функция, определенная на X . Будем обозначать $N_f = \{x \in X : f(x) > 0\}$.

Справедлива следующая теорема [2].

Теорема 1. *НИГ U допустим для ЗНП $I = \langle X, V, \eta \rangle$ тогда и только тогда, когда*

- 1) *для любой записи $y \in V$ такой, что $O(y, \eta) \neq \emptyset$, выполняется: $L_U(y) \neq \emptyset$, и $\max_{\alpha \in L_U(y)} \varphi_\alpha(x) \equiv \eta(x, y)$;*
- 2) *для любой записи $y \in V$ такой, что $O(y, \eta) = \emptyset$, выполняется: либо $L_U(y) = \emptyset$, либо $\max_{\alpha \in L_U(y)} \varphi_\alpha(x) \equiv \eta(x, y) \equiv 0$.*

Базовое множество предикатов F можно рассматривать как базовое множество нечетких функций F , а любой ИГ U над F — как НИГ U над F . В таком случае для любого $x \in X$ для любого $c \in (0, 1]$ выполняется равенство $\{y \in V : \rho(x, y) \geq c\} = \{y \in V : x\rho y\}$, следовательно для ЗИП $I = \langle X, Y, \rho \rangle$ $\mathcal{J}_U(x) \equiv \mathcal{J}_U(x, c)$ для любого $c \in (0, 1]$.

Поэтому ИГ U допустим для ЗИП $I = \langle X, Y, \rho \rangle$ тогда и только тогда, когда НИГ U допустим для ЗНП $I = \langle X, Y, \rho(x, y) \rangle$. Поэтому теорема 1 справедлива для ИГ.

3. Сведение задачи нечеткого информационного поиска к задаче информационного поиска с сохранением размерности пространства

Будем считать, что множество запросов $X = [0, 1]$, и множество записей $Y = [0, 1]$.

Введем некоторый способ перехода от задачи информационного поиска к задаче нечеткого поиска, такой, что для замкнутого ρ и числа $c = 1$ задача нечеткого поиска совпадает с исходной задачей информационного поиска.

Рассмотрим некоторую функцию $\mu : \mathbb{R}^2 \rightarrow [0, 1]$, такую, что $\mu(x, y) = 1 \Leftrightarrow (x, y) = (0, 0)$ и для любых $x \in [-1, 1]$, $y \in [-1, 1]$ и любого $\lambda \in (0, 1)$ выполняется неравенство $\mu(\lambda x, 0) > \mu(x, 0)$, и если $y \neq 0$, то $\mu(\lambda x, \lambda y) \geq \mu(x, y)$. Рассмотрим некоторую функцию $\mu' : \mathbb{R} \rightarrow [0, 1]$, такую, что $\mu'(x) = 1 \Leftrightarrow x = 0$ и для любого $x \in [-1, 1]$ и любого $\lambda \in (0, 1)$ выполняется неравенство $\mu'(\lambda x) > \mu'(x)$. Пару $\phi = \langle \mu, \mu' \rangle$ назовем *основанием* перехода. Основание $\phi = \langle \mu, \mu' \rangle$ позволяет задать оператор перехода ϕ , который каждой ЗИП позволяет сопоставить некоторую ЗНП, а каждому ИГ некоторый НИГ.

Пусть $\rho \subseteq [0, 1]^2$ — отношение поиска. Через $\phi(\rho)$ обозначим нечеткое отношение $\eta(x, y) = \sup_{(x_0, y_0) \in \rho} \mu(x - x_0, y - y_0)$. Заметим, что если $\rho \neq \emptyset$, то множество $\{(x, y) \in [0, 1]^2 : \eta(x, y) = 1\}$ замкнуто.

Каждой ЗИП $I = \langle [0, 1], V, \rho \rangle$ сопоставим ЗНП $\phi(I) = \langle [0, 1], V, \phi(\rho) \rangle$.

Пусть F — некоторое базовое множество предикатов на $[0, 1]$. Если $f \in F$, то обозначим $N_f = \{x \in [0, 1] : f(x) = 1\}$. Если $A \subseteq [0, 1]$, то обозначим $\mu'(A)(x) = \sup_{x_0 \in A} \mu'(x - x_0)$. Предикату $f \in F$ сопоставим функцию $\phi(f)(x) = \mu'(N_f)(x)$. Множеству F сопоставим базовое

множество нечетких функций $\phi(F) = \{\phi(f) : f \in F\}$. Заметим, что если $N_f \neq \emptyset$, то множество $\{x \in [0, 1] : \phi(f)(x) = 1\}$ замкнуто.

Пусть U — ИГ над базовым множеством F . Сопоставим ему НИГ $\phi(U)$ над базовым множеством $\phi(F)$, полученный из ИГ U заменой для каждого ребра ИГ U предиката f , приписанного этому ребру, на функцию $\phi(f)$.

Возникает естественный вопрос: в каких случаях основание перехода ϕ от ЗИП I к ЗНП $\phi(I)$ обеспечивает переход от ИГ U , допустимого для ЗИП I , к НИГ $\phi(U)$, допустимому для ЗНП $\phi(I)$?

Для произвольного базового множества предикатов F рассмотрим всевозможные конечные подмножества $F_\lambda \subseteq F$. Для каждого F_λ и произвольного основания перехода $\phi = \langle \mu, \mu' \rangle$ рассмотрим функцию $g_{\lambda, \phi}(x) = \min_{f \in F_\lambda} \phi(f)(x)$. Заметим, что если для всех $F_\lambda \subseteq F$, для всех $x \in [0, 1]$ $g_{\lambda, \phi}(x) < 1$, то либо базовое множество F пусто, либо для всех предикатов $f \in F$ $N_f = \emptyset$.

Фиксируем основание перехода ϕ . Для тех индексов λ , для которых функция $g_{\lambda, \phi}$ принимает значение 1, обозначим

$$\begin{aligned}
 a'_\lambda &= \sup_{\substack{0 \leq x \leq 1 \\ \forall t \in [0, x] \ g_{\lambda, \phi}(t) = \mu'(t-x)}} x, & a''_\lambda &= \sup_{\substack{0 \leq x \leq 1 \\ \forall t \in [x, 1] \ g_{\lambda, \phi}(t) = \mu'(t-x)}} (1-x). \\
 b'_\lambda &= \sup_{\substack{0 \leq t \leq x \leq 1 \\ \forall t' \in (t, x] \ g_{\lambda, \phi}(t') = \mu'(t'-x)}} (x-t), & b''_\lambda &= \sup_{\substack{0 \leq x \leq t \leq 1 \\ \forall t' \in [x, t] \ g_{\lambda, \phi}(t') = \mu'(t'-x)}} (t-x).
 \end{aligned}$$

Если $g_{\lambda, \phi}(x) < 1$ для всех $x \in [0, 1]$, то положим $a'_\lambda = a''_\lambda = b'_\lambda = b''_\lambda = 0$.

Положим

$$a'_F = \sup_\lambda a'_\lambda, \quad a''_F = \sup_\lambda a''_\lambda, \quad b'_F = \sup_\lambda b'_\lambda, \quad b''_F = \sup_\lambda b''_\lambda.$$

Заметим, что всегда $a'_F \leq b'_F$, $a''_F \leq b''_F$.

Если для некоторого λ $b'_F = b'_\lambda$ или $b''_F = b''_\lambda$, то положим $u_F = 1$. Если такого λ не существует, то положим $u_F = 0$. Если $a'_F = b'_F$ и для некоторого λ $a'_F = a'_\lambda$, то положим $u'_F = 1$, иначе положим $u'_F = 0$. Если $a''_F = b''_F$ и для некоторого λ $a''_F = a''_\lambda$, то положим $u''_F = 1$, иначе положим $u''_F = 0$.

Для каждого подмножества $F_\lambda \subseteq F$, для которого $g_{\lambda,\phi}$ принимает значение 1, рассмотрим множество точек $V_\lambda = \{v_1^\lambda, v_2^\lambda, \dots\}$ таких, что $g_{\lambda,\phi}(v_i^\lambda) < 1$, и существует $x \in [0, 1]$, для которого $g_{\lambda,\phi}(x) = 1$, и либо для всех $t \in [x, v_i^\lambda]$ $g_{\lambda,\phi}(t) = \mu'(t - x)$, и в некоторой окрестности точки v_i^λ при $t > v_i^\lambda$ выполняется $g_{\lambda,\phi}(t) \neq \mu'(t - x)$, либо для всех $t \in (v_i^\lambda, x]$ $g_{\lambda,\phi}(t) = \mu'(t - x)$, и в некоторой окрестности точки v_i^λ при $t < v_i^\lambda$ выполняется $g_{\lambda,\phi}(t) \neq \mu'(t - x)$. Обозначим $c_i^\lambda = g_{\lambda,\phi}(v_i^\lambda)$. Если для некоторого индекса λ $0 \in V_\lambda$, то положим $b' = \inf_{\lambda, i: v_i^\lambda=0} c_i^\lambda$, иначе положим $b' = \infty$. Если для некоторого индекса

λ $1 \in V_\lambda$, то положим $b'' = \inf_{\lambda, i: v_i^\lambda=1} c_i^\lambda$, иначе положим $b'' = \infty$. Рас-

смотрим множества $V'_\lambda = V_\lambda \setminus \{0, 1\}$. Если для некоторого индекса λ $V'_\lambda \neq \emptyset$, то положим $b = \inf_{\lambda, i: v_i^\lambda \in V'_\lambda} c_i^\lambda$, иначе положим $b = \infty$. Положим $\varepsilon'_F = \min\{b', b\}$, $\varepsilon''_F = \min\{b, b''\}$.

Скажем, что базовое множество предикатов F и основание $\phi = \langle \mu, \mu' \rangle$ согласованы, если

1) Для любых $A_1, \dots, A_m \in \{N_f : f \in F\}$ таких, что $\bigcap_{i=1}^m A_i \neq \emptyset$,

$$\text{справедливо } \min\{\mu'(A_1), \dots, \mu'(A_m)\} \equiv \mu'(\bigcap_{i=1}^m A_i),$$

2) $\mu(x, 0) \equiv \mu'(x)$ для всех $x \in (-b'_F, b''_F)$, причем

а) если $u'_F = 1$, то $\mu(-b'_F, 0) = \mu'(-b'_F)$, если $u''_F = 1$, то $\mu(b''_F, 0) = \mu'(b''_F)$,

б) если $u'_F = u''_F = 0$ и $u_F = 1$, то $\max\{\mu(-b'_F, 0), \mu(b''_F, 0)\} = \max\{\mu'(-b'_F), \mu'(b''_F)\}$,

3) для всех $y \neq 0$ для всех $x \in [-1, 0]$ справедливо $\mu(x, y) \leq \varepsilon'_F$,
для всех $y \neq 0$ для всех $x \in [0, 1]$ справедливо $\mu(x, y) \leq \varepsilon''_F$.

Назовем ИГ *неизбыточным*, если в нем нет ориентированных цепей, проводимость которых тождественно равна нулю.

Множество избыточных ИГ над базовым множеством F , допустимых для ЗИП I , обозначим $\mathcal{U}'(I, \mathcal{F})$.

Скажем, что базовое множество предикатов F и основание $\phi = \langle \mu, \mu' \rangle$ вырождены, если для любого отношения

$\rho \subseteq [0, 1]^2$, любой ЗИП I типа $\langle [0, 1], [0, 1], \rho \rangle$ выполняется: $(U \in \mathcal{U}'(I, F) \implies \phi(U) \in \mathcal{U}(\phi(I), \phi(F)))$.

Теорема 2. *Базовое множество предикатов F и основание $\phi = \langle \mu, \mu' \rangle$ вырождены тогда и только тогда, когда они согласованы.*

Доказательство этой теоремы можно найти в [2].

Из теоремы 2 следует, что требование по сохранению свойства допустимости выполняется лишь для очень узкого класса оснований перехода ϕ и, кроме того, накладывает сильные ограничения на базовое множество предикатов F .

4. Сведение задачи нечеткого информационного поиска к задаче информационного поиска большей размерности

Скажем, что ЗНП $I = \langle X, V, \eta \rangle$ и ЗИП $I' = \langle X \times [0, 1], V, \rho \rangle$ эквивалентны, если $(x, c) \rho y \Leftrightarrow \eta(x, y) \geq c$.

Скажем, что нечеткая функция $f(x)$ и предикат $g(x, c)$ эквивалентны, если $g(x, c) = 1 \Leftrightarrow f(x) \geq c$. Будем обозначать: $f \sim g$.

Любому базовому множеству нечетких функций $\mathcal{F} = \{f_\kappa\}$ можно единственным образом поставить в соответствие базовое множество предикатов $G = \{g_\kappa : g_\kappa \sim f_\kappa, f_\kappa \in \mathcal{F}\}$. Базовые множества \mathcal{F} и G будем называть эквивалентными.

Возьмем НИГ U над базовым множеством нечетких функций \mathcal{F} . Занумеруем произвольным образом вершины U . Пусть (i, j) — ребро U (ориентированное ребро с началом в вершине i и концом в вершине j), f_{ij} — функция, приписанная этому ребру. Заменяем f_{ij} на эквивалентный ей предикат g_{ij} . Выполним такую замену для всех ребер U . В результате получим ИГ U' над базовым множеством предикатов G , эквивалентном \mathcal{F} . Полученный ИГ U' будем называть эквивалентным НИГ U .

Теорема 3. *Если НИГ U над базовым множеством нечетких функций \mathcal{F} допустим для ЗНП I , то ИГ U' , эквивалентный НИГ U ,*

над эквивалентным по отношению к \mathcal{F} базовым множеством предикатов G допустим для ЗИП I' , эквивалентной ЗИП I , причем $T(U) = T(U')$.

Доказательство. Занумеруем одинаковым образом вершины НИГ U и ИГ U' .

Понятно, что проводимость любой цепи $\theta(C)$ и функция фильтра любой вершины φ_β ИГ U' представляют собой предикат. Действительно,

$$\theta(C)(x, c) = \min_{d \in C} \theta(d)(x, c) \equiv \bigwedge_{d \in C} g_d(x, c),$$

$$\varphi_\beta(x, c) = \max_{C \in \mathcal{C}_\beta} \theta(C)(x, c) \equiv \bigvee_{C \in \mathcal{C}_\beta} \bigwedge_{d \in C} g_d(x, c),$$

где \mathcal{C}_β — множество ориентированных цепей ИГ U' из корня в вершину β , d — ребро ИГ U' , $\theta(d)$ — проводимость ребра d , по определению равная функции, приписанной ребру d (в данном случае — предикату g_d).

Покажем, что функции фильтров соответствующих листьев (то есть листьев с одинаковыми номерами) НИГ U и ИГ U' эквивалентны.

Действительно, соответствие между ребрами НИГ U и ИГ U' взаимно однозначно: каждому ребру (i, j) одного графа соответствует ребро (i, j) другого графа, и наоборот. По определению эквивалентности НИГ U и ИГ U' , проводимости соответствующих ребер эквивалентны. Рассмотрим цепь $C = \{d_1, \dots, d_m\}$ из корня в вершину β НИГ U и соответствующую цепь C' ИГ U' . Будем считать, что ребру d_i приписана функция f_i , а функции f_i эквивалентен предикат g_i . Верны соотношения:

$$\theta(C) \equiv \min_{d \in C} \theta(d)(x) \equiv \min_{i=\overline{1, m}} f_i(x) \geq c \Leftrightarrow \forall i = \overline{1, m} \quad f_i(x) \geq c \Leftrightarrow$$

$$\Leftrightarrow \forall i = \overline{1, m} \quad g_i(x, c) = 1 \Leftrightarrow \bigwedge_{i=1}^m g_i(x, c) \equiv \theta(C')(x, c) = 1.$$

Пусть $\mathcal{C}_\beta = \{C_1, \dots, C_s\}$ — множество цепей из корня НИГ U в вершину β . Для функции фильтра φ_β вершины β НИГ U и функции фильтра $\varphi_{\beta'}$ соответствующей вершины β' ИГ U' выполняются соотношения:

$$\begin{aligned} \varphi_\beta(x) \equiv \max_{C \in \mathcal{C}_\beta} \theta(C)(x) &\equiv \max_{i=1, s} \theta(C_i)(x) \geq c \Leftrightarrow \\ &\Leftrightarrow \exists j \in \overline{1, m} : \theta(C_j)(x) \geq c \Leftrightarrow \theta(C'_j)(x, c) = 1 \Leftrightarrow \\ &\Leftrightarrow \bigvee_{i=1}^s \theta(C'_i)(x, c) \equiv \varphi_{\beta'}(x, c) = 1. \end{aligned}$$

Пусть НИГ U допустим для ЗНП $I = \langle X, V, \eta \rangle$. Из эквивалентности функций фильтра соответствующих вершин НИГ U и ИГ U' следует, что $\mathcal{J}_U(x, c) \equiv \mathcal{J}_{U'}(x')$, где $x' = (x, c)$. Следовательно, из определения эквивалентности ЗНП и ЗИП, ИГ U' допустим для ЗИП $I' = \langle X', V, \rho \rangle$.

Из эквивалентности функций фильтра соответствующих вершин НИГ U и ИГ U' также сразу следует, что $T(U) = T(U')$.

Теорема доказана.

В доказательстве теоремы 3 показано, что для любого НИГ, допустимого для ЗНП, существует допустимый ИГ для эквивалентной ЗИП. Вообще говоря, может найтись другой ИГ, допустимый для ЗИП. Поэтому, справедливо

Следствие 1. *Если ЗИП I и ЗНП I' эквивалентны, множество нечетких функций \mathcal{F} эквивалентно множеству предикатов G , то $T(I, \mathcal{F}) \geq T(I', G)$.*

Следовательно, можно получить решение ЗНП с множеством запросов X и множеством записей Y , построив допустимый ИГ для эквивалентной ЗИП с множеством запросов $X \times (0, 1]$ и множеством записей Y . В общем случае, это решение не будет уступать по сложности решению исходной задачи.

5. Пример решения ЗНП через эквивалентную ЗИП

Рассмотрим в качестве примера ЗНП $I = \langle X, V, \eta \rangle$, где $X = Y = [0, 1]$, V — конечное подмножество Y , $\eta(x, y) = \max\{1 - 2|y - x|, 0\}$.

Для решения ЗНП I можно использовать алгоритм для задачи интервального поиска, описанный в [3].

В задаче интервального поиска множество запросов $X = \{[u, v] \subseteq [0, 1]\}$, множество записей $Y = [0, 1]$, отношение поиска ρ задается как

$$[u, v]\rho y \iff y \in [u, v].$$

Фактически, мы имеем запараметризованную задачу интервального поиска $I' = \langle X \times [0, 1], V, \rho \rangle$ с отношением ρ :

$$(x, c)\rho y \iff y \in [u, v],$$

где

$$u = \max\left\{0, x - \frac{1-c}{2}\right\}, \quad v = \min\left\{x + \frac{1-c}{2}, 1\right\}.$$

Поэтому, после соответствующей параметризации, для решения I становится возможным использование ИГ, допустимого для задачи интервального поиска.

Символом $f_{a,b}(x, c)$, где $a, b \in [0, 1]$, будем обозначать предикат, принимающий значение 1 в том и только в том случае, если

$$\left[x - \frac{1-c}{2}, x + \frac{1-c}{2}\right] \cap [a, b] \neq \emptyset,$$

символом $\bar{f}_{0,a}(x, c)$, где $a \in [0, 1]$ — предикат, принимающий значение 1 в том и только в том случае, если

$$\left[x - \frac{1-c}{2}, x + \frac{1-c}{2}\right] \cap [0, a] = \emptyset,$$

символом $f_a(x, c)$ — предикат, принимающий значение 1 в том и только в том случае, если

$$\min\left\{x + \frac{1-c}{2}, 1\right\} - \max\left\{0, x - \frac{1-c}{2}\right\} \leq a,$$

символом $\bar{f}_a(x, c)$ — предикат, принимающий значение 1 в том и только в том случае, если

$$\min \left\{ x + \frac{1-c}{2}, 1 \right\} - \max \left\{ 0, x - \frac{1-c}{2} \right\} > a.$$

В качестве базового множества возьмем

$$\mathcal{F} = \{f_{a,b}: a, b \in [0, 1]\} \cup \{\bar{f}_{0,a}: a \in [0, 1]\} \cup \{f_a: a > 0\} \cup \{\bar{f}_a: a > 0\}.$$

Упорядочим записи в библиотеке $V = \{y_1, \dots, y_k\}$ так, чтобы выполнялось $y_1 \leq y_2 \leq \dots \leq y_k$.

Возьмем произвольное ориентированное бинарное сбалансированное дерево с корнем и k висячими вершинами высоты $\lceil \log_2 k \rceil$, у которого все ребра ориентированы от корня к висячим вершинам. Объявим висячие вершины листьями. Из каждой внутренней (не висячей) вершины дерева исходит два ребра, назовем одно из них левым (Л), а другое правым (П). Тогда каждому листу можно сопоставить слово из символов Л и П, в соответствии с ребрами, встречающимися в цепи от корня к данному листу, причем первый символ слова соответствует ребру, исходящему из корня, а последний — ребру, входящему в лист. Упорядочим листья согласно лексикографическому порядку слов, соответствующих листьям, считая, что символ Л меньше символа П. Теперь i -ому листу (обозначим его α_i) сопоставим запись y_i , $i = 1, \dots, k$. Рассмотрим произвольную внутреннюю вершину β . Обозначим через V_β множество записей, соответствующих листьям ветви, растущей из вершины β . Отметим, что V_β состоит из подряд идущих записей, то есть если y_i — запись с минимальным номером в V_β , а y_j — запись с максимальным, то $y_l \in V_\beta$ для всех $l = i, \dots, j$. Обозначим через β' и β'' вершины, в которые ведут соответственно левое и правое ребра, исходящие из β . Пусть y_j — запись с максимальным номером в $V_{\beta'}$. Если β' — не лист, то сопоставим левому ребру, исходящему из β , предикат f_{0,y_j} , а если лист, то предикат f_{y_j,y_j} . Если β'' — не лист, то сопоставим правому ребру предикат \bar{f}_{0,y_j} , а если лист, то предикат $f_{y_{j+1},y_{j+1}}$. Такую операцию проделаем для всех внутренних вершин, тем самым определим нагрузку всех ребер. Из листа α_i выпустим ребро в лист α_{i+1} и припишем этому ребру предикат

$f_{y_{i+1}, y_{i+1}}$. Проведем эту операцию для всех номеров i от 1 до $k-1$. Корень полученного графа обозначим β_1 .

Пусть $S = \{s_1, \dots, s_m\}$, где $s_i = i/(m+1)$, $i = 1, \dots, m$. Здесь и далее $m = \lfloor \log_2 k \rfloor$.

Для каждого s_i , $i = 1, \dots, m$, найдем два целых числа l_i и r_i , первое из которых является номером ближайшей к числу s_i записи из V , меньшей, чем s_i , причем, если такой записи не существует, то примем $l_i = 1$, а второе — номером ближайшей к числу s_i записи из V , не меньшей, чем s_i , причем, если такой записи не существует, то примем $l_i = k$.

Объявим β_1 обычной внутренней вершиной. Проведем в β_1 ребро, начало его обозначим β_0 и объявим корнем полученного графа. Припишем этому ребру предикат $f_{1/(m+1)}$. Выпустим из корня еще одно ребро, конец которого обозначим через β_2 , и припишем этому ребру предикат $\bar{f}_{1/(m+1)}$. Построим ориентированное сбалансированное бинарное дерево с m висячими вершинами с корнем в вершине β_2 , все ребра которого ориентированы от β_2 к концевым вершинам. Из каждой внутренней вершины дерева исходит два ребра, одно из которых назовем левым (Л), а другое правым (П). Каждой концевой вершине можно сопоставить слово из символов Л и П, в соответствии с ребрами, встречающимися в цепи от корня к концевой вершине, причем первый символ слова соответствует ребру, исходящему из корня, а последний — ребру, входящему в концевую вершину. Занумеруем концевые вершины в порядке возрастания в соответствии с лексикографическим порядком слов, и обозначим их через $\beta'_1, \dots, \beta'_m$, то есть вершине β'_1 соответствует слово (Л, Л, ..., Л), а вершине β'_m — слово (П, П, ..., П). Пусть β — произвольная внутренняя вершина дерева. Пусть β'_j — концевая вершина с максимальным номером в ветви, растущей из вершины, в которое ведет левое ребро, исходящее из β . Тогда левому ребру, исходящему из β , припишем предикат f_{0, s_j} , а правому \bar{f}_{0, s_j} . Такую операцию проведем для всех внутренних вершин дерева и полностью определим нагрузку его ребер.

Построим еще k вершин, обозначим их $\alpha'_1, \dots, \alpha'_k$, и объявим их листьями, и припишем каждому листу α'_i ($i = 1, \dots, k$) запись y_i . Теперь для каждого $i = 2, \dots, k$ из листа α'_i построим ребро в лист α'_{i-1} и припишем ребру $(\alpha'_i, \alpha'_{i-1})$ предикат $f_{y_{i-1}, y_{i-1}}$. Затем для каждого

$i = 2, \dots, m$ из вершины β'_i выпустим два ребра: одно ребро в лист α_{r_i} и припишем ему предикат $f_{y_{r_i}, y_{r_i}}$, а второе в лист α'_i и припишем ему предикат f_{y_i, y_i} . Полученный граф обозначим через U .

Покажем, что ИГ U допустим для ЗИП $I' = \langle X \times [0, 1], V, \rho \rangle$. Рассмотрим произвольную запись $y_i \in V$. Ясно, что $L_U(y_i) = \{\alpha_i, \alpha'_i\}$.

В каждый из листьев α_i и α'_i ведут только ребра, которым приписан предикат f_{y_i, y_i} , следовательно,

$$N_{\varphi_{\alpha_i} \vee \varphi_{\alpha'_i}} \subseteq N_{f_{y_i, y_i}} = O(y_i, \rho). \quad (1)$$

Пусть (x, c) — произвольный запрос, такой, что

$$\max \left\{ 0, x - \frac{1-c}{2} \right\} \leq y_i \leq \min \left\{ x + \frac{1-c}{2}, 1 \right\}, \quad (2)$$

то есть $(x, c) \in O(y_i, \rho)$. Покажем, что

$$\varphi_{\alpha_i}(x, c) \vee \varphi_{\alpha'_i}(x, c) = 1.$$

Рассмотрим сначала случай, когда $(x, c) \in N_{f_{1/m+1}}$. Тогда проводимость ребра (β_0, β_1) равна единице, а ребра (β_0, β_2) — нулю. Поскольку любая цепь, ведущая в лист α'_i , проходит через ребро (β_0, β_2) , то $\varphi_{\alpha'_i}(x, c) = 0$.

Так как проводимость ребра (β_0, β_1) на запросе (x, c) равна 1, то мы выходим в вершину β_1 . Покажем, что $\varphi_{\alpha_i}(x, c) = 1$.

Обозначим через G'_β подсеть сети U , состоящую из цепей, исходящих из вершины β .

При $i = 1$ в лист α_1 сети G'_{β_1} ведет единственное ребро из вершины, которую обозначим через β' , функция фильтра листа α_1 в сети G'_{β_1}

$$\varphi_{\alpha_1}|_{G'_{\beta_1}} = \varphi_{\beta'} \wedge f_{y_1, y_1}.$$

Через y_j обозначим максимальную запись в библиотеке $V_{\beta'}$. Примем $c = y_j$, если $j \neq k$, и $c = 1$, если $j = k$. Очевидно, $c \geq y_1$. Тогда

$$\begin{aligned} N_{\varphi_{\alpha_1}|_{G'_{\beta_1}}} &= \\ &= \left\{ (x, c) \in N_{\varphi_{\beta'}} : \max \left\{ 0, x - \frac{1-c}{2} \right\} \leq y_1 \leq \min \left\{ x + \frac{1-c}{2}, 1 \right\} \right\} = \end{aligned}$$

$$\begin{aligned}
&= \left\{ (x, c) \in X \times [0, 1] : \max \left\{ 0, x - \frac{1-c}{2} \right\} \leq c, \right. \\
&\quad \left. \max \left\{ 0, x - \frac{1-c}{2} \right\} \leq y_1 \leq \min \left\{ x + \frac{1-c}{2}, 1 \right\} \right\} = \\
&= \left\{ (x, c) \in X \times [0, 1] : \max \left\{ 0, x - \frac{1-c}{2} \right\} \leq y_1 \leq \min \left\{ x + \frac{1-c}{2}, 1 \right\} \right\} = \\
&\quad = N_{f_{y_1, y_1}},
\end{aligned}$$

то есть

$$\varphi_{\alpha_1}|_{G'_{\beta_1}} = f_{y_1, y_1}.$$

Предположим теперь, что для некоторого $i \geq 1$ функция фильтра листа $\alpha_i|_{G'_{\beta_1}}$ равна $\varphi_{\alpha_i}|_{G'_{\beta_1}} = f_{y_1, y_1}$.

Рассмотрим лист α_{i+1} , $i+1 \leq k$. В него ведут два ребра: из листа α_i и некоторой внутренней вершины β . Пусть y_l и y_j — соответственно минимальная и максимальная записи в библиотеке V_β . Пусть

$$c_1 = \begin{cases} y_{l-1}, & \text{если } l \neq 1, \\ 0, & \text{если } l = 1, \end{cases} \quad c_2 = \begin{cases} y_j, & \text{если } j \neq k, \\ 0, & \text{если } j = k. \end{cases}$$

Очевидно, что $c_1 \leq y_i$ и $c_2 \geq y_{i+1}$. Тогда

$$\varphi_{\alpha_{i+1}}|_{G'_{\beta_1}} = \varphi_{\alpha_i}|_{G'_{\beta_1}} \wedge f_{y_{i+1}, y_{i+1}} \vee \varphi_\beta \wedge f_{y_{i+1}, y_{i+1}},$$

или

$$\begin{aligned}
&N_{\varphi_{\alpha_{i+1}}|_{G'_{\beta_1}}} = \\
&= \left(\left\{ (x, c) \in X \times [0, 1] : \max \left\{ 0, x - \frac{1-c}{2} \right\} \leq y_i \leq \min \left\{ x + \frac{1-c}{2}, 1 \right\} \right\} \cap \right. \\
&\cap \left. \left\{ (x, c) \in X \times [0, 1] : \max \left\{ 0, x - \frac{1-c}{2} \right\} \leq y_{i+1} \leq \min \left\{ x + \frac{1-c}{2}, 1 \right\} \right\} \right) \cup \\
&\quad \cup \left(\left\{ (x, c) \in X \times [0, 1] : c_1 \leq \max \left\{ 0, x - \frac{1-c}{2} \right\} \leq c_2 \right\} \cap \right. \\
&\cap \left. \left\{ (x, c) \in X \times [0, 1] : \max \left\{ 0, x - \frac{1-c}{2} \right\} \leq y_{i+1} \leq \min \left\{ x + \frac{1-c}{2}, 1 \right\} \right\} \right) =
\end{aligned}$$

$$\begin{aligned}
 &= \left\{ (x, c) \in X \times [0, 1] : \max\left\{0, x - \frac{1-c}{2}\right\} \leq y_i, \min\left\{x + \frac{1-c}{2}, 1\right\} \geq y_{i+1} \right\} \cup \\
 &\quad \cup \left\{ (x, c) \in X \times [0, 1] : c_1 < \max\left\{0, x - \frac{1-c}{2}\right\} \leq y_{i+1}, \right. \\
 &\quad \quad \left. \min\left\{x + \frac{1-c}{2}, 1\right\} \geq y_{i+1} \right\} = \\
 &= \left\{ (x, c) \in X \times [0, 1] : \max\left\{0, x - \frac{1-c}{2}\right\} \leq y_{i+1} \leq \min\left\{x + \frac{1-c}{2}, 1\right\} \right\} = \\
 &\quad = N_{f_{y_{i+1}, y_{i+1}}}.
 \end{aligned}$$

Следовательно, при условии (2) $\varphi_{\alpha_i}(x, c) = 1$.

Теперь рассмотрим случай, когда $(x, c) \notin N_{f_{1/m+1}}$. Тогда проводимость ребра (β_0, β_1) равна нулю, а ребра (β_0, β_2) — единице, то есть мы попадем в вершину β_2 , а проводимость всех цепей, проходящих через (β_0, β_1) , равна 0. Из вершины β_2 сначала растет дерево, и оно обладает тем свойством, что из каждой его внутренней вершины исходит два ребра, и им приписаны противоположные предикаты, то есть всегда проводимость только одного из двух ребер равна 1. Отсюда следует, что для любого запроса всегда существует единственная цепь, соединяющая β_2 с некоторой вершиной β'_j , проводимость которой равна 1. Таким образом, после прохождения дерева мы попадаем в некоторую вершину β'_j , причем, по построению этого дерева, эта вершина такова, что s_j является минимальным числом из S , таким, что $s_j \geq \max\left\{0, x - \frac{1-c}{2}\right\}$.

Предположим сначала, что $y_j < s_j$. Тогда $i \leq l_j$, и поскольку

$$\max\left\{0, x - \frac{1-c}{2}\right\} \leq y_i \leq y_l < s_j \leq v,$$

то

$$f_{y_l, y_l}(x, c) = 1,$$

то есть проводимость ребра (β'_j, α'_l) равна 1. Из вершины α'_l в вершину α'_i ведет цепочка ребер, которым приписаны функции f_{y_m, y_m} , где $m \in \{i, \dots, l_j\}$, и, поскольку

$$\max\left\{0, x - \frac{1-c}{2}\right\} \leq y_i \leq y_m \leq y_l < \min\left\{x + \frac{1-c}{2}, 1\right\},$$

то

$$f_{y_m, y_m}(x, c) = 1, \quad m \in \{i, \dots, l_j\},$$

то есть проводимость цепи из α'_{l_j} в α'_i равна 1 на запросе (x, c) . Таким образом, мы показали, что существует цепь, ведущая из корня в лист α'_i , проводимость которой равна 1 на запросе (x, c) , то есть $\varphi_{\alpha'_i}(x, c) = 1$. Отметим, что если даже $f_{y_m, y_m}(x, c) = 1$, то есть из вершины β'_j мы попадем еще и в вершину α_{r_j} , то все равно $\varphi_{\alpha_i}(x, c) = 0$, так как из листа α_{r_i} в лист α_i нет цепи.

Осталось рассмотреть случай, когда $y_i \geq s_j$. Тогда $i \geq r_j$, и, поскольку $\max\{0, x - \frac{1-c}{2}\} \leq s_j \leq y_{r_j} \leq y_i \leq \min\{x + \frac{1-c}{2}, 1\}$, то $f_{y_{r_j}, y_{r_j}}(x, c) = 1$, и мы попадем в лист α_{r_j} . Из листа α_{r_j} в лист α_i существует цепь, и ее проводимость по аналогии с предыдущим случаем равна 1 на запросе (x, c) , то есть $\varphi_{\alpha_i}(x, c) = 1$. Также аналогично предыдущему случаю $\varphi_{\alpha'_i}(x, c) = 0$.

Таким образом мы показали, что если $(x, c) \in O(y_i, \rho)$, то

$$\varphi_{\alpha_i}(x, c) \bigvee \varphi_{\alpha'_i}(x, c) = 1.$$

Более того, мы показали, что при $(x, c) \in O(y_i, \rho)$

$$\varphi_{\alpha_i}(x, c) \bigwedge \varphi_{\alpha'_i}(x, c) = 0. \quad (3)$$

Учитывая (1), получаем, что

$$\varphi_{\alpha_i} \bigvee \varphi_{\alpha'_i} = \chi_{y_i, \rho}; \quad (4)$$

более того, с учетом (3)

$$N_{\varphi_{\alpha_i}} \bigcap N_{\varphi_{\alpha'_i}} = \emptyset. \quad (5)$$

В силу произвольности записи y_i получаем, что U допустим для ЗИП I .

Оценим сложность ИГ U . Обозначим через \mathcal{M} множество всех ребер ИГ U . Пусть \mathcal{M}_1 — множество ребер ИГ U , исходящих из листьев, и $\mathcal{M}_2 = \mathcal{M} \setminus \mathcal{M}_1$. Если обозначить через $T(\mathcal{M})$ сумму сложностей ребер из множества \mathcal{M} , то

$$T(U) = T(\mathcal{M}) = T(\mathcal{M}_1) + T(\mathcal{M}_2). \quad (6)$$

Оценим сначала $T(\mathcal{M}_1)$. Множество \mathcal{M}_1 состоит из $2(k-1)$ ребер, исходящих из листьев $\alpha_i, i = 1, \dots, k-1$ и $\alpha'_j, j = 2, \dots, k$, следовательно, учитывая (5), а затем (4), находим, что

$$\begin{aligned} T(\mathcal{M}_1) &= \sum_{i=1}^{k-1} \mathbf{P}(N_{\varphi_{\alpha_i}}) + \sum_{i=2}^k \mathbf{P}(N_{\varphi_{\alpha'_i}}) \leq \sum_{i=1}^k (\mathbf{P}(N_{\varphi_{\alpha_i}}) + \mathbf{P}(N_{\varphi_{\alpha'_i}})) = \\ &= \sum_{i=1}^k \mathbf{P}(N_{\varphi_{\alpha_i} \vee \varphi_{\alpha'_i}}) = \sum_{i=1}^k \mathbf{P}(O(y_i, \rho)). \quad (7) \end{aligned}$$

Обозначим через G''_{β_1} (G''_{β_2}) подсеть сети G'_{β_1} (G'_{β_2}), состоящую из ребер, не принадлежащих \mathcal{M}_1 .

Ярусом сети назовем множество вершин с одинаковой высотой (высота вершины есть длина наименьшей цепи от корня к данной вершине), а значение высоты назовем номером яруса. Пусть ξ_1 и ξ_2 — две внутренние вершины из одного яруса сети G''_{β_1} . Очевидно, что $V_{\xi_1} \cap V_{\xi_2} = \emptyset$, откуда следует, что $N_{\varphi_{\xi_1}} \cap N_{\varphi_{\xi_2}} = \emptyset$. Поэтому

$$\sum_{\xi} \mathbf{P}(N_{\varphi_{\xi}}) = \mathbf{P} \left(\bigcup_{\xi} N_{\varphi_{\xi}} \right),$$

где суммирование берется по всем внутренним вершинам одного яруса.

Легко показать, что сумма сложностей ребер, исходящих из вершин одного яруса дерева G''_{β_1} , не превышает

$$2\mathbf{P}(N_{f_{1/(m+1)}}) \leq \frac{6}{\log_2 k}. \quad (8)$$

Аналогично показывается, что сумма сложностей ребер, исходящих из вершин одного яруса дерева G''_{β_2} , не превышает

$$2\mathbf{P}(X \setminus N_{f_{1/(m+1)}}) \leq 2.$$

Поскольку количество ярусов в деревьях G''_{β_1} и G''_{β_2} не превышает соответственно $\lceil \log k \rceil$ и $\lceil \log \log k \rceil + 1$, в силу (8) с учетом ребер, исходящих из корня,

$$T(\mathcal{M}_2) \leq 2 \log_2 \log_2 k + 12.$$

С учетом (6) и (7) получаем

$$T(U, F) \leq \sum_{i=1}^k \mathbf{P}(O(y_i, \rho)) + 2 \log_2 \log_2 k + 12.$$

Список литературы

- [1] Гасанов Э. Э. Информационно-графовая модель хранения и поиска данных // Интеллектуальные системы. 1998. Т. 3. Вып. 3–4. С. 163–192.
- [2] Фещук А. А. К вопросу анализа нечетких информационных графов // Дискретная математика. 2002. Т. 14. Вып. 2. С. 65–84.
- [3] Гасанов Э. Э., Кудрявцев В. Б. Теория хранения и поиска информации. М.: Физматлит, 2002.