

ПРИМЕНЕНИЕ ОНТОЛОГИЙ В СКОРИНГОВЫХ СИСТЕМАХ КРЕДИТОВАНИЯ

А. Ю. Кайда, А. В. Черный, А.Ф. Тузовский
Томский политехнический университет
anastasiakaida@gmail.com

Введение

В 1941 г. Дэвид Дюран впервые применил методику классификации кредитов на «хорошие» и «плохие». Он определил не только группы факторов, позволяющие максимально определить степень кредитного риска, но и коэффициенты, характеризующие кредитоспособность частного клиента [1]. Однако учитывая только статистику с 1991 г. лавина доля банков прекратила свое существование, обанкротившись [2].

Методика Дюрана, породившая применение скоринговых карт, не исключает тот факт, что сами карты не позволяют представить полную картину о заемщике, а лишь дают базовые представления, которые могут привести к тому, что кредит выдан неплатежеспособному по тем или иным причинам клиенту банка. Классическая система работает таким образом: правильно заполненная анкета отправляется из подсистемы хранения в подсистему анализа, где подвергается нескольким этапам проверки. Такими этапами являются стоп-факторы, черные списки, скоринг и проверка кредитной истории [3]. Сокращение процента «плохих» кредитов возможно за счет увеличения количества различных факторов, влияющих на принятие решения банком. С попыткой решения этой проблемы возникает лавинообразный рост обрабатываемых объемов данных на фоне стремительного увеличения уровня загрузки систем хранения. Происходит «обрастание» конкретного лица различного рода факторами и критериями, некоторые из которых хоть и являются неявными, но вполне очевидными и важными. Для того чтобы извлечь из имеющейся информации максимум пользы, а также сделать данные более связанными и согласованными необходима более гибкая, связанная и распределенная модель данных. Именно такими преимуществами перед реляционными базами данных и обладает онтология. Также она позволяет учесть не только явные, но и неявные критерии путем логического вывода в запросах и ряда характеристик свойств.

Методы повышения точности скоринговых систем за счет онтологий

■ Онтология, также называемая семантической моделью, формализует знания в виде иерархии понятий в области определения, используя обширный словарь для обозначения типов, свойств и взаимосвязи этих понятий [4]. Главное ее преимущество — это представление

объектов реального мира в форме, доступной для машинного распознавания. Для решения этих задач используется язык OWL (Ontology Web Language) — стандарт W3C. Основой данного языка является представление данных в виде «объект-свойство»[5].

■ На сегодняшний день SQL остаётся единственным механизмом связи между прикладным программным обеспечением и базой данных. Архитектура системы предусматривает перенос данных из базы данных в онтологию посредством интеграционного модуля, основной принцип работы которого показан на рисунке 1. Каждый столбец в базе является свойством определенного понятия, а каждая строка — экземпляром [6].

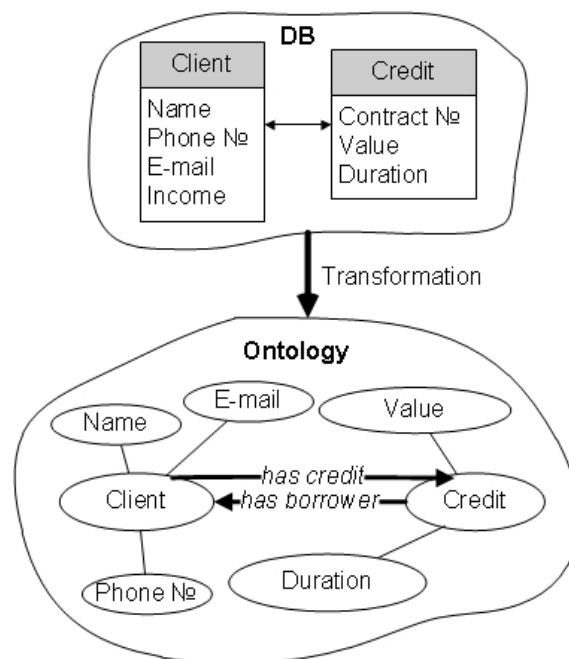


Рис. 1. Интеграционный модуль

Критерии и правила

Помимо явных критериев при одобрении кредита для конкретного заемщика предложена концепция использования также неявных критериев. Примером необходимости такого критерия может служить заемщик, постоянно меняющий место работы. Какова вероятность того, что он, взяв кредит сегодня, не оставит свое рабочее место завтра? Насколько важно учитывать тот факт, что некоторый заемщик при отсутствии высшего образования имеет высокооплачиваемую работу, однако работает на сомнительную организацию и

по определению не может рассматриваться как востребованный на рынке труда сотрудник?

Еще одной необходимостью являются комплексные критерии. Данные критерии создаются на основе положительных тенденций после многочисленного повторения одного и того же случая или модели поведения заемщика. Примером могут служить наличие богатого родственника или сезонный «кризис» работников сферы туризма и фермеров. Такого рода критерии способны значительно расширить возможности грамотной оценки платежеспособности заемщика.

Еще одной особенностью данной разработки являются неявные критерии, которые, в свою очередь, выявляются посредством анализа всех прошедших через банк клиентов, как успешно взявших кредит, так и получивших отказ. В каждой из двух групп имеются характерные, устойчивые тенденции, не являющиеся, как правило, четко прописанным пунктом традиционной скоринговой карты, однако их выявление позволит с большей точностью определять надежность заемщика. При большом числе повторений этой тенденции в одной группе и критически низком — в другой можно будет определить ее как неявный критерий для той или иной группы заемщиков.

Все это необходимо идентифицировать и зафиксировать для улучшения планирования стратегии банка. Использование языков правил SWRL [7] и RIF [8] специфицирует условия и ситуации, что дает кредитору представление о платежеспособности заемщика более точно, чем классическая скоринговая карта. SWRL — семантический язык правил, использующий логические правила вкупе с OWL DL. Также он имеет обширный перечень встроенных модулей [9] для создания более сложных структур. Простейший пример правил показан в формулах (1) и (2).

$$\text{Person}(?x)^{\text{isInBlackList} = 1} \rightarrow \text{UnreliablePayer}(?x) \quad (1)$$
$$\text{Person}(?x)^{\text{hasCredit}(?x, ?c)} \cap \text{Credit}(?c)^{\text{is}(?c, \text{UnrepaidLoan}(?c))} \rightarrow \text{UnreliablePayer}(?x) \quad (2)$$

Особенности разрабатываемой онтологии

1. Привязка онтологии ко времени для возможности наблюдать за изменениями в режиме реального времени и выявления определенных тенденций. Применяется стандарт W3C — OWL-Time [10].
2. Использование семантических языков правил для закрепления выявленных закономерностей и возможности их распознавание системой. Модуль обработки правил.
3. Взаимодействие с социальными сетями. Концепция создания собственного классификатора на основе Наивного Байесовского Классификатора

(Naive Bayes) [11], использующего для анализа в качестве исходного материала хештэги и авторские комментарии.

4. Аннотационный модуль, дающий возможность дополнить онтологию информацией на естественном языке для понимания пользователя.

5. Обмен информацией между банками и ограничение доступа к данным различных групп лиц.

Заключение

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.

Литература

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: <http://www.banki.ru/banks/memory/>, свободный
3. Компоненты кредитно-скоринговой системы [Электронный ресурс]. - Режим доступа: http://www.basegroup.ru/solutions/case/banks/system_components/, свободный
4. T.Gruber(June 1993). "[A translation approach to portable ontology specifications](#)" (PDF). [Knowledge Acquisition](#) 5(2):199–220.[doi:10.1006/knac.1993.1008](#).
5. Allemang D., Hendler J., Semantic Web for The Working Ontologist Modeling in RDF, RDFS and OWL – Morgan Kaufmann Publishers, 2008. - pp. 14-50.
6. Beaulieu, Alan (April 2009). Mary E Treseler, ed. Learning SQL (2nd ed.). Sebastapol, CA, USA: O'Reilly. ISBN 978-0-596-52083-0.
7. SWRL: A Semantic Web Rule Language Combining OWL and RuleML [Электронный ресурс]. - Режим доступа: <http://www.w3.org/Submission/SWRL/>, свободный
8. RIF Overview (Second Edition) [Электронный ресурс]. - Режим доступа: <http://www.w3.org/TR/rif-overview/>, свободный
9. Built-ins for SWRL [Электронный ресурс]. - Режим доступа: <http://www.daml.org/2004/04/swrl/builtins.html>, свободный
10. Time Ontology in OWL [Электронный ресурс]. - Режим доступа: <http://www.w3.org/TR/owl-time/>, свободный
11. A. Pak, P. Paroubek: Twitter as a Corpus for Sentiment Analysis and Opinion Mining, The International Conference on Language Resources and Evaluation, 2010.