

УДК 519.71

*Е.И. Кучеренко, Д.А. Павлов*Харьковский национальный университет радиоэлектроники, Харьков, Украина
dmitry_pavlov@ukr.net

Некоторые аспекты анализа развития нечетких онтологий

В работе вводится понятие нечеткой онтологии. Рассматривается ряд вопросов, связанных с представлением в онтологии нечетких знаний и отношений, а также делается ряд важных обобщений. Предлагается развитие методов критериальной оценки адекватности изменений онтологий при развитии для нечеткого пространства состояний. Теоретические результаты подтверждены рекомендациями к практическому использованию.

Введение

Представление нечеткой информации в онтологии является предметом ряда важных исследований [1-3]. Это актуально в связи с тем, что онтология, будучи формальным отображением реальности, требует четкого представления информации, а реальность не всегда может быть четко выражена. Такое несоответствие порождает сложности концептуального характера. Среди таких сложностей – реализация процедур адекватного нечеткого логического вывода, реализация адекватных блоков интеграции онтологической информации и прочее. Но несмотря на указанные проблемы, существует ряд задач, для которых объективно существует нечеткое представление знаний. Среди них: работа с естественным языком, описание сложных процессов, создание словарей и тезаурусов и многое другое в условиях субъективного отображения реальности. Из вышесказанного следует, что использование нечетких онтологий, возможно в рамках некоторых ограничений, будет иметь место и, соответственно, их исследование представляет как научный, так и практический интерес.

Создание нечеткой онтологии так же, как и онтологии, функционирующей в четком пространстве состояний, сопряжено с итеративным процессом накопления и представления знаний, в результате которого мы имеем ряд версий нечетких онтологий $\{O\tilde{n}t_i\}, i \in I$, упорядоченный в пространстве времени I . Сопоставление версий с целью выявления их адекватности предметной области и отличий между собой является полезным средством, позволяющим повысить эффективность работы разработчика онтологий, что наглядно демонстрируют современные системы контроля версий [4]. Целью данной работы является повышение эффективности представления и использования знаний на основе создания формальных подходов к сопоставлению нечетких онтологий и выявления их адекватности в процессе развития.

Постановка задачи

Эта работа является очередным этапом в ряде исследований, посвященных анализу развития онтологий и созданию системы контроля версий развивающихся онтологий. В ней принимаются все утверждения и определения, данные в [5].

Пусть существуют версии развивающихся нечетких онтологий $\{O\tilde{n}t_i\}, i \in I$, отображающие помимо четких, нечеткие знания, включая представленные в нечетком пространстве состояния.

Требуется:

1. провести анализ тенденций с целью обобщения подходов построения нечетких онтологий;
2. предложить эффективные средства выявления показателей, сформированных в [5], для случая нечетких онтологических пространств, а именно:
- предложить и обосновать формальные критерии выявления неполноты ($N\tilde{p}l$) нечеткой онтологической информации:

$$N\tilde{p}l = \text{true}, \forall O\tilde{n}t_i \in O\tilde{n}t, i \in I, \quad (1)$$

где $O\tilde{n}t_i$ – версия нечеткой онтологии $O\tilde{n}t$ в момент времени i , I – пространство времени, в котором существует альтернативная версия онтологии;

- предложить и обосновать формальные критерии выявления избыточности ($\tilde{I}z$) нечеткой онтологической информации:

$$\tilde{I}z = \text{true}, \forall O\tilde{n}t_i \in O\tilde{n}t, i \in I; \quad (2)$$

- разработать подходы к выявлению и локализации свойств (1) и (2) в условиях нечетко представленных знаний в онтологиях;
3. сформулировать рекомендации по практическому использованию теоретических положений работы.

1 Формальная онтология

Определим формальную модель онтологии, для которой разрабатывается механизм сравнения версий.

Утверждение 1. Всякая рассматриваемая онтология $O\tilde{n}t$ может быть представлена в виде

$$O\tilde{n}t = \langle \tilde{T}, \tilde{R}, \tilde{A} \rangle, \quad (3),$$

где $\tilde{T}$ – иерархия терминологии, $\tilde{R}$ – набор отношений, $\tilde{A}$ – набор утверждений, основанный на $\tilde{T}$ и $\tilde{R}$.

Утверждение 1 ориентировано на работы, в которых структура онтологии проецируется на существующие подходы дескриптивной логики (ДЛ) [6]. На практике этот подход подтвердил свою эффективность.

Различная степень выразительности онтологий накладывает ряд ограничений на элементы $\tilde{T}, \tilde{R}, \tilde{A}$ в (3), а также на их взаимодействие. В данной работе вводится ограничение на существование в (3) терминологических циклов, в которых некий термин описывается с помощью самого себя, что не всегда приводит к адекватности отображения.

Замечание 1. Набор утверждений $\tilde{A}$ может использовать также элементы-предикаты ρ в качестве значений свойств, которые тождественны некоторым функциям принадлежности $\mu(k)$.

Важно отметить, что расширение онтологии элементами-предикатами является одним из развивающихся направлений развития онтологического представления знаний [7] и используется в работе как средство снижения уровня абстракции примеров.

2 Анализ подходов к представлению нечеткости в онтологии

Существующие подходы к анализу и сопоставлению версий онтологий [8] имеют ряд функциональных ограничений и принципиальных недостатков, к которым следует отнести: отсутствие возможности работы с нечеткими данными, высокая сложность процедур анализа. Существующие подходы к использованию нечеткости в онтологии носят разобщенный характер и не содержат общепризнанных методов [1], [2], [7], [9], [10].

Отражение нечеткости в онтологии тесно связано с задачей формирования мета-описаний об онтологической информации. Тогда как экземпляром знаний (утверждением) для всех компонентов из (3) является тройка $a_j = \langle s, p, o \rangle$, где s – «субъект», p – «предикат», o – «объект». Мета-описанным утверждением является пара

$$a_m = \langle a_j, meta_a \rangle, \quad (4)$$

где $meta_a$ – мета-информация об a_j .

В настоящее время наиболее распространенными методами введения мета-описаний (4) являются следующие подходы [9], [10]: реификация, введение мета-отношения и введение мета-класса.

Рассмотрим каждый из подходов.

Реификация заключается в превращении простого утверждения в развернутую структуру

$$\langle s, p, o \rangle \rightarrow [\langle x, subject, s \rangle, \langle x, predicate, p \rangle, \langle x, object, o \rangle], \quad (5)$$

где x – некоторый абстрактный объект, $subject$, $predicate$ и $object$ – зарезервированные отношения, обозначающие субъект, предикат и объект соответственно.

Это дает возможность введения дополнительных утверждений об объекте x , который семантически является идентификатором данной тройки.

Введение мета-отношения является реификацией особого рода. Оно заключается в замещении одиночного предиката p групповым предикатом P_r , который формально является представителем определенного рода класса. Например, для простого предиката «работать_в», связывающего некоторый хозяйственный объект и некоторого человека, может потребоваться указание даты, с которой актуальна данная запись. Тогда, создавая некоторый абстрактный класс «Работать_в», можно у него определить набор свойств, среди которых будет «сотрудник» и «предприятие», а также все мета-информационные атрибуты. Для

мета-отношений обязательными свойствами являются свойства «ОТ» и «В» (в примере «сотрудник» и «предприятие»). При таком подходе простое утверждение преобразуется в

$$\langle s, p, o \rangle \rightarrow [\langle Pr, OT, s \rangle, \langle Pr, B, o \rangle], \quad (6)$$

где «В» и «ОТ» – зарезервированные отношения, семантически обозначающие начало и конец замещаемого одиночного предиката.

С одной стороны, данный подход является более выразительным, нежели подход реификации, так как в описании отношения (6) используются расширенные возможности описания классов. С другой стороны, такой подход в значительной степени затрудняет логический вывод, так как теряются стандартные возможности описания отношений, что является существенным ограничением в реализации.

Введение мета-класса предлагает двухэтапное описание утверждения. Первый этап описания связывает субъект с представителем мета-класса. В свою очередь, этот представитель содержит непосредственно объект и дополнительную информацию об этом конкретном отношении. То есть

$$\langle s, p, o \rangle \rightarrow [\langle s, p', O \rangle, \langle O, is, o \rangle], \quad (7)$$

где O – представитель мета-класса, а p' – семантически эквивалентно p с учетом специфики значения свойства, « is » – зарезервированный предикат, обозначающий эквивалентность объектов. Примером мета-класса может служить следующее: для простого предиката «работать_в», взятого из вышеописанного примера, может потребоваться указание даты, с которой актуальна данная запись. Тогда, создавая некоторый мета-класс «Сотрудник_на_предприятии», одним из атрибутов которого может быть непосредственно предприятие, а другим – дата зачисления на службу на это предприятие, получаем возможность разделить конкретную связь «работать_в».

Все три описанных подхода широко применяются и имеют как положительные, так и отрицательные стороны [9], но важным аспектом их существования является то, что формально каждое из приведенных описаний может быть преобразовано в одно из двух других. Докажем это.

Теорема 1. Для всех мета-отношений вида (6) найдется хотя бы одно, эквивалентное ему реифицированное утверждение вида [(5), $\langle x, hasMeta, y \rangle$], где $hasMeta$ – отношение, обозначающее то, что объект является мета-информацией о субъекте.

Докажем теорему 1. Пусть существует некоторое мета-отношение Pr , с помощью которого вводится утверждение (6). Очевидно, что $Pr = [\langle s, relation, o \rangle, meta]$ есть совокупность отношения между объектом и субъектом и мета-описанием этого отношения. Рассмотрим некоторый объект из (3), равный мета-описанию $y = meta$. Введем такое отношение p , что $p^I = Pr^I - y^I$, то есть семантика отношения p является мета-отношением без учета мета-информации. Введем ряд простых утверждений $\langle x, subject, s \rangle, \langle x, predicate, p \rangle, \langle x, object, o \rangle, \langle x, hasMeta, y \rangle$. Совокупность данных утверждений не что иное, как реифицированное утверждение (4). В силу проведенных эквивалентных преобразований оно эквивалентно изначальному набору утверждений. Считаем теорему доказанной.

Теорема 2. Для всех мета-отношений вида (7) найдется хотя бы одно, эквивалентное ему реифицированное утверждение вида $[(5), \langle x, hasMeta, y \rangle]$.

Докажем теорему 2. Пусть существует утверждение вида (7). Определим свойство p , которое семантически эквивалентно p' , но соединяющее непосредственно субъект и объект. Введем некоторый объект y , такой, что $y' = O' - \langle O, is, o \rangle'$, то есть y содержит всю информацию из O , кроме связности с конкретным объектом o . Аналогично положениям доказательства теоремы 1. Результирующие утверждения эквивалентны изначальному набору утверждений. Считаем теорему доказанной.

Следствие 1 (из теорем 1 и 2). Любые реифицированные утверждения можно рассматривать в общем виде (4), не изменяя семантики логических рассуждений.

Опишем нечеткое утверждение $\tilde{a}_j$. Мета-элементом утверждения является функция принадлежности $\mu(k)$, введенная для нечетких множеств Заде (L. Zadeh), [11]:

$$\tilde{a}_j = \langle a_j, \mu_a \rangle. \quad (8)$$

μ_a в данном случае определяет степень уверенности в существовании утверждения a_j .

Утверждение 2. Любую онтологию (3), содержащую хотя бы одно утверждение вида (8), будем считать *нечеткой*.

Замечание 1. В случае, если значение $\mu_a(k) = 1$ в (8), то данное утверждение будет четким, и в таком случае его присутствие не влечет за собой нечеткость онтологии.

Замечание 2. В случае, если значение $\mu_a(k) = 0$ в (8), то это утверждение может быть извлечено из онтологии без влияния на ее семантику.

Результаты Утверждения 2 и Замечаний 1 и 2 имеют важное практическое значение.

3 Неполнота и избыточность в условиях нечеткости

Алгоритмы анализа адекватности версий онтологий [5] ориентированы на работу с простыми утверждениями. Для эффективной работы с нечеткими онтологиями требуется ряд специфических особенностей, учитывающих, в частности, (8).

Рассмотрим алгоритм рекурсивного анализа неполноты и избыточности версий онтологии в ракурсе присутствия нечеткости. Значения критериев неполноты и избыточности в четкой онтологии рассчитываются по формуле [5]:

$$\text{diff_}X(A, B) = \frac{(\sum \text{diff_}X(a_k, b_k))}{n}, k = \overline{1, n}, \quad (9)$$

где a_k – свойство объекта A , b_k – свойство объекта B , $A = \{a_1, a_2, \dots, a_n\}$, $B = \{b_1, b_2, \dots, b_n\}$, $0 \leq \text{diff_}X(A, B) \leq 1$, $\text{diff_}X$ – одна из функций $\text{diff_}Iz$ либо $\text{diff_}Npl$.

Для решения задачи в условиях нечеткости определим функции.

Утверждение 3. Отличие по неполноте объекта нечеткой онтологии есть нормализованная сумма максимумов между уменьшением в уверенности существования и неполноты данного атрибута данного объекта:

$$\text{diff_} \tilde{Npl}(\tilde{A}, \tilde{B}) = \frac{(\sum \max((\mu_{k_B} - \mu_{k_A}), \text{diff_} \tilde{Npl}(\tilde{a}_k, \tilde{b}_k)))}{n}, k = \overline{1, n}. \quad (10)$$

Действительно, если при изменении версии онтологии уменьшается уверенность в существовании атрибута $\tilde{a}_k$ у объекта $\tilde{A}$, это приводит к увеличению неполноты информации обо всем объекте. В то же время, если увеличивается непосредственно неполнота атрибута $\tilde{a}_k$, это также увеличивает общую неполноту. Увеличение уверенности в существовании связи между объектом $\tilde{A}$ и атрибутом $\tilde{a}_k$ не влияет на результат вычислений.

Утверждение 4. Отличие по избыточности объекта нечеткой онтологии есть нормализованная сумма максимумов между увеличением в уверенности существования и избыточности данного атрибута данного объекта:

$$\text{diff_} \tilde{Iz}(\tilde{A}, \tilde{B}) = \frac{(\sum \max((\mu_{k_A} - \mu_{k_B}), \text{diff_} \tilde{Iz}(\tilde{a}_k, \tilde{b}_k)))}{n}, k = \overline{1, n}. \quad (11)$$

Действительно, суть Утверждения 4 раскрывается аналогично Утверждению 3.

Учитывая замечание 1, для применения формализмов (10 – 11) будем преобразовывать утверждения онтологии к нечеткому виду (8) и выполнять решение задачи в общем виде.

4 Практическая реализация

Пусть существует онтология, обслуживающая производственный объект с системой вентиляции, использующий сводки погоды для снижения расхода энергии.

Обозначим ряд свойств объекта *Погода*, влияющих на учет и прогнозирование: *температура, влажность, облачность, ветер*.

Для каждого из свойств введем лингвистические переменные: *низкое, среднее, высокое*. Значения свойств погоды вводятся как нечеткие утверждения с некоторым значением, соответствуют функции принадлежности μ_x . Для повышения наглядности примера будем рассматривать значения данных свойств как предикаты p , зависящие от фактического значения показателей температуры, влажности, облачности и ветра.

$\{< \text{Погода, температура, низкая_температура} >, p_{\text{низкое}}(\text{факт_знач_тем})\}$

Не уходя от общности, введем конкретное представление предиката как

$$p_{\text{низкое}} = \max(0, 0.01 * t). \quad (12)$$

Для термина *Погода*, или C , определим ряд подтерминов *Душная, Нормальная, Бодрящая* или C_1, C_2, C_3 соответственно. Все подтермины попарно несовместны, то есть $x \notin C_j \forall x \in C_i$, если $i \neq j$, где $i = \overline{1, 3}, j = \overline{1, 3}$. Причем $C = \cup C_i, i = \overline{1, 3}$. Пусть для каждого из подтерминов существует набор необходимых и достаточных ограничений, оперирующий со свойствами, которые могут быть представлены как:

$\text{Душная} = \text{Погода} \cap \exists \text{температура.высокое} \cap \exists \text{влажность.высокое} \cap \exists \text{ветер.низкое}$

Введем понятие *Затраты_энергии_на_вентиляцию(зэнав)*, со свойством *при_погоде*, для которого существует набор из трех подклассов: *низкие_зэнав*, *средние_зэнав*, *высокие_зэнав*. Причем существуют необходимые и достаточные ограничения, такие, как:

высокие_зэнав = Затраты_энергии_на_вентиляцию $\cap$ Эпри_погоде.Душная,
средние_зэнав = Затраты_энергии_на_вентиляцию $\cap$ Эпри_погоде.Нормальная,
низкие_зэнав = Затраты_энергии_на_вентиляцию $\cap$ Эпри_погоде.Бодрящая.

Потребность в онтологии подобного вида не подлежит сомнению в связи с тем, что она позволит решать задачи оптимизации процессов в общем виде. Блок нечеткого логического анализа позволит эффективно извлекать полезную информацию из малого набора утверждений.

Дальнейшее развитие данной онтологии возможно по нескольким ветвям. Рассмотрим один из этих формализмов.

Интерпретация значения *температуры* подвержена влиянию разных обстоятельств, в частности времени года. Так, для периода *зима* высокой можно считать температуру 15 градусов, тогда как для периода *лето* такая температура будет нормальной, а возможно, даже и низкой. Тогда имеет смысл определить для объекта *температура* свойство времени замера, которое влияет на смещение нечетких понятий относительно оси измерений. Тогда (12) примет другой вид:

$$\rho''_{\text{низкое}} = \max(0, 0,01 * t - n(T)), \quad (13)$$

где $n(T)$ – некоторая функция зависимости от времени.

Переход от (12) к (13):

- изменит вычисляемое значение коэффициента μ_k , который станет более адекватным данной ситуации. Это приведет к изменению значений функций (10) – (11) для данного утверждения.
- свяжет утверждения о температуре с утверждениями о времени, что также отразится на численных значениях (10) – (11).

Выводы

1. Выполнен анализ тенденций и общих подходов к построению нечетких онтологий. Сделан ряд важных обобщений.
2. Впервые предложены формальные критерии и отображения нечеткой онтологии на основе обобщенных мета-элементов.
3. Получили дальнейшее развитие формальные критерии выявления Npl и Iz для случаев нечетких отношений.
4. Определено, что данное направление является перспективным при реализации в условиях нечетких знаний и нечетких отношений, что определяет практическую значимость результатов работы.

Литература

1. Straccia U. Towards a Fuzzy Description Logic for the Semantic Web // 2nd European Semantic Web Conference, 2005. – P. 167-181.
2. Quan T.T., Hui S.C., Cao T.H. FOGA: A Fuzzy Ontology Generation Framework for Scholarly Semantic Web // <http://olp.dfki.de/pkdd04/quant-final.pdf> (05.07.2005)

3. Mazziere M. A Fuzzy RDF Semantics to Represent Trust Metadata // Proc. Workshop Fuzzy Logic and the Semantic Web. – Marseille. – 2005.
4. Fish S. The New Breed of Version Control Systems // http://www.onlamp.com/pub/a/onlamp/2004/01/29/scm_overview.html (04.07.2005)
5. Кучеренко Е.И., Павлов Д.А. О проблемах выявления неполноты и избыточности информации в онтологическом пространстве // Прикладная радиоэлектроника. – 2005. – № 2.
6. Ian Horrocks, Peter Patel-Scheider. Reducing OWL Entailment to Description Logic Satisfiability // J. Web Sem. – 2004. – 1(4). – P. 345-357.
7. Straccia U. Description Logics with Fuzzy Concrete Domains // <http://dienst.iei.pi.cnr.it/Dienst/UI/2.0/Describe/ercim.cnr.isti/2005-TR-03?tiposearch=cnr&langver=> (05.07.2005)
8. PROMPT // <http://protege.stanford.edu/plugins/prompt/prompt.html> (04.07.2005)
9. Defining N-ary Relations on the Semantic Web // www.w3.org/TR/2004/WD-swbp-n-aryRelations-20040721/ (04.07.2005)
10. RDF/XML Syntax Specification // www.w3.org/TR/rdf-syntax-grammar/ (04.07.2005)
11. Zadeh L.A. Fuzzy Sets // Information and Control. – 1965. – 8(3). – P. 338-353.

Є.І. Кучеренко, Д.О. Павлов

Деякі аспекти аналізу розвитку нечітких онтологій

У роботі вводиться поняття нечіткої онтології. Розглянуто ряд питань, пов'язаних із представленням в онтології нечітких знань та відношень, а також виконано ряд важливих узагальнень. Пропонується розвиток методів критеріальної оцінки адекватності змін онтологій під час розвитку для нечіткого простору станів. Теоретичні результати підтверджено рекомендаціями щодо практичного використання.

E.I. Kucherenko, D.A. Pavlov

Some Aspects of Fuzzy Ontologies Development Analysis

Fuzzy ontology concept is declared in the work. There is spotted a number of questions concerned with fuzzy knowledge and relations representation in an ontology and also there is made a number of significant generalizations. Improvement of criteria methods for ontology conformity in evolution checking is suggested for a fuzzy space of conditions. The theoretical results are proved by recommendations to the practical use.

Стаття поступила в редакцію 08.07.2005.