



OSTIS-2014

(Open Semantic Technologies for Intelligent Systems)

УДК 004.822:514

ЭКСПЕРТНЫЕ СИСТЕМЫ БЕЗ ЭКСПЕРТОВ

Клещев А.С., Смагин С.В.

*Институт автоматизации и процессов управления Дальневосточного отделения
Российской академии наук, г. Владивосток, Россия*

kleshev@iacp.dvo.ru

smagin@iacp.dvo.ru

В работе проведен анализ традиционных способов формирования баз знаний для экспертных систем, выделены их ключевые достоинства и недостатки. Рассмотрен подход, при котором в качестве модели зависимости между знаниями и данными используются онтологии предметных областей, имеющие вид небогатых систем логических соотношений с параметрами. Предложена схема индуктивного формирования и сопровождения хорошо интерпретируемых баз знаний для таких моделей, которая сохраняет все достоинства традиционных подходов и исключает их недостатки.

Ключевые слова: индуктивное формирование баз знаний; экспертная система; хорошая интерпретируемость; модель зависимости с параметрами; онтология предметной области.

Введение

Традиционно базы знаний для интеллектуальных (экспертных) систем формируются экспертами предметных областей вручную – при помощи соответствующих редакторов знаний [Protégé, 2013] [Клещев и др., 2006]. Как показали многолетние исследования, именно такое, экспертное, формирование баз знаний является наиболее «узким местом» в эффективном использовании подобных систем на практике [Гаврилова и др., 2004]. При своих очевидных преимуществах, такой подход имеет ряд серьезных недостатков, которые во многом являются причинами сложности разработки и сопровождения, а также высокой стоимости, как самих баз знаний, так и экспертных систем в целом.

Альтернативным способом создания баз знаний является интеллектуальный анализ данных (Data Mining), в частности автоматическое (индуктивное) формирование баз знаний на основе эмпирических данных (обучающих выборок) [Вагин и др., 2008]. Индуктивное формирование знаний представляет собой извлечение и переработку семантики из одной формы в другую, а именно из обучающей выборки – в базу знаний. Однако, все известные на сегодняшний день из литературы попытки двигаться в данном направлении не привели к прорывным результатам, потому что, отчасти снижая сложность и стоимость разработки и сопровождения, такой подход неизбежно приводил к формированию плохо интерпретируемых баз знаний [Витяев, 2006].

Причиной этого является то, что в традиционных постановках задач этого и смежных направлений (таких как машинное обучение (Machine Learning), обнаружение знаний в базах данных (Knowledge Discovery in Databases) и др.) считается, что модель зависимости между знаниями и данными (классами (кластерами) и объектами предметной области) – не известна. В этом предположении в качестве таких моделей зависимости выбираются те или иные проблемно-независимые математические модели [Загоруйко, 1999]. В результате формируемая при таком подходе база знаний получается формально пригодной для использования ее в экспертных системах, но при этом сама остается непонятной экспертам предметной области. При этом залогом эффективного использования экспертных систем является хорошая интерпретируемость (в системе понятий предметной области) их результата пользователями, а также понимание ими того, почему был получен именно этот результат. Автоматически (индуктивно) сформированные базы знаний (т.е. описания классов или кластеров) могут быть использованы в интеллектуальных системах только в том случае, если они понятны экспертам соответствующих предметных областей [Michie, 1980]. В этом случае специалисты предметных областей смогут не только сами пользоваться такими базами знаний в своей профессиональной деятельности, но и будут доверять экспертным системам, использующим модели этих знаний, а также смогут проверять выводы таких систем, сформированные подсистемами объяснений.

В работе [Клещев и др., 2012] рассматривается подход, при котором в качестве модели зависимости между знаниями и данными используются онтологии предметных областей, имеющие вид небогатых систем логических соотношений с параметрами. В таких моделях описанием классов (кластеров) является набор значений параметров, названный базой знаний, который является хорошо интерпретируемым по построению. Цель данной работы состоит в том, чтобы предложить схему индуктивного формирования и сопровождения хорошо интерпретируемых баз знаний для таких моделей, которая сохраняет все достоинства традиционных подходов и исключает их недостатки.

1. Экспертное формирование баз знаний

На Рисунке 1 представлена схема формирования и сопровождения баз знаний с участием эксперта предметной области. Эксперт предметной области при помощи редактора знаний формирует начальное состояние базы знаний. Поток объектов названо множеством описаний объектов, поступающих на вход экспертной системе при ее практическом использовании. Решатель задач, получая на вход описание некоторого объекта и используя в своей работе базу знаний, формирует результат, а также его объяснение, которые передаются пользователю. Далее результат поступает в систему верификации, где верифицированные результаты накапливаются, чтобы затем быть переданными эксперту, который, анализируя их, осуществляет сопровождение базы знаний (формируя ее текущее состояние).

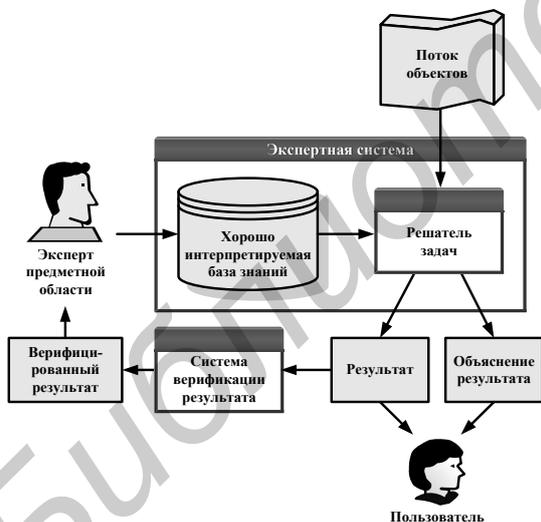


Рисунок 1 – Схема формирования и сопровождения баз знаний с участием эксперта предметной области

Перечислим наиболее важные преимущества и недостатки подобного подхода.

1.1. Преимущества экспертного формирования баз знаний

- База знаний хорошо интерпретируема, т.е. понятна специалистам предметной области.
- Уровень доверия к базе знаний зависит от квалификации и авторитета эксперта – чем выше авторитет, тем выше уровень доверия.

1.2. Недостатки экспертного формирования баз знаний

- Формирование и сопровождение базы знаний – процесс трудоемкий и по ряду причин (таких как неудобство, а также непривычность использования, низкая адаптивность программных средств) вызывает затруднения у экспертов.
- База знаний является субъективной, т.к. мнения групп экспертов одной и той же предметной области могут различаться.
- База знаний является неполной, т.к. опыт любого, даже первоклассного эксперта, всегда уже совокупности всех знаний о предметной области.
- Стоимость базы знаний (а, следовательно, и стоимость всей экспертной системы) высока, т.к. работа высококлассного специалиста стоит дорого – как на этапе формирования начального состояния базы знаний, так и на этапе ее сопровождения.
- Формирование и сопровождение базы знаний осуществляется вручную и требует участия того эксперта, который ее формировал, в течение всего периода использования экспертной системы.
- Замена одного эксперта на другого может привести к формированию новой базы знаний.

2. Индуктивное формирование и сопровождение хорошо интерпретируемых баз знаний

На Рисунке 2 представлена схема формирования и сопровождения баз знаний без участия эксперта предметной области.

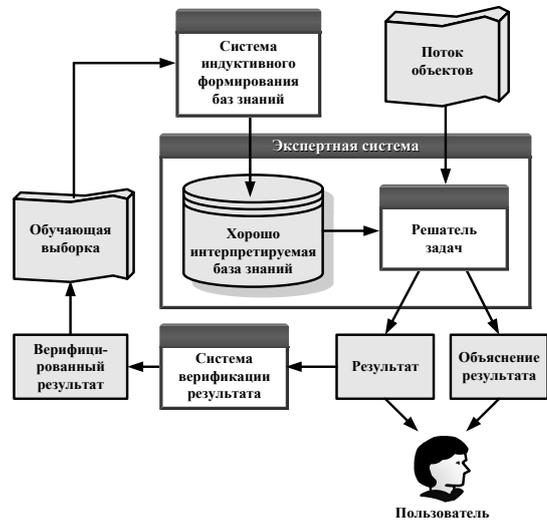


Рисунок 2 – Схема формирования и сопровождения баз знаний без участия эксперта предметной области

Принципиальным отличием данной схемы от предыдущей, является замена эксперта на систему индуктивного формирования баз знаний, на вход которой поступает обучающая выборка, а на выходе получается хорошо интерпретируемая база знаний. Тем не менее, на начальном этапе такая выборка, состоящая из максимально возможного числа наиболее полно описанных объектов предметной области, должна быть отобрана специалистом или экспертом. Перечислим преимущества этой схемы.

2.1. Преимущества предлагаемой схемы индуктивного формирования и сопровождения баз знаний

- База знаний так же понятна специалистами предметной области, как и в случае экспертного ее формирования.
- Формирование базы знаний происходит автоматически – по обучающей выборке.
- Уровень доверия к базе знаний зависит от объема и репрезентативности обучающей выборки.
- База знаний объективна, если обучающая выборка репрезентативна и имеет большой объем.
- База знаний полна настолько, насколько велик объем обучающей выборки, при условии, что обучающая выборка репрезентативна.
- Оценку полноты описаний объектов обучающей выборки может осуществить не только эксперт, но и достаточно опытный специалист предметной области, что снижает соответствующие расходы на формирование и сопровождение.
- В процессе практического использования экспертной системы качество базы знаний растет, т.к. результаты, которые прошли этап верификации, автоматически пополняют обучающую выборку.
- Ввиду особенностей системы индуктивного формирования (для указанных выше моделей), получаемая база знаний является абсолютно правильной и максимально точной относительно обучающей выборки, на основе которой она была сформирована, и при этом может обладать формой представления, принятой в предметной области.
- Алгоритмы, лежащие в основе системы индуктивного формирования знаний, независимы от экспертной системы и могут сопровождаться без необходимости в новой обучающей выборке.

3. Пример применения предлагаемой схемы

Рассмотрим в качестве адекватной и хорошо интерпретируемой модели зависимости онтологию медицинской диагностики, приближенную к реальной, которая представлена в виде небогатой системы логических соотношений с параметрами. Эта модель является важным частным случаем онтологии, опубликованной в работе [Клещев и др., 2005]. В работе [Клещев и др., 2012] предложены новые постановки основных задач индуктивного формирования баз знаний для таких моделей, а также представлен алгоритм обучения, решающий указанные задачи в этих постановках.

В работе [Смагин, 2013] представлено описание комплекса программ InForMedKB v1.1 (INductive FORmation of MEDical Knowledge Bases), в котором реализован указанный выше алгоритм обучения. Комплекс позволяет создавать обучающие выборки (состоящие из историй болезни (ИБ) различных разделов медицины) и на их основе индуктивно формировать базы медицинских знаний (в форме, принятой в медицинской литературе) хорошего уровня интерпретируемости и такого уровня

качества, который эксперты предметной области медицинской диагностики оценивают как достаточный для решения в ней практических задач. Также комплекс позволяет формировать объяснения баз знаний на основе описаний ИБ из используемых обучающих выборок. В разработанном комплексе нет формальной разницы между экспертными и индуктивно сформированными знаниями, т.к. форма представления одинакова – основана на онтологии и ориентирована на медицинские представления.

Описания заболеваний, входящих в базу медицинских знаний, состоят из описаний клинических проявлений признаков, входящих в клинические картины этих заболеваний. В таком описании для каждого признака указывается количество ИБ, в которых он наблюдался, число вариантов его клинического проявления (ВКП), а также описание этих вариантов. ВКП соответствует типу реакций организма на данное заболевание по данному признаку. Описание ВКП содержит информацию о числе периодов динамики в нем, о значениях признака в этих периодах и о границах длительности этих периодов. При этом для каждого ВКП формируется его объяснение – указывается количество ИБ, которые его поддерживают (в которых этот ВКП проявился), а также номера ИБ, которые входят в определяющее подмножество ИБ данного ВКП (исключение этих ИБ из обучающей выборки не позволит сформировать данный ВКП). Совокупность объяснений всех сформированных ВКП образует объяснение заболевания, а совокупность объяснений всех заболеваний образует объяснение базы медицинских знаний.

При помощи данного комплекса, на основе обучающей выборки реальных данных, содержащей истории болезни из раздела медицины «острый живот», была индуктивно сформирована база медицинских знаний, получившая в работе экспертную оценку. Ниже представлен фрагмент описания клинического проявления признака «Боли в животе (Локализация)» при аппендиците.

Признак «Боли в животе (Локализация)»

Количество историй болезни, в которых наблюдался признак, равно 22. Число вариантов динамики равно 9.

1 вариант: число периодов динамики = 2.

- *эпигастральная область*, затем через 5-7 часов

- *правая подвздошная область*

Вариант поддерживают истории болезни количеством 6.

2 вариант: число периодов динамики = 1.

- *правая подвздошная область*

Вариант поддерживают истории болезни количеством 3.

3 вариант: число периодов динамики = 2.

- *весь живот*, затем через 3-12 часов

- *правая подвздошная область*

Вариант поддерживают истории болезни количеством 6.

По мнению эксперта, полученная база знаний обладает высоким уровнем интерпретируемости для практикующего врача, входящие в нее описания заболеваний соответствуют знаниям, имеющимся в научной и учебной медицинской литературе, а в ряде случаев дополняют их описанием динамики клинических проявлений.

Заключение

Ключевым недостатком традиционного экспертного формирования баз знаний является сложность их сопровождения, в случае, когда на вход экспертной системе поступает поток объектов. Полученные и верифицированные результаты работы неизбежно выявляют дефекты в базе знаний, и чем таких объектов больше, тем сложнее эксперту ее модифицировать. Ключевым недостатком традиционного индуктивного формирования и сопровождения баз знаний является их плохая интерпретируемость. В аналогичных предыдущему случаю ситуациях, невозможность формирования понятного объяснения становится причиной того, что реально на практике экспертные системы не используются. Это, в свою очередь, влечет невозможность улучшения используемых баз знаний, т.к. поток объектов через экспертные системы не движется, а никакая обучающая выборка не способна адекватно заменить поток объектов реальной предметной области.

В работе представлена схема индуктивного формирования хорошо интерпретируемых баз знаний (для моделей зависимости с параметрами, являющихся онтологиями предметных областей), которая сохраняет все достоинства традиционных подходов, исключает их недостатки. Использование таких моделей зависимости и систем индуктивного формирования баз знаний, разработанных для них, позволяет формировать хорошо интерпретируемые базы знаний без участия экспертов предметной области. Работа выполнена при финансовой поддержке РФФИ (проекты № 12-07-00179-а и 14-07-00270-а) и ДВО РАН (проект № 12-И-П15-03).

Библиографический список

- [Вагин и др., 2008] Достоверный и правдоподобный вывод в интеллектуальных системах / Вагин В.Н. [и др.]; – М.: ФИЗМАТЛИТ. – 2008. – 704 с.
- [Витяев, 2006] Витяев Е.Е. Извлечение знаний из данных. Компьютерное познание. Модели когнитивных процессов. – Новосибирск: НГУ. – 2006. – 293 с.
- [Гаврилова и др., 2004] Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский – СПб: Питер. – 2000. – 384 с.
- [Загоруйко, 1999] Загоруйко Н.Г. Прикладные методы анализа данных и знаний. Новосибирск: Изд-во Ин-та математики. – 1999. – 270 с.
- [Клещев и др., 2005] Клещев А.С., Москаленко Ф.М., Черняховская М.Ю. Онтология и модель онтологии предметной области «Медицинская диагностика». – Владивосток: ИАПУ ДВО РАН. – 2005. – 44 с.
- [Клещев и др., 2006] Клещев А.С., Орлов В.А. Компьютерные банки знаний. Универсальный подход к решению проблемы редактирования информации // Информационные технологии. – М: Новые технологии. – 2006. – №5. – С. 25-31.
- [Клещев и др., 2012] Клещев А.С., Смагин С.В. Задачи индуктивного формирования знаний для онтологии медицинской диагностики // Научно-техническая информация. Серия 2. – М.: ВИНТИ РАН. – 2012. – №1. – С. 9-21.
- [Смагин, 2013] Смагин С.В. Комплекс программ для индуктивного формирования баз медицинских знаний в форме, принятой в медицинской литературе // Ломоносов-2013: Материалы XX Международной научной конференции студентов, аспирантов и молодых ученых. – М.: Издательский отдел факультета ВМК МГУ. – 2013. – С. 36-37.

[Michie, 1980] Michie D. Expert Systems // The Computer Journal. – 1980. – Vol. 23. № 4. – pp. 369-376.

[Protégé, 2013] Protégé – a free, open source ontology editor and knowledge-base framework [http://protege.stanford.edu]

EXPERT SYSTEMS WITHOUT EXPERTS

Kleschev A.S., Smagin S.V.

*Institute of Automation and Control Processes,
Far Eastern Branch of the Russian Academy of
Sciences, Vladivostok, Russia*

kleschev@iacp.dvo.ru

smagin@iacp.dvo.ru

Introduction

The article provides the analysis of traditional ways of knowledge bases formation for expert systems and highlights their key advantages and disadvantages. An approach is explored in which domain ontologies, having a form of unenriched systems of logical relationships with parameters, are used as a knowledge-data dependence model. A framework is suggested for inductive formation and maintenance of easily interpretable knowledge bases for such models, which retains all the advantages of traditional approaches and avoids their disadvantages.

Main Part

The key disadvantage of the traditional knowledge bases formation by experts is the complexity of maintaining them, when a stream of objects comes as input data into the expert system. Received and verified results inevitably reveal defects in the knowledge base, and the more such objects there are, the more difficult it becomes for the expert to modify the knowledge base.

The key disadvantage of traditional inductive formation and maintenance of knowledge bases is their poor interpretability. In such cases the inability of the expert system to form an understandable explanation, precludes the system from being used in actual practice. This, in turn, entails the impossibility to improve the knowledge base, as due to the fact that such expert systems are not being used, the stream of objects is not getting formed, and no learning sample is able to adequately replace the real stream of objects.

Conclusion

In the dependence models which are considered in this article the description of the classes and clusters of a domain is a set of parameter values (called a knowledge base), which is easily interpretable by way of construction. Use of such models and program systems for inductive formation of knowledge bases, developed using the models, will allow to form easily interpretable knowledge bases without the participation of domain experts.