

УДК 004.89



ИНТЕЛЛЕКТУАЛЬНЫЕ МЕТОДЫ ОЦЕНКИ ФИНАНСОВЫХ РИСКОВ, СВЯЗАННЫХ С ОПЛАТОЙ ЗА ОБУЧЕНИЕ В ВУЗЕ

Н.А. Дергачева¹, И.В. Шостак², А.С. Топал³

¹ НАУ им. Н.Е. Жуковского «ХАИ», г. Харьков, Украина

² НАУ им. Н.Е. Жуковского «ХАИ», г. Харьков, Украина, iv_shostak@rambler.ru

³ НАУ им. Н.Е. Жуковского «ХАИ», г. Харьков, Украина

В статье сформулированы основные задачи поиска знаний в базах данных для оценки финансовых рисков, связанных с оплатой за обучение в ВУЗе. Приведены возможные методы решения задач, основанные на применении технологий Data Mining.

ПОИСК АССОЦИАТИВНЫХ ЗНАНИЙ, ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ, ТЕХНОЛОГИИ DATA MINING, АССОЦИАТИВНЫЕ ПРАВИЛА, ДЕРЕВЬЯ РЕШЕНИЙ

Введение

В настоящее время в ВУЗах широко распространена контрактная форма подготовки специалистов, которая позволяет увеличить количество студентов и обеспечить качественный уровень работы профессорско-преподавательского и инженерно-технического составов. Применение данной формы обучения связано с рядом финансовых рисков для ВУЗов и банков, кредитующих обучение студентов, связанных с несвоевременной оплатой за обучение и погашением кредитных задолженностей. Таким образом, становится актуальным решение задачи разработки методики оценки финансовых рисков, связанных с оплатой за контрактное обучение.

Одним из важных этапов управления рисками является сбор и обработка первичных данных [1]. Полученная информация является основой применения различных технологий интеллектуального анализа данных (ИАД) [2, 3]. В последнее время получили распространение такие методы ИАД как: методы поиска ассоциативных правил [4], применение нейронных сетей [5, 6], деревьев решений [7], самоорганизующихся карт [8, 9]. Для повышения эффективности использования методов ИАД к рассматриваемым данным целесообразно применять сразу несколько алгоритмов [2]: деревья решений, нейронные сети и выводы путем сопоставления (memoory-based reasoning). Если несколько моделей приводят к одинаковым результатам, значит, найден наилучший сценарий. Это особенно полезно, когда инструмент выдает ответ, но не предлагает интуитивного объяснения (наиболее часто встречающаяся проблема нейронных сетей).

Целью статьи является обоснование актуальности и постановка задачи оценки финансового риска по характеристикам сведений о студентах, обучающихся в ВУЗах.

1. Постановка задачи исследования

Рассмотрим основные бизнес-процессы [10] в системе оплаты контрактного обучения (рис. 1). Оплата может проводиться прямыми платежами (1), при помощи кредитования (2-3), а также применением смешанной формы оплаты.

Прямые платежи производятся непосредственно студентом полностью за весь период обучения либо периодически по частям (раз в месяц, раз в триместр, раз в год). Финансовые риски в этом случае связаны с несвоевременным внесением оплаты за обучение самим студентом. Если студент оплачивает всю сумму за обучение сразу, то финансовый риск в процессе обучения студента отсутствует. Если оплата осуществляется по частям, то финансовый риск распределяется на весь период обучения.

Студент может получить банковский кредит на обучение как физическое лицо (2). В большинстве случаев для получения кредита необходим поручитель, а финансовым рискам подвергаются банки, выдавшие кредит на обучение.

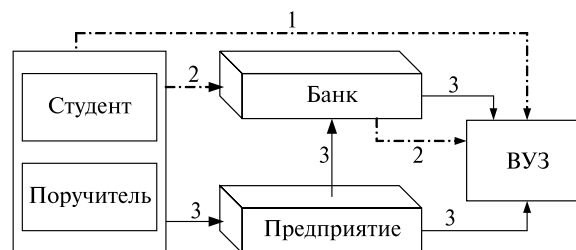


Рис. 1. Бизнес-процессы при оплате за обучение по контракту

При оформлении кредита на обучение студента могут привлекаться предприятия с различной формой собственности (3). При этом заключается договор между ВУЗом, студентом и предприятием на получение кредита. Предприятие выступает в качестве поручителя и оплачивает обучение студента в соответствии с условиями договора. Финансовым рискам подвергается непосредственно ВУЗ как субъект хозяйственной деятельности.

Анализ бизнес-процессов в системе оплаты за контрактное обучение показал, что финансовые риски при различных формах оплаты имеют различные источники, степени и величины риска, вероятности наступления рисков событий, следовательно, необходимы различные приемы управления риском и способами его снижения. Таким образом, оценку финансовых рисков необходимо проводить для каждого варианта оплаты за обучение в отдельности.

Процесс управления финансовым риском можно разделить на два этапа: этап анализа риска; этап принятия мер по устранению и минимизации. Анализ финансового риска включает: сбор и обработку данных по аспектам риска, качественный и количественный анализ риска.

Меры по устранению и минимизации финансового риска включают: выбор и обоснование предельно допустимых уровней риска, выбор методов снижения риска, формирование вариантов рискованного вложения капитала, оценку их оптимальности на основе сопоставления ожидаемой прибыли и величины риска.

При оценке финансовых рисков может быть применена технология поиска знаний в базах данных Data Mining (DM). Технология DM заняла свои позиции в последнее десятилетие, получив центральную роль во многих сферах бизнеса. Эта технология позволяет существенно повысить возможности организации в достижении целей. Ее популярность растет все больше, поскольку инструменты совершенствуются, получают все более широкое применение, дешевеют и становятся проще в использовании.

В основу технологии поиска знаний в данных положена концепция шаблонов (образов).

Эти шаблоны представляют собой функциональные, логические и другие закономерности, свойственные подвыборкам данных и могут быть компактно выражены в понятной человеку форме. Важным требованием к разыскиваемым шаблонам является их неожиданность, нетривиальность, неочевидность (unexpected regularities). Весь процесс выделения знаний из данных принято делить на этапы, представленные на рис. 2.

Большинство авторов приводит классификацию задач Data Mining по типам производимой информации [2, 3] и выделяет четыре наиболее типичные задачи извлечения знаний из данных в виде нахождения регулярностей в больших массивах информации:

- 1) построение классификатора;
- 2) кластеризация;
- 3) вывод правил ассоциации;
- 4) восстановление моделей зависимости.

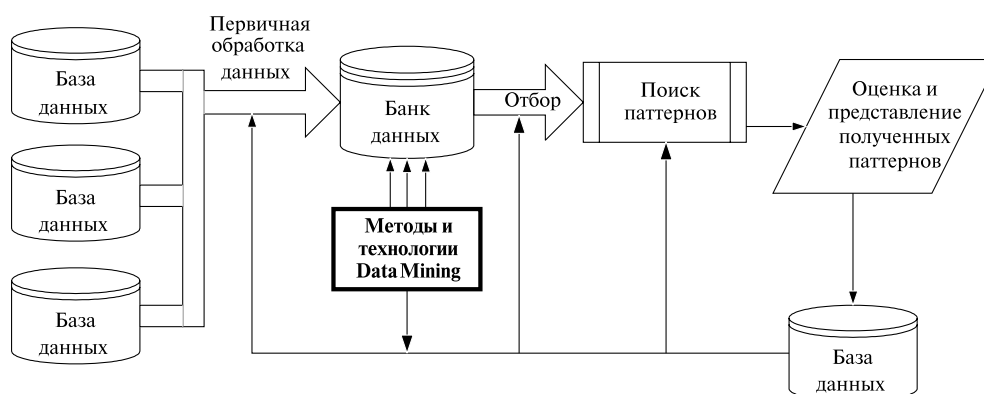


Рис. 2. Процесс обнаружения знаний в базах и хранилищах данных

Для решения реальной практической задачи, как правило, требуется применение нескольких подходов и методов.

Задача построения классификатора заключается в построении решающего правила на основании уже классифицированных объектов (обучающей выборки). С помощью решающего правила (набора правил) объект неизвестного класса может быть отнесен к какому-либо из имеющихся классов.

Под задачей кластеризации понимается разбиение множества объектов на заданное или неизвестное число классов на основании некоторого математического критерия качества классификации, это определяет непосредственную связь задачи кластеризации с задачей классификации.

Вывод ассоциативных правил заключается в нахождении шаблонов имплицитивного вида, отражающих набор взаимно сопутствующих событий. Каждое правило должно иметь меру определенности не ниже заданного уровня. Ассоциативное правило является, таким образом, логической закономерностью вида «Если...то...», свойственной подвыборке данных и обладающей статистической состоятельностью. Для поиска ассоциативных правил применяют методы ограниченного перебора или построения деревьев покрытий

Так, в случае оплаты за обучение при непосредственных прямых платежах возможно применение методов построения ассоциативных правил с целью получения ассоциативного портрета «потенциального неплательщика за обучение» на основе анализа базы данных о сроках оплаты и должниках по оплате за обучение. Или же наоборот, построение ассоциативного портрета «идеального плательщика». Подобные решения могут быть получены на основе применения нейронных сетей обратного распространения (back propagation) или самоорганизующихся карт Кохонена, которые помогут, имея информацию о некоторой из части исследуемых объектов (студентах, совершающих оплату за обучение), достаточно достоверно судить по ряду общих признаков об объектах, с которыми мы мало знакомы.

Рассматривая финансовые риски при кредитовании физических лиц для оплаты за обучение, необходимо отметить, что известна классическая задача Data Mining – задача о предоставлении кредита и ее решение методом построения дерева решений. На сегодняшний день известно достаточно много методик кредитного скоринга. Одной из самых известных является модель Дюрана. Однако известные методы принятия решения в большинстве случаев неприменимы [7] по причине плохой адаптируемости, ее высокой стоимости и большой вероятности ошибки модели при определении кредитоспособности потенциального заемщика, обусловленной субъективным мнением специалиста. Таким образом, необходим новый надежный инструмент для эффективного решения задач о предоставлении кредита, основывающийся на применении технологий Data Mining. Также, по мнению авторов, может быть интересной задача об оценке финансовых рисков при предоставлении кредита в условиях недостоверной информации (например, о доходах за последнее время).

При анализе финансового риска предоставления кредита для оплаты за обучение студента с привлечением различных предприятий, наряду с классическим решением задачи о предоставлении кредита, необходим анализ финансово-экономической деятельности предприятия, выступающего гарантом возвращения кредита за обучение студента с привлечением технологий Data Mining. На основе анализа имеющихся данных о выплатах предприятием задолженностей по выплатам, их известных финансово-экономических показателей возможно построение ассоциативных правил и на их основе – ассоциативного портрета «предприятия, не являющегося гарантом выплаты кредита» за обучение студента и наоборот «предприятия-надежного гаранта выплаты кредита».

2. Пример использования деревьев решений при оценке кредитоспособности физических лиц

В данном примере [7] рассмотрено применение метода деревьев решений как одного из вариантов устранения недостатков скоринговой системы оценки кредитоспособности заемщика. Одним из вариантов решения подобной задачи является применение алгоритмов, решающих задачи классификации, то есть задачи отнесения какого-либо объекта (в данном случае потенциального заемщика) к одному из заранее известных классов (Предоставлять/Не предоставлять кредит). Деревья решений – один из методов автоматического анализа данных. Получаемая модель – это способ представления правил в иерархической, последовательной структуре, где каждому объекту соответствует единственный узел, дающий решение.

Сущность метода дерева решений заключается в следующем:

1. На основе данных за прошлые периоды строится дерево. При этом класс каждой из ситуаций, на основе которых строится дерево, заранее известен. Для случая кредитования должно быть известно, была ли возвращена основная сумма долга и проценты и не было ли просрочек в платежах. При построении дерева все известные ситуации обучающей выборки сначала попадают в верхний узел, а потом распределяются по узлам, которые в свою очередь также могут быть разбиты на дочерние узлы. Критерий разбиения – это различные значения какого-либо входного фактора. Для определения поля, по которому будет происходить разбиение, используется показатель энтропии (мера неопределенности). Выбирается то поле, при разбиении по которому устраняется больше неопределенности. Неопределенность тем выше, чем больше примесей (объектов, относящихся к различным классам) находятся в одном узле. Энтропия равна нулю, если в узле будут находиться объекты, относящиеся к одному классу.

2. Полученную модель используют при определении класса («Предоставлять кредит»/«Не предоставлять кредит») для вновь возникших ситуаций (при поступлении заявки на получение кредита).

3. При существенном изменении текущей ситуации дерево можно (и даже нужно) перестроить, то есть адаптировать к существующей обстановке.

Для практического примера была использована программа Tree Analyzer из пакета Deductor (v.3). В качестве исходных данных была взята выборка, состоящая из 1000 записей, где каждая запись – это описание характеристик заемщика и параметр, описывающий его поведение во время погашения займа. При обучении дерева (кроме стандартных параметров) использовались следующие факторы, определяющие заемщика: ‘Размер ссуды’; ‘Срок ссуды’; ‘Цель ссуды’; ‘Среднемесячный доход’; ‘Среднемесячный расход’; ‘Основное направление расходов’; ‘Наличие недвижимости’; ‘Наличие автотранспорта’; ‘Наличие банковского счета’; ‘Наличие страховки’; ‘Отраслевая принадлежность предприятия’; ‘Срок работы на данном предприятии’; ‘Направление деятельности заемщика’; ‘Срок работы на данном направлении’; ‘Пол’; ‘Семейное положение’; ‘Возраст’; ‘Срок проживания в данной местности’; ‘Обеспеченность займа’; ‘Предоставлять кредит’. Некоторые поля (‘№ Паспорта’, ‘ФИО’, ‘Адрес’, ‘Название организации’) определены алгоритмом уже до начала построения дерева решений как непригодные по причине практически уникальности каждого из значений. Целевым полем является поле ‘Предоставлять кредит’, принимающее значения ‘Да’ и ‘Нет’.

После процесса построения дерева решений при помощи программы Tree Analyzer получена модель оценки кредитоспособности физических лиц, описывающая ситуацию, относящуюся к определенному банку. Эта модель представляется в виде иерархической структуры правил – дерева решений.

Анализируя полученное дерево решений, можно сделать следующие выводы:

1. При помощи дерева решений можно проводить анализ значащих факторов, так как при определении параметра на каждом уровне иерархии используется критерий наибольшего устранения неопределенности. Таким образом, более значимые факторы, по которым проводится классификация, находятся на более близком расстоянии (глубине) от корня дерева, чем менее значимые. Например фактор ‘Обеспеченность займа’ более значим, чем фактор ‘Срок проживания в данной местности’. А фактор ‘Основное направление расходов’ значим только в сочетании с другими факторами.

2. Такие показатели как ‘Размер ссуды’, ‘Срок ссуды’, ‘Среднемесячный доход’ и ‘Среднемесячный расход’ вообще отсутствуют в полученном дереве. Данный факт можно объяснить тем, что в исходных данных присутствует такой показатель как ‘Обеспеченность займа’, и так как этот фактор является точным обобщением этих четырех показателей, алгоритм построения дерева решений выбрал именно его.

Очень важной особенностью построенной модели является то, что правила, по которым определяется принадлежность заемщика к той или иной группе, записаны на естественном языке. Например на основе построенной модели получаются правила следующего вида:

ЕСЛИ Обеспеченность займа = Да И Срок проживания в данной местности, лет > 5.5 И Возраст > 19.5 И Наличие недвижимости = Да И Наличие банковского счета = Да ТО Предоставлять кредит = Да (Достоверно на 98%).

ЕСЛИ Обеспеченность займа = Да И Срок проживания в данной местности, лет > 5.5 И Наличие недвижимости = Да И Возраст > 21.5 И Срок работы на данном направлении, лет <= 5.5 И Пол = Муж И Наличие банковского счета = Нет И Основное направление расходов = Одежда, продукты питания и т.п. ТО Предоставлять кредит = Нет (Достоверно на 88%).

Правильно построенное на данных прошлых периодов дерево решения обладает важной особенностью — ‘способностью к обобщению’, то есть при возникновении новой ситуации (обращение потенциального заемщика) можно с большой вероятностью утверждать, что вновь обратившийся

заемщик поведет себя так же, как и те заемщики, характеристики которых схожи с характеристиками вновь обратившегося.

На основе построенной модели также можно определять принадлежность потенциального заемщика к одному из классов. Для этого необходимо воспользоваться диалоговым окном ‘Эксперимент’ программы Tree Analyzer, в котором, последовательно отвечая на вопросы, можно в итоге получить ответ на вопрос:

‘Предоставлять ли кредит’.

На основе приведенного примера можно сделать вывод о достоинствах метода деревьев решений (устраняющих принципиальные недостатки скоринговой системы):

1. Стоимость адаптации сводится к минимуму по причине того, что алгоритмы построения модели классификации (дерево решений) – это самоадаптирующиеся модели (вмешательство человека минимально)

2. Качество результата достаточно велико за счет того, что выбираются наиболее значимые факторы для определения конечного ответа, и полученный результат является статистически обоснованным.

Выводы

В статье сформулированы основные задачи поиска знаний в базах данных для оценки финансовых рисков, связанных с оплатой за обучение в ВУЗе. Приведены возможные методы решения указанных задач, основанные на применении технологии Data Mining.

Список литературы: 1. Бюджетирование, финансовое планирование и анализ, управленческий учет. – [Электрон. ресурс]. Метод доступа: http://www.cis2000.ru/pub-Ksh/books/boolc_71/page21.shtml. 2. Data Mining – добыча данных. [Электрон, ресурс]. – Метод доступа: <http://www.basegroup.ru/tasks/datamining.htm>. 4. Введение в анализ ассоциативных правил – [Электрон. ресурс]. – Метод доступа: <http://www.basegroup.ru/rulcs/intro.htm>. 5. Нейронные сети – математический аппарат. – [Электрон. ресурс]. – Метод доступа: <http://www.basegroup.ru/neural/math.htm>. 6. Нейронные сети как средство добычи данных – [Электрон. ресурс]. – Метод доступа <http://www.basegroup.ru/neural/ns.htm>. 7. Использование деревьев решений для оценки кредитоспособности физических лиц – [Электрон. ресурс]. – Метод доступа <http://www.basegroup.ru/practice/solvency.htm>. 8. Самоорганизующиеся карты – математический аппарат – [Электрон. ресурс]. – Метод доступа: <http://www.basegroup.ru/neural/som.htm>. 9. Пастухов Е.С., Кутыин В.М. Методика оценки изменений в банковской среде на основе технологии самоорганизующихся карт признаков – [Электрон. ресурс]. – Метод доступа <http://www.bankclub.ru/seminar-article.htm>. 10. Торнтуэйт У. Как начать внедрение технологии data mining, не теряя ее ценных возможностей – [Электрон. ресурс]. – Метод доступа: <http://citcity.ru/12996>.

Поступила в редколлегию 12.10.2007