

УДК 519.854

О.В. Калиниченко, С.Ю. Шабанов-Кушнаренко, А.В. Ярмак

Харьковский национальный университет радиоэлектроники, Харьков

О ПРЕДИКАТНЫХ МОДЕЛЯХ НЕЯВНЫХ ЗНАНИЙ В ЗАДАЧАХ АНАЛИЗА ИНФОРМАЦИОННЫХ ПРОЦЕССОВ

Усовершенствована предикатная модель выделения неявных зависимостей при трансформации явного представления знаний. Данная модель основана на представлении явных знаний в виде систем бинарных предикатов и определяет последовательность логических операций по преобразованию для пары графов. Формализованное преобразование данных в неявные знания обеспечивает теоретическую основу для выявления неявных знаний, и построения более достоверного представления знаний в системах искусственного интеллекта.

Ключевые слова: предикатная модель, неявное знание, инженерия знаний, информационный процесс

Вступление

Проблема представления знаний рассматривается в разделе искусственного интеллекта – инженерии знаний [1]. Для представления явных знаний разработано значительное количество методов и моделей, которые позволяют отобразить знания как в процедурной, так и в декларативной форме с целью их использования в системах искусственного интеллекта [2]. По иному обстоит дело с неявными знаниями в силу их природы – непрявленности, т.е. неотделимости от человека. Концепция использования неявных знаний требует дальнейшего развития. В то же время выделение и формализация неявных знаний может существенно расширить возможности систем искусственного интеллекта, приблизив их возможности к возможностям интеллекта человека. Традиционные подходы к анализу данных и процессов основаны на известной иерархической модели DIKW «данные-информация-знания-мудрость (метазнания)» [4].

Применение данной модели требует выделения взаимосвязей между данными и информацией, затем – между информацией и явными знаниями, затем – между неявными и явными знаниями. Тогда основное внимание уделяется механизму выявления неявных знаний путем преобразования информации и ранее полученных явных знаний.

Для получения исходной информации относительно структурированных объектов и процессов можно воспользоваться методами интеллектуального анализа данных и, в особенности, процессов. Интеллектуальный анализ процессов (process mining), как показано в работе [5], предназначен для построения моделей процессов. Исходными данными для анализа являются записи о выполнении процессов, представленные в виде файлов – логов событий. Такие логи содержат сведения о последовательности произошедших событий в

некоторой информационной системе с обязательными метками времени.

Логи событий могут, например, фиксировать выполнение бизнес-процессов, технологических процессов, поведение пользователей в социальных сетях. Методы process mining базируются и развивают методы интеллектуального анализа данных.

Постановка задачи

Структура процессов может изменяться вследствие влияния следующих факторов, связанных с использованием неявных знаний:

- скрытые неформализованные знания о выполнении процесса, влияющие на последовательность его активностей;
- опыт исполнителей процесса, приводящий к изменению внутренней структуры активностей и, следовательно, изменяющий их результат;
- порядок взаимодействия (часто неформальный) исполнителей – людей либо организаций при выполнении процесса;

Недостаточная разработанность концепции представления и использования неявных знаний в системах искусственного интеллекта требует анализа подходов к выявлению и использованию неявных знаний человеком и разработки модели представления таких знаний. Вышеизложенное определяет возможность получения неявных знаний путем интеллектуального анализа информационных объектов и процессов и подтверждает актуальность проблемы, рассматриваемой в статье. Решение данной проблемы позволит построить более адекватные модели информационных объектов и процессов, отражающих скрытые ранее неявные знания.

Целью работы является исследование и усовершенствование методов поиска неявных зависимостей при анализе информационных объектов, для преобразования знаний в явную форму и дополнения модели информационного объекта.

О методах выделения и преобразования неявных знаний

Между неявными и явными знаниями существуют следующие отличия:

- в случае явного знания человек может объяснить, записать, формализовать используемые закономерности;
- в случае неявного знания мы можем выделить практический результат его применения без объяснения того, как этот результат был получен.

Т.о., мы не видим внутренних закономерностей и можем лишь повторять внешние проявления использования скрытого знания.

Так, невозможно стать квалифицированным поваром, просто повторяя рецепты приготовления блюд (используя формализованные знания).

Необходим тот самый неявный контекст, отражающий особенности применения формализованных знаний (рецептов).

Изложенные особенности неявных знаний показывают, что решение проблемы их выделения и формализации сопряжено со значительными трудностями.

Поэтому перед детальным рассмотрением данной проблемы необходимо проанализировать подходы к обучению с использованием неявных знаний, которые трудно отделить от человека.

Извлечение и формализация скрытых зависимостей из структурированных массивов данных в задачах data-, process-, web - mining основывается на неявном влиянии таких зависимостей на сформированную структуру данных. Полученные в результате решения указанных задач структуры, например, модели процессов, представляются в виде графа и, фактически, представляют собой явные знания о процессах (либо структурированных объектах) предметной области.

В то же время, модели идентичных процессов (объектов), полученные в результате решения задач интеллектуального анализа массивов данных за различные интервалы времени, во многих случаях отличаются. Такие отличия в структуре полученных графов определяются применением неформализованных, скрытых знаний при выполнении процессов и формировании исходных массивов данных. Это и указывает на принципиальную возможность выделения неявных зависимостей при анализе структурированных объектов, полученных в результате исследования больших массивов данных.

Традиционные подходы к анализу данных и процессов основаны на известной иерархической модели DIKW «данные – информация – знания – мудрость (метазнания)»

Концепция DIKW опирается в первую очередь на наблюдения за естественным интеллектом [4]. В

данной работе эта модель представляет интерес в силу того, что неявные знания характерны именно для человеческого интеллекта.

Преобразование неявных знаний в явную форму позволяет использовать их в системах искусственного интеллекта.

Первый уровень модели DIKW содержит данные в виде набора символов, которые могут быть использованы, однако сами по себе не имеют значения. Данные являются просто результатом наблюдений и не всегда представлены в удобном для использования формате.

$$\{p_1, p_2, \dots, p_n\}, \quad (1)$$

$p_i, i = \overline{1, n}$ - множество данных.

На уровне информации между данными определяются связи, которые задают значение совокупности данных. На данном уровне не определяется, будет ли это значение использоваться, однако на основе информации могут быть сделаны определенные выводы.

При предикатном представлении информации предикат I задает связи между данными $\{p_1, p_2, \dots, p_n\}$, чем и определяет структуру информации.

$$I(p_1, p_2, \dots, p_n). \quad (2)$$

Предикат I при необходимости можно представить в виде системы бинарных предикатов $I_i, i = \overline{1, m}$, попарно отражающих связи между элементами данных. Число m заранее не известно, оно определяется конкретной задачей (во-первых, самим предикатом I , во-вторых - процессом бинаризации предиката I , т.к. бинаризация практически всегда требует введения дополнительных переменных).

Система бинарных предикатов $I_i, i = \overline{1, m}$ может быть представлена в виде реляционной сети для параллельной обработки.

На третьем уровне на основе взаимосвязанной структурированной информации определяется и накапливается знание. На данном уровне знание имеет полезное значение в смысле его дальнейшего использования. Однако интеграция знания с целью создания нового знания на данном уровне не рассматривается.

В соответствии с предлагаемым подходом, накопление знаний на основе данных и информации о них можно представить формально как добавление новых предикатов, заданных на множестве исходных данных $\{p_1, p_2, \dots, p_n\}$, и создание системы

$$\{I_i, i = \overline{1, k}\} \quad (3)$$

предикатов, каждый из которых определяет информацию о данных $\{p_1, p_2, \dots, p_n\}$.

Уровень метазнаний объединяет все предыдущие, позволяя находить принципиально новое понимание из существующего знания. Уровень метазнаний представим в виде предиката второго порядка, заданного на множестве $\{I_i, i = \overline{1, k}\}$.

$$M(I_1, I_2, \dots, I_k), \quad (4)$$

который связывает всю формализованную информацию о данных $\{p_1, p_2, \dots, p_n\}$ и позволяет вычислять новую информацию.

Следующее свойство знания, которое используется в естественном интеллекте, заключается в представлении знания как процесса, в отличие от знания как объекта в системах искусственного интеллекта.

Взаимосвязь данных, информации и знаний в явной и неявной формах определяют процессный аспект знаний.

Знание при процессном подходе является не только объектом, но и может выступать в роли управляющего воздействия.

Так, источник знаний представляет свои неявные знания в структурированном виде, в форме информации.

В процессе структуризации используется явное знание – например, о требуемой структуре и виде представления информации. Полученная информация содержит в себе знания в скрытом виде. Она может быть преобразована в знания лишь после ее интерпретации с использованием дополнительных знаний – как неявных, так и явных.

Приемник знания выполняет интерпретацию связей между данными, получая информацию, затем интерпретацию шаблонов, получая неявные знания.

На последнем (опциональном) этапе путем формализации полученных ранее шаблонов выполняется преобразование знания из неявной в явную форму.

Данный этап завершает процесс передачи знаний в естественном интеллекте. Полученные явные знания в дальнейшем могут быть использованы в системах ИИ.

Важная особенность рассмотренного процесса передачи знаний заключается в следующем. Для того чтобы два человека смогли передать друг другу неявные знания, они должны обладать коллективным знанием (как явным, так и неявным). Это означает, что их системы структуризации и интерпретации знаний должны соответствовать друг другу.

Многоуровневый предикатный подход к выделению неявных связей

Предлагаемый подход формально представлен в виде системы предикатов (5), которая охватывает

три ключевых составляющих, отражающих процессный характер формирования и трансформации знания.

$$\{I_i, i = \overline{1, k}; I'_j, j = \overline{1, l}\}, \quad (5)$$

$$M^*(I_1, I_2, \dots, I_k, I'_1, I'_2, \dots, I'_l),$$

где модель (4) дополняется системой предикатов $I'_j, j = \overline{1, l}$, формализующей неявные знания, а модель (3) дополняется включением в число переменных предиката M системы предикатов $I'_j, j = \overline{1, l}$.

В соответствии с предлагаемым подходом, первоначально выделяются взаимосвязи между данными и информацией, затем – между информацией и явными знаниями, затем – между неявными и явными знаниями.

Ключевое отличие предлагаемого подхода заключается в том, что основное внимание уделяется механизму выявления неявных знаний путем преобразования информации и ранее полученных явных знаний.

При таком подходе явные знания представляются преимущественно в декларативной форме, а неявные выделяются в процедурной, либо в виде последовательности причинно-следственных зависимостей.

Отмеченное отличие связано с тем, что неявные знания обычно связаны с естественным интеллектом и в большинстве случаев отражают опыт и приобретенные навыки носителя.

Предикатная модель процесса выявления неявных зависимостей, которая разработана в соответствии с данным подходом, включает в себя следующие преобразования:

- данных в информацию;
- информации в явные знания;
- данных и явных знаний в неявные знания в форме предикатов.

При первом преобразовании, данных в информацию, исходные данные представляются в виде наборов разнородных исходных элементов $p_i, q_i, r_i, i = \overline{1, K}$, которые подлежат агрегированию, уточнению, дополнению и преобразованию в информацию:

$$\{p_1, p_2, \dots, p_k\}, \quad (6)$$

$$\{q_1, q_2, \dots, q_k\}, \dots, \{r_1, r_2, \dots, r_k\}.$$

Преобразование осуществляется путем интерпретации на основе использования явных знаний. Последние могут иметь различные формы представления, однако в обобщенном виде их целесообразно представить в виде набора предикатов, описывающих граф преобразования данных в информацию.

Графовое представление обладает тем преимуществом, что сочетает декларативное описание знаний с элементами процедурного.

Прохождение по ветвям графа моделирует процесс преобразования данных в информацию.

Граф преобразования G_I^D представляется в виде следующего множества

$$G_I^D = \{V_I^D, E_I^D\}, \quad (7)$$

где $V_I^D = \{v_j\}$ – множество вершин графа G_I^D ,

D – данные,

I – информация,

v_j – j -я вершина графа преобразования данных

в информацию G_I^D ,

$E_I^D = \{e_k\}$ – множество дуг (edges) графа G_I^D .

Каждая дуга графа e_k , соединяющая вершины графа v_i и v_j , при построении его алгебологической модели описывается бинарным предикатом вида:

$$P_k(x_i, x_j)=1, \quad (8)$$

где x_i, x_j – атрибуты вершин графа v_i и v_j , которые представляют собой символьные переменные, заданные на областях определения данных в вершинах графа G_I^D .

Дуги графа G_I^D , заданные предикатами алгебры конечных предикатов, обладают следующей особенностью.

Предикаты из множества $\{P_k\}$, $k=\overline{1, m}$, где m – число дуг графа, задаются на множестве допустимых значений, которое представляется произвольным набором символов.

Поэтому такой предикат может задать условное действие по преобразованию, которое моделируется дугой графа G_I^D

Выводы

Формализация используемых человеком неявных знаний приводит нас к построению системы явных знаний. Для систем искусственного интеллекта последовательность преобразований будет иной: информация – явные знания – формализация дополнительных неявных знаний, поскольку мы не можем использовать неявные знания непосредственно в силу их природы – неотделимости от человека. В то же время мы имеем возможность выделить и формализовать неявные знания по «следам использования» уже имеющихся явных знаний.

Очевидно, что явные знания об одном и том же объекте предметной области могут отличаться. Отличия между этими знаниями вызваны неявной составляющей, которая пока еще не формализована. Например, отличие между фрагментами знаний об одном объекте могут быть связаны с личностью исполнителя, который пополнял базу знаний интеллектуальной системы.

Список литературы

1. Бондаренко М.Ф. Теория интеллекта [Текст]: учеб. / М.Ф. Бондаренко, Ю.П. Шабанов-Кушнаренко. – Х.: Компания СМИТ, 2006. – 576 с.
2. Нонака, И. Компания - создатель знания. Зарождение и развитие инноваций в японских фирмах [Текст] / И. Нонака, Х. Такеучи – М., 2003. – 384 с.
3. Reber, A.S. Implicit learning and tacit knowledge: An essay on the cognitive unconscious [Текст] / A.S. Reber. – NY: Oxford University Press, 1993. – 190 с.
4. Ackoff, R.L. From Data to Wisdom / R.L. Ackoff. // Journal of Applied Systems Analysis. – 1989. – Vol. 16. – P. 3-9.
5. Aalst van der W. Process Mining: Discovery, Conformance and Enhancement of Business Processes / W.M.P. van der Aalst. – N.-Y., Springer Verlag, 2011. – 370 p.

Поступила в редколлегию 25.03.2015

Рецензент: д-р техн. наук, проф. С.Ф. Чалый, Харьковский национальный университет радиоэлектроники, Харьков.

ПРО ПРЕДИКАТНІ МОДЕЛІ НЕЯВНИХ ЗНАНЬ В ЗАДАЧАХ АНАЛІЗУ ІНФОРМАЦІЙНИХ ПРОЦЕСІВ

О.В. Калиниченко, С.Ю. Шабанов-Кушнаренко, А.В. Ярмач

Удосконалено предикатну модель виділення неявних залежностей при трансформації явного представлення знань. Дана модель заснована на представленні явних знань у вигляді систем бінарних предикатів і визначає послідовність логічних операцій з перетворення пари графів. Формалізоване перетворення даних в неявні знання забезпечує теоретичну основу для виявлення неявних знань і побудови більш достовірного подання знань в системах штучного інтелекту.

Ключові слова: предикатна модель, неявне знання, інженерія знань, інформаційний процес

ON PREDICATE MODEL OF TACIT KNOWLEDGE IN THE PROBLEMS OF THE INFORMATION PROCESSES ANALYSIS

O.V. Kalynychenko, S.Yu. Shabanov-Kushnarenko, A.V. Yarmak

Improved model selection predicate implicit dependencies in the transformation of explicit knowledge representation. This model is based on the representation of explicit knowledge in the form of systems of binary predicates and determines the sequence of logical operations to transform the pair of graphs. Formalized converting data into tacit knowledge provides the theoretical basis for the reveal-ment of tacit knowledge, and build a more reliable knowledge representation in artificial intelligence systems.

Keywords: predicate model, tacit knowledge, knowledge engineering, information process