ОПРЕДЕЛЕНИЕ ПАРАМЕТРОВ НЕЧЕТКОЙ ГРАФОВОЙ МОДЕЛИ ГИПЕРСЕТИ КОНТРОЛЯ ДОСТОВЕРНОСТИ ЭЛЕКТРОННЫХ ТЕКСТОВ

Ахатов А.Р., Рахмонкулов Ф.
Самаркандский государственный университет

Концептуальной основой использования нечеткой семантической гиперсети является извлечение свойств информации, закономерностей статистических распределений n-кратных ошибок, полученных из большого контекста материалов. При решении задачи осуществляется переход от исходной модели к нечеткой модели, в которых фиксируются множества вершин и ребер графа, описывающие в общем виде сетевую структуру гиперсети [1].

Строится ориентированный нечеткий граф для поиска объекта, извлечения нечетких правил для контроля искаженной словоформы путем выбора альтернативных словоформ с целью коррекции ошибок в текстах.

Для оптимизации структуры нечеткой графовой модели рассмотрим три варианта архитектуры сетевой модели по нечетким критериям, в которых в качестве переменных параметров рассматриваются вершины графа, ребра графа, вершины и ребра графа.

В настоящей работе разработаны научно-методические основы построения нечеткой гиперсети для поиска, размещения поисковых образов, контроля достоверности передачи текстов электронных документов в системах автоматизированного документооборота предприятий и организаций [2]. Предложена методика проектирования гиперсети поиска и анализа словоформ естественного языка с различной архитектурой графовых моделей для реализации в системе контроля и коррекции орфографии на основе нейронных сетей, моделей нечетких выводов, баз знаний и баз данных. Разработаны модели оптимизации расчета параметров нечетких графовых моделей и показаны примеры получения численных результатов.

Методы оптимизации параметров нечетких графовых моделей. Задачи оптимизации параметров моделей представляют собой минимаксные задачи размещения поисковых объектов. Рассмотрим построение нечеткого графа. Путь p в нечетком графе, представляется последовательностью дуг $(a_1, a_2, ..., a_q)$, за его нечеткую длину принимается нечеткое число $\tilde{l}(p)$, равное сумме длин всех дуг, входящих в p, т.е.

$$\tilde{l}(p) = \sum_{(x_i, x_j) \in p} \tilde{c}_{ij} ,$$

где $\widetilde{c}_{ij}$ - нечеткой число, представляющее длину дуги, соединяющей вершины x_i и x_j.

Для кратчайшего пути $\overline{p}_{ik}$ между вершинами x_i и $x_k \in X$ требуется выполнение условия

$$\widetilde{l}\,(\overline{p}_{ik}) = \min_r \widetilde{l}_r(p_{ik}),$$

где p_{ik} - путь между вершинами $x_i, x_k \in X$;

$r = 1,2,\ldots,s$, s - число различных путей между вершинами $x_i, x_k \in X$ графа $\widetilde{G}$.

Решение задачи опирается на операции над интервальными числами. Пусть заданы два интервала $A=[d_{s1},d_{l1}]$ и $B=[d_{s2},d_{l2}]$. Сумма двух интервальных чисел $A=[d_{s1},d_{l1}]$ и $B=[d_{s2},d_{l2}]$ определяется по формуле

$$A+B=[d_{s1}+d_{s2}, d_{l1}+d_{l2}],$$

где величина $w(A)=d_{l1}-d_{s2}$ оценивает ширину интервала $A=[d_{s1},d_{l1}]$.

Центр интервала $A=[d_{s1},d_{l1}]$ вычисляется как

$$m_A=(d_{s1}+d_{l1})/2.$$

Предложим следующие способы оптимизации графовой модели, основанные на сравнение интервалов.

Способ 1. Сравнение левых границ интервалов. Пусть заданы два интервала $A=[d_{s1},d_{l1}]$ и $B=[d_{s2},d_{l2}]$. Тогда $A<B$, если $d_{s1}<d_{s2}$.

Способ 2. Сравнение правых границ интервалов. Пусть $A=[d_{s1},d_{l1}]$ и $B=[d_{s2},d_{l2}]$. Тогда $A<B$, если $d_{l1}<d_{l2}$.

Отметим, что в этих способах не учитывается длина интервала и при сравнении используется только одна из границ. Причем, если в первом случае $d_{s1}=d_{s2}$, а во втором - $d_{l1}=d_{l2}$, то необходимо производить сравнение по ширине интервалов.

Способ 3. Сравнение центров интервалов: $A<B$, если $m_A<m_B$.

Этот способ лучше учитывает размер интервалов и их левую и правую границы, но может возникнуть ситуация, когда $m_A=m_B$. В соответствии с вышесказанным, предложим обобщенный способ сравнения интервалов.

Способ 4. Для определения минимального из двух интервалов проверяется следующее условие.

Если $d_{s1}<d_{s2}$ и $d_{l1}<d_{l2}$, то $A<B$.

Если оно не выполняется, то необходимо проверить второе условие

$$A<B, \text{ если } m_A<m_B,$$

где $m_A=(d_{s1}+d_{l1})/2$ и $m_B=(d_{s2}+d_{l2})/2$.

Если $m_A=m_B$, то проверяется третье условие: $A<B$ и $w(A)<w(B)$, где $w(A)=d_{l1}-d_{s1}$ и $w(B)=d_{l2}-d_{s2}$.

Таким образом, сначала проверяется, выполняются ли одновременно условия 1 и 2, затем проверяется условие 3. Если и этого недостаточно, сравнивается ширина интервалов.

Вычисление параметров графовой модели. Для каждой вершины $x_i \in X$ приведенного графа $\widetilde{G}$ определим два нечетких чисел – внешнего и внутреннего разделения вершины x_i.

Тогда получим следующие выражения для вычисления нечетких чисел внешнего и внутреннего разделения:

$$\widetilde{s}_0(x_i) = \max_{x_j \in X}[\widetilde{d}(x_i, x_j)],$$

$$\widetilde{s}_t(x_i) = \max_{x_j \in X}[\widetilde{d}(x_j, x_i)].$$

Вершина x_0^*, для которой

$$\widetilde{s}_0(x_0^*) = \min_{x_i \in X}[\widetilde{s}_0(x_i)],$$

представляет внешний центр графа $\widetilde{G}$.

А вершина x_t^*, для которой

$$\widetilde{s}_t(x_t^*) = \min_{x_i \in X}[\widetilde{s}_t(x_i)],$$

представляет внутренний центр графа $\widetilde{G}$.

Вершины x_0^*, являющиеся внешним центром, представляют нечеткий внутренний радиус графа

$$\widetilde{p}_t = \widetilde{s}_t(x_t^*).$$

Когда нечеткое число внешнего и внутреннего разделения вершины x_i определяется выражением

$$\widetilde{s}_{0,t}(x_i) = \max_{x_j \in X}\{\widetilde{d}(x_i, x_j) + \widetilde{d}(x_j, x_i)\},$$

то считается, что на вершине $x_{0,t}^*$, достигается минимум выражения

$$\widetilde{s}_{0,t}(x_{0,t}^*) = \min_{x_i \in X}[\widetilde{s}_{0,t}(x_i)],$$

который представляет внешне-внутренний центр графа, а значение $\widetilde{\rho}_{0,t} = \widetilde{s}_{0,t}(x_{0,t}^*)$ - нечеткий внешне-внутренний радиус графа $\widetilde{G}$.

Предположим, что вершины графа заданы в виде интервалов

$$[d_s(x_i, x_j), d_l(x_i, x_j)],$$

где $d_s(x_i, x_j)$ и $d_l(x_i, x_j)$ – соответственно расстояния между наиболее близкими и наиболее удаленными друг от друга точками полигонов x_i и x_j.

Тогда формулы расчета нечетких чисел внешнего и внутреннего разделения примут вид:

$$\tilde{s}_0(x_i) = \max_{x_j \in X}\{[d_s(x_i, x_j), d_l(x_i, x_j)]\},$$
$$\tilde{s}_t(x_i) = \max_{x_j \in X}\{[d_s(x_j, x_i), d_l(x_j, x_i)]\}.$$

Так как в правых частях приведенных равенств присутствуют интервальные числа, то и значения чисел внешнего и внутреннего разделения будут представлять собой интервалы. В таком случае, внешний радиус графа определяется следующим образом:

$$\tilde{s}_0(x_0^*) = [s_{0s}(x_0^*), s_{0l}(x_0^*)] = \min_{x_i \in X}\{[s_{0s}(x_i), s_{0l}(x_i)]\},$$

а внутренний радиус определяется как

$$\tilde{s}_t(x_t^*) = [s_{ts}(x_t^*), s_{tl}(x_t^*)] = \min_{x_i \in X}\{[s_{ts}(x_i), s_{tl}(x_i)]\}.$$

Найдены кратчайшие нечеткие расстояния между вершинами. При определении минимальных и максимальных интервальных чисел использованы способы сравнения 1-3, описанные выше.

Нечеткие числа внешнего и внутреннего разделения вершин занесены в матрицы расстояний, которые расположены в последнем столбце и в последней строке матриц.

Определено, что рассмотренный способ сравнения интервалов лучше применять, когда оптимизация размещения центров объекта должна производиться по наименее удаленным друг от друга точкам полигонов, второй способ – когда оптимизация осуществляется по наиболее удаленным друг от друга точкам полигонов.

В качестве критерия сравнения интервалов лучше использовать их средние значения. При этом точка y является центром графа $\tilde{G}$ и значение радиуса будет меньше всех других значений $\tilde{s}_0(x_i)$ и $\tilde{s}_t(x_i)$, $x_i \in X$, $i = 1,2,3,4$.

В большинстве случаев могут быть заданы нечеткие веса вершин, характеризующие их важность. Нечеткие веса вершин или ребер графа могут быть представлены в виде нечетких треугольных чисел, чисел с колоколообразной функцией принадлежности, в виде лингвистических переменных.

Литература

1. Вохминцев А.В., Мельников А.В. Модель знаний на основе нечетких семантических гиперсетей для представления отношений между объектами в естественном тексте // Интеллектика, логистика и системология. – Челябинск, 2002. - Вып. 7. - с. 21-33.

2. Ахатов А.Р., Тишликов С.А. Оптимизация параметров гиперсети базs данных и базы знаний систем контроля и коррекции орфографии // Илмий тадқиқотлар ахборотномаси, СамДУ. – Самарканд, 2013. - № 5 (81). – с. 47-55

Сведения об авторах

Ахатов Акмал Рустамович, кандидат технических наук, доцент кафедры информационных технологий СамГУ,
Тел.: +998902716418,
e-mail: a-axatov@samdu.uz

Рахмонкулов Феруз, магистрант Самаркандского государственного университета по специальности «Информационные технологии в образовании»
Tel:+998933080022
feruz0123@mail.ru

Секция: Техника йўналишидаги олий ўқув юртларида замонавий ахборот-коммуникацион технологиялари ва таълим жараёнида улардан фойдаланишнинг долзарб муаммолари;