

Е. А. Хала, г.Киев

ПОСТРОЕНИЕ ОНТОЛОГИЙ НА ОСНОВЕ ТЕКСТОВОГО КОНТЕНТА С ИСПОЛЬЗОВАНИЕМ НЕЧЕТКОЙ ЛОГИКИ

In article the way of introduction of indistinct logic directly in ontology on the set subject domain, and enrichments of the actual functions is considered. Two opportunities by means of which the expert can add knowledge in ontology are described. Proposed solutions allow to represent and argue on uncertain information.

Ключевые слова: нечеткий вывод, нечеткая логика, структура онтологии, анализ текста, текстовый контент.

Введение. Онтология – формальное описание конкретной прикладной области интересов, разделенной между гетерогенными приложениями. С поддержкой онтологий, пользователи и системы могут общаться друг с другом, поддерживая информационный обмен, базирующийся на семантике, а не только на синтаксисе.

Существуют онтологические приложения, где информация часто является неопределенной и неточной. Например, семантические приложения на основе семантических сетей, таких как электронная коммерция, управление знаниями, веб-порталы, и т.д. Действительно, концептуального формализма на основе стандартной онтологии может быть недостаточно, чтобы представить неточную информацию это довольно часто встречается во многих прикладных областях. Например, ключевые слова, полученные многократными запросами в одной и той же самой области нельзя рассмотреть с одинаковой актуальностью, так как некоторые ключевые слова могут быть более значимыми, чем другие. Таким образом, возникает необходимость в различной интерпретации в зависимости от контекста.

Одно из возможных решений, для обработки неопределенных данных и, следовательно, для решения таких проблем, заключается в том, чтобы обеспечить включение нечеткой логики в онтологию [1]. Цель теории нечетких множеств заключается в том, чтобы описать неопределенное понятие через обобщенное понятие множества, в соответствии с которым объект может принадлежать к определенной степени множества (как правило, вещественное число из интервала $[0,1]$). Например, семантическое содержание заявления, такого как "Coca-Cola безалкогольный сильногазированный ароматизированный напиток", может иметь степень, или значение истинности, 0,6. В настоящее время, нечеткие множества и онтологии совместно используются для устранения неопределенной информацией в различных областях, например, для поиска в документах или для создания научной онтологии из базы данных в ESKIMO и FOGA структурах. Тем не менее, ни в одном из этих примеров нет слияния теории

нечетких множеств и онтологии [1], но поскольку онтологии все больше набирают популярности, этот вопрос не мог остаться без внимания. Сейчас активно ведутся исследования и даже есть некоторые наработки применения теории нечеткой логики в онтологиях.

Целью данной статьи является представления способа интеграции нечеткой логики в онтологию для того, чтобы получить расширенную онтологию, которая больше подходит для решения проблем неопределенности рассуждений. Это первый шаг на пути к реализации теоретической модели и полноценной структуры, основанной на онтологии, которая в состоянии различить нюансы естественных языков.

Нечеткая логика и определение нечеткого значения

В этом разделе представляется логическая основа для работы с неопределенностью, что является основным аспектом для всех приложений онтологии, где пользователь заинтересован в информации, которая часто содержит неточное и расплывчатое описание понятий. Например, можно интересоваться поиском “сильно пряное красное вино” или в рассуждении с такими понятиями как “холодное место”, “дорогой товар”, “быстрый мотоцикл”, и т.д.

Для того чтобы справиться с этими проблемами предлагается подход основан на теории нечетких множеств. Идея состоит в расширении редактора онтологии, так чтобы непосредственно обрабатывать неопределенности во время определения онтологии, чтобы обогатить предметную область.

Сначала, вспомним определение нечеткого множества. Рассмотрим непустое множество объектов U , называемое универсумом. Нечеткое множество определяется как функция $[0,1]$ на U , $f: U \rightarrow [0,1]$. Полагая, что элемент $x \in U$, $f(x)$ представляет принадлежность значения x на множестве f [2].

Первая проблема, для решения заключается в том, как назначить нечеткое значение субъекта онтологии. В компромиссе между понятностью и точностью, получение высокой степени понятности и точности является противоречивой целью и, на практике, одно из двух свойств преобладает над другим. В зависимости от того, какое требование реализуется в первую очередь, нечеткую область моделирования можно разделить на две различные области:

1. Лингвистическое нечеткое моделирование - основной целью является получение нечетких моделей с хорошей интерпретируемостью;

2. Точное нечеткое моделирование - основной целью является получение нечеткой модели с хорошей точностью.

Необходимо быть как можно более объективными, к обеим вероятностям, что даются эксперту: определить точное значение или определить язык. В первом случай эксперт, при создании онтологии, определяет функция $f(\text{Понятие} \sqcup \text{Экземпляр}) \times \text{Свойство} \rightarrow \text{Значение свойства} \times [0,1]$ с учетом того что $f(o, p)$ – это значение, понятия или

экземпляра o принимающее свойство p с соответствующей степенью. Например, в гипотетической онтологии о попугаях, $f(\text{Кеша, цвет}) = (\text{зеленый}, 0,8)$ означает, что для свойства *цвет*, экземпляр Кеша, будет иметь значение зеленый со степенью 0,8 [4,5].

Очевидно, что могут существовать ситуации, в которых нет *Значений свойства* необходимых для данного свойства. Например, "У Кеша есть чувство юмора со значением 0,9" не может быть корректно выражено с помощью только формализма. В этой ситуации, необходимо сопоставление пары (понятия/экземпляр, свойство) на $[0,1]$, то есть $f':(\text{Понятие} \sqcup \text{Экземпляр}) \times \text{Свойство} \rightarrow [0,1]$ и выше упомянутый будет выглядеть $f'(\text{Кеша, чувство юмора}) = 0,9$ [5].

Для упрощения обозначения, определим функцию $g:(\text{Понятие} \sqcup \text{Экземпляр}) \times (\text{Свойство} \sqcup \text{Значение свойства}) \rightarrow [0,1]$. Таким образом, «Кеша, зеленого цвета со значением 0,8" будет $g(\text{Кеша, зеленый}) = 0,8$. Используя такую функцию g , у эксперта есть шанс выбрать ценность значения с бесконечной точностью, то есть точность предпочтительней интерпретируемости.

С другой стороны вторая возможность заключается в том, чтобы выбрать в качестве членов значения, метку в заданном множестве. Возьмем множество $L = \{\text{мало, достаточно, умеренно, вполне, очень, полностью}\}$ которое, очевидно, не является исчерпывающим из всех возможных меток, но которое может быть интуитивно изменено.

В этом случае значение $g(o, p)$ автоматически присваивается в соответствии с таблицей 1.

Таблица 1. Сопоставление нечеткого значения метке.

Метка	Значение
мало	0.2
достаточно	0.4
умеренно	0.6
вполне	0.7
очень	0.8
полностью	1

Подводя итог, мы даем возможность добавить ценность значения в паре (понятия/экземпляр, свойство) двумя различными способами: через точное значение $v \in [0,1]$ или выбрав метку в определенном заранее множестве L . Таким образом, через функцию g определим новое отношение в онтологии.

Другая возможность состоит в том, чтобы присвоить нечеткое значение объект (понятие или экземпляр). В этом случае можно определить функцию $h:\text{Понятие} \sqcup \text{Экземпляр} \rightarrow [0,1]$.

Отметим, что нечеткое значение задается с помощью одно из двух функций g и h , числом в единичном интервале $[0,1]$. Это позволяет приложениям, основанным на нечеткой онтологии, использовать стандарты и

хорошо изученные инструменты. Например, для того, чтобы задать два (или более) различных нечетких значения, может использоваться оператор накопления. Самый простой пример t -норма и t -конорма, то есть двоичное отображение, которое дает семантику с “OR”, “AND” операторами. Наиболее известные геделевские норма и конорма, т. е. минимаксные операторы. Учитывая сказанное выше, есть необходимо вычислить значение истинности заявления “Кеша зеленого цвета и имеет чувство юмора”. Если известно, что $f(\text{Кеша, зеленый}) = 0,8$ и $f(\text{Кеша, чувство юмора}) = 0,9$, то $[f(\text{Кеша, зеленый}) \text{ и } f(\text{Кеша, чувство юмора})] = \min\{0,8, 0,9\} = 0,8$ [5,6].

Определение для нечеткой онтологии будет следующим: нечеткая онтология – онтология, расширенная нечеткими значениями, которые задаются двумя функциями:

$$g: (\text{Понятие} \sqcup \text{Экземпляр}) \times (\text{Свойство} \sqcup \text{Значение свойства}) \rightarrow [0,1] \text{ и} \\ h: \text{Понятие} \sqcup \text{Экземпляр} \rightarrow [0,1].$$

Обновление нечеткого значения

После того, как эксперт создал нечеткую онтологию, неверно предполагать, что она хороша и, что любое нечеткое значение четко определено и подходит к любой обстановке. Таким образом, механизм изменения нечетких значений необходим для того, чтобы наилучшим образом подогнать их к конкретной среде или обобщить.

Предположим, что текущее нечеткое значение f и, как результат запроса оно должно быть обновлено до f_{new} . Простейший способ, это задать $f: = f_{new}$. Закономерно предположить, что после того, как некий запрос нечеткого свойства достигнет стабильного значения, будет бесполезно изменять его на f_{new} , теряя всю историю приобретенных знаний. Решение этой проблемы может быть в уменьшении важности f_{new} при любом изменении:

$$f: = f + (f_{new} - f)/(Q + 1), \quad (1)$$

где Q – количество обновлений выполненных для заданного значения. Понятно, что значение Q должно храниться в онтологии для любого определения нечеткого значения [7].

Теперь вопрос заключается в том, чтобы вычислить новое нечеткое значение f_{new} . Маловероятно найти в документе точное определение нечеткого значения, но, как правило, можно найти языковой классификатор. Например, мы не нашли “Coca-Cola со сладким вкусом со значением 0,6”, но смысл такого заявления в том, что “У Coca-Cola очень сладкий вкус”. Для этого рассматривается подход на основе модификаторов понятия.

Модификаторы понятий имеют влияние на изменение нечеткого значения свойства. Зададим ряд лингвистических ограничений таких как “очень”, “более или менее”, “едва”, тогда модификатор понятия - система одного или более ограничений, таких как “очень немного” или “очень очень немного”. Любому (лингвистическому) модификатору понятия необходимо сопоставить (числовой) член модификатора.

Член модификатора – это значение $\beta > 0$, которое используется в качестве показателя для изменения значения функции принадлежности f как f^β .

В соответствии с их воздействие на нечеткое значение, ограничения можно разделить на две группы: типа концентрации и растяжения. Влияние модификатора понятия заключается в снижении степени принадлежности значения. Таким образом, в данном случае, он должен быть $\beta > 1$. Например, для ограничения "очень", назначается $\beta = 2$. Таким образом, если мы знаем, что $g(\text{Coca-Cola, сладкий вкус}) = 0,8$, т.е. "У Coca-Cola сладкий вкус со значением 0,8", то Coca-Cola имеет очень сладкий вкус со значением $0,8^2 = 0,64$. Напротив, ограничение типа расширения может повышать значение принадлежности, то есть $\beta \in (0,1)$. Например, если *слегка* присваивается $\beta = 0,25$ и $g(\text{Coca-Cola, сладкий вкус}) = 0,8$, то у Coca-Cola слегка сладкий вкус со значением $0,8^{0,25} = 0,95$, интуитивно понятно, что если что-то является "сладким", тогда оно в какой-то степени и "слегка сладкое". Заметим, что такой подход отличается от оригинала Заде, где "слегка", а также другие модификаторы, обрабатываются более сложным образом. Преимущество этого метода в том, что он обеспечивает универсальный и простой способ управления модификаторами понятий, даже если при углублении таким способом в семантику необходимо обращаться к цепям модификатора [4,7,8].

Таким образом, модификаторы понятий используется в литературе для определения нового нечеткого члена с учетом уже существующих. Например, если мы знаем нечеткое значение свойства *красный*, можно вывести нечеткое свойство члена *очень красный* простым возведением в степень 2 значения *красный*. Вместе с тем, здесь противоположная ситуация. В качестве примера, предположим, что знаем из онтологии свойство *красный* а также, как результат запроса, что некий объект "очень красный". Таким образом, из *очень красный* мы должны вывести новое свойство красного для этого объекта. Так, если в онтологии $g(o, \text{красный}) = 0,7$, необходимо увеличить это значение, например $g(o, \text{красный}) = 0,7^{0,5}$. Эффект *очень* повышает значение свойства не уменьшая его величину. Таким образом можно сказать, что в обычном случае выполняется вывод:

Красный \rightarrow *Очень красный*

в то время как в данной ситуации:

Красный и Очень красный \rightarrow *Красный*

Этот аргумент также распространяется на все другие концепции модификаторов. Таким образом, то, что обычно считают модификатором концентрации, становится растяжением и наоборот.

Необходимо решить следующее: определить какой модификатор рассматривается (и каково его значение) и определить метод вычисления β , значения для цепей концепций модификаторов.

Выберем множество ограничений $H = \{\text{очень, гораздо больше, больше, больше или меньше, средне, слегка}\}$. Это только один из возможных вариантов. Множество H может быть изменено в зависимости от своих

потребностей, при условии, что оно удовлетворяет следующим двум свойствам.

- H множество упорядочено, т. е. *очень* $< \dots <$ *слегка* и только значение β для меньших и больших элементов фиксировано, соответственно как 0,5 и 2.
- Два подмножеств концентрирующих ограничений или {более или менее, средне, слегка} растягивающих ограничений {очень, гораздо больше, более} имеют одинаковую мощность (одинаковое количество элементов).

Эти два условия необходимы для расчета принадлежности модификатора последовательности ограничений. Единственная разница в задании инверсии ограничения, связанное с использованием модификаторов понятий [8].

Открытым вопросом является то, что множество H модификаторов понятий, не исчерпывает всех нюансов естественного языка. Многие элементы могут быть добавлены в H , а это потребует новых алгоритмов для их обработки, поскольку не все существующие модификаторы понятий могут быть полностью заданы или точно разделены на две подгруппы одинаковой мощности.

Примеры применения

Приведем два примера возможного использования нечетких онтологий. Первый основан на нечетком значении связанной (случай, свойства) пары и второй - способ использует понятия с нечетким значение, чтобы исправить проблему перегрузки.

Расширение запросов. При выполнении запроса на документ, это обычная практика, чтобы расширить набор понятия, что уже присутствует в запросе с остальных которые могут быть получены из онтологии. Очевидно для данного значения, а также ее родители и потомков можно добавить в запрос, а затем искать в документе. Возможность использования нечетких онтологий расширяет запросы, кроме того, потомки и родители, случаев из понятий, которые в некоторой степени удовлетворяют запрос.

Чтобы принять решение, в каком случаи производить добавление к расширенному запросу, мы должны решить, как лингвистические метки соотносятся с численными значениями. Решение такое же, как ранние описанное, только метки принадлежащие множеству L допускаются в запросах, и они преобразуются в числовые значения в соответствии с табл. 1. Если c - понятие, p - собственность и l метка тогда $\mu(c, p, l)$ значение, данной метки l для собственности p и понятие c . Теперь рассмотрим все случаи i понятия c , и они включены в расширенное в расширенную запроса, если и только если:

$$|\mu(c,p,l) - g(i,p)| \leq e, \quad (2)$$

где e из $[0.1]$ уровень терпимости.

Очевидно, количество экземпляров, которые будут добавлены к расширенному запросу, зависит от значения e , тем больше, что e большинство экземпляров. Граница случаев $e = 0$ содержит только те случаи, которые точно соответствуют запросу, и входит в него, и $e = 1$, все экземпляры включены.

Понятно, что это простейший случай, когда только одно свойство присутствует в запросе. Если два или более запроса должны быть удовлетворены, необходимо обобщить уравнение (2). Предположим, что в запросе есть n свойств, ссылающиеся на те же понятия, что и экземпляр c , тогда

$$\sum_{j=1}^n |\mu(c, p, l) - g(i, p)| / n \leq e. \quad (3)$$

То есть требуем, чтобы среднее значение расстояний между *Значениями свойств* в запросе и *Значения свойств* в экземпляр было меньше допустимого e .

Перегрузка понятия. Как писалось выше возможность использования нечетких значение, связанного понятия является ограничение проблемы из-за перегрузки понятия в онтологии.

Давайте предположим, что понятие c присутствует в различных частях онтологии, цель состоит в том, чтобы дать указания о том, какое место является более значительным, по отношению к определенному домену. На первом этапе, любому понятию, которое присутствует в нескольких местах, дается равное нечеткое значение, таким образом, что они в сумме дают 1.

Для любого c_i множество его местных условий L_i , т. е. родителей и потомки вычисляется в онтологии. Затем все элементы $l \in L_i$ ищутся в документах при анализе и назначенным на них весом w_i^j в соответствии с значимостью, которую они имеют в документе. Допустим, что в рассматриваемом документе, есть l_j вхождение для элемента l . Тогда для любого понятия c_i и для любого документ d вычисляются следующая функции:

$$\mu_d(c_i) = \sum_j l_j w_i^j \quad (4)$$

Сумма по всем n документам μ_d обозначена как μ :

$$\mu(c_i) = \sum_{d=1}^n \mu_d.$$

Затем новое значение принадлежности для понятия c_i и документа d :

$$h_{new}(c_i) = \frac{\mu_d(c_i)}{\mu(c_i)}. \quad (5)$$

Затем эти значения используются для обновления $h(c_i)$ в соответствии в уравнении (1), и они применяются в порядке релевантности, так что значение из более значимых документов обрабатываются в первую очередь и имеют большее влияние, чем следующие.

Заключение. В статье был описан способ введения нечеткой логику напрямую в онтологию по заданной предметной области и обогащения

фактических функций. Предлагаемое решение позволяет представлять и рассуждать о неопределенной информации. И это актуальная проблема для всех тех областей, где приложения основаны на онтологии.

Также был представлен метод, основанный на модификаторах понятий, для автоматического обновления степени при обращении.

1. Асаки К., Ватада Д., Иваи С. и др. Прикладные нечеткие системы/ Пер. с япон.; под ред. Тэрано Т., Сугэно М. – М.: Мир, 1993. - 368 с.

2. Заде Л.А. Основы нового подхода к анализу сложных систем и процессов принятия решений// Математика сегодня: Сборник статей; пер. с англ. – М.: Знание, 1974. – С.5-49.

3. Заде Л. Понятие лингвистической переменной и ее применение к принятию приближенных решений/ Пер. с англ. Колмогоров А.Н., Новиков С.П. — М.: Мир, 1976. -165 с.

4. Заде Л.А. Роль мягких вычислений и нечеткой логики в понимании, конструировании и развитии информационных/ интеллектуальных систем/Пер. с англ. И.З. Батыршина// «Новости Искусственного Интеллекта», 2001,№2-3, С. 7-11.

5. Лихтарников Л.М., Сукачева Т.Г. Математическая логика: Курс лекций. Задачник-практикум и решения.— СПб.: Лань, 1998.- 285 с.

6. Новак В., Перфильева И., Мочкорж И. Математические принципы нечеткой логики/ Пер. с англ. под ред. Аверкина А.Н. – М.: ФИЗМАТЛИТ, 2006. – 352 с.

7. Орловский С.А. Проблемы принятия решений при нечеткой исходной информации. — М.: Радио и связь, 1981. — 286 с.

8. Штовба С.Д. Введение в теорию нечетких множеств и нечеткую логику.

Поступила 18.9.2013р.

УДК 621.314:519.22

Т. Л. Щербак, г. Киев

КЛАССИФИКАЦИЯ РЕЖИМОВ ЭЛЕКТРОПОТРЕБЛЕНИЯ ОРГАНИЗАЦИЙ

Abstract. Proposed a classification normal and abnormal processes electricity consumption modes based on analysis of the main factors of their formation. Considered problems of computer modeling in the study of the dynamics of such regimes.

Keywords. The process electricity consumption, normal mode, abnormal mode, the processes of change-point.

Введение. Научно-техническая проблематика исследований электропотребления различного уровня является разноплановой и многогранной, о чем свидетельствуют результаты большого количества