

ПРИМЕНЕНИЕ СЕМАНТИЧЕСКОГО АНАЛИЗА СОЦИАЛЬНЫХ СВЯЗЕЙ В БАНКОВСКИХ СКОРИНГОВЫХ СИСТЕМАХ

А.Ю. Кайда

(г. Томск, Томский политехнический университет)

E-mail: anastasiakaida@gmail.com

THE USE OF SEMANTIC ANALYSIS FOR SOCIAL TIES IN THE BANKING SCORING SYSTEM

A.Y. Kaida

(Tomsk, Tomsk Polytechnic University)

Annotation: This article describes the expert system for credit scoring systems based on ontological data model. The standards of Semantic Web with the fact extractor for analysis of social ties are a powerful tool for increasing the percentage of refundable loans as the add-on to the traditional system.

Введение. В 1941 г. Дэвид Дюран впервые применил методику классификации кредитов на «хорошие» и «плохие». Именно его методы породили создание скоринговых таблиц. Однако учитывая только статистику с 1991 г. львиная доля банков прекратила свое существование, обанкротившись. Скоринговые карты не позволяют представить полную картину о заемщике, а лишь дают базовые представления, оценка которых может привести к тому, что кредит будет выдан неплатежеспособному по тем или иным причинам клиенту банка. Классическая система работает таким образом: правильно заполненная анкета отправляется из подсистемы хранения в подсистему анализа, далее подвергается нескольким этапам анализа: стоп-факторы, черные списки, скоринг и проверка кредитной истории [2]. Сокращение процента «плохих» кредитов возможно за счет увеличения количества различных факторов, влияющих на принятие решения банком. Стоит учитывать, что задачи на сильно связанных данных или графовые задачи – один из классов трудно решаемых на реляционной модели задач. Попытки их решения на реляционной модели приводят к непредсказуемому количеству соединений в запросах, однако с этим прекрасно справляется онтология – формализованное представление знаний предметной области в виде связанных между собой экземпляров классов по принципу «субъект-свойство-объект».

Преимущества онтологической системы. В первую очередь, онтологическая модель данных является более гибкой и структурированной чем реляционная. Она более проста в редактировании, а объекты реального мира представляются в доступной для машинного распознавания форме на базе RDF (Resource Description Framework) [3]. Поскольку сегодня SQL (Structured Query Language) остается единственным механизмом связи между прикладным программным обеспечением и базой данных, архитектура данной системы предусматривает перенос данных из базы данных в онтологию посредством интеграционного модуля, где каждый столбец в базе является свойством определенного понятия, а каждая строка – экземпляром [4]. Онтологии поддерживают язык логического вывода SPARQL (SPARQL Protocol and RDF Query Language), который в совокупности с языками правил SWRL (Semantic Web Rule Language) [5] или RIF (Rule Interchange Format) [6] является мощным инструментом для создания экспертной системы по оценке надежности заемщика посредством выявления неявных или же комплексных критериев. Неявные критерии выявляются посредством анализа всех ранее подававших заявку на получения кредита в том или ином банке заемщиков, как получивших одобрение, так и получивших отказ. Среди каждой из групп определяются характерные, устойчивые тенденции, не являющиеся, как правило, четко прописанным пунктом скоринговой карты. Анализ RDF-триплетов с применением RIF-документа позволит создать новые триплеты, который будет распознаваться системой как новая информация для анализа. Простейший пример правила показан в формуле (1).

$$\begin{aligned} & \text{Forall } ?x ?y ?z (?x[\text{ex: fatherOf} \rightarrow ?z]: - \\ & \quad \text{And } (?x[\text{ex: brotherOf} \rightarrow ?y] \\ & \quad \quad ?y[\text{ex: uncleOf} \rightarrow ?z])) \end{aligned} \quad (1)$$

Анализ социальных связей. Огромное количество информации, присутствующее в социальных сетях, находится в открытом доступе и не защищено никакими настройками приватности. Использование модуля RCO Fx Ru [7], который является лингвистическим анализатором текстов на русском языке, позволяет вычлени из страницы пользователя его социальные связи для последующего анализа. Библиотека модуля осуществляет полный синтактико-семантический разбор русского текста. В состав лингвистического обеспечения пакета, помимо общих словарей и правил русского языка, входят правила выделения специальных объектов (дат, адресов, документов, телефонов, денежных сумм, марок автомобилей и пр.), шаблоны для распознавания различных классов событий и фактов (сделок, экономических показателей, конфликтов, биографических фактов и пр.). Применение данной технологии позволит заполнить онтологическую модель информацией о социальных связях и контактных данных, что сделает возможным оценку платежеспособности и надежности потенциального заемщика по множеству явных (есть ли у него также близкие родственники-должники, о которых ранее не была указана информация; есть ли задержки по заработной плате на его работе и т. п.) и неявных критериев (стабильность работ, соразмерность роста потребления и карьерного роста и т. д.).

Заключение. Разработка основывается на результатах исследований, проведенных Чернием А.В. [8] и Губиным М.Ю. [9] в рамках программы «УМНИК». Эти исследования создали основу для реализации такой системы в области финансов, где данный подход ранее не применялся.

На данном этапе разработка проводится и тестируется в условиях лаборатории. В дальнейшем планируется внедрение системы в режиме бета-теста в один из банков. На данный момент имеются несколько заинтересованных финансовых учреждений.

Список литературы

1. Модель кредитного скоринга Дюрана [Электронный ресурс]. – Режим доступа: http://afdanalyse.ru/publ/finansovyj_analiz/1/model_kreditnogo_skoringa_djurana/16-1-0-140, свободный
2. Прекратившие существование банки [Электронный ресурс]. – Режим доступа:
3. <http://www.banki.ru/banks/memory/>, свободный
4. Resource Description Framework [Электронный ресурс]. – Режим доступа: <http://www.w3.org/RDF/>, свободный
5. Кайда А.Ю., Черний А.В. Интеграция гетерогенной информации в кредитно-скоринговых системах // Молодой ученый. – 2014 – № 21 (80). – С. 20–25.
6. SWRL: A Semantic Web Rule Language Combining OWL and RuleML [Электронный ресурс]. – Режим доступа <http://www.w3.org/Submission/SWRL/>, свободный
7. RIF Overview (Second Edition) [Электронный ресурс]. – Режим доступа: <http://www.w3.org/TR/rif-overview/>, свободный
8. RCO Fact Extractor SDK [Электронный ресурс]. – Режим доступа: http://www.rco.ru/product.asp?ob_no=5047, свободный
9. А.В. Черний, А.Ф. Тузовский, Semantic Web масштаба организации // Журнал «Информационные технологии» № 6 2010 г., Издательство Новые технологии. – С. 50–55
10. Губин М.Ю., Разин В.В., Тузовский А.Ф. Методы создания семантических метаописаний документов с применением семантических сетей, фреймовых моделей и частотных характеристик // Доклады ТУСУРа, № 2 (22), часть 2, декабрь 2010