



УДК 519.872



Евгений Станиславович СЕМЕНКИН,
профессор кафедры системного анализа
и исследования операций Сибирского
государственного аэрокосмического университета
им. акад. М.Ф.Решетнева (г. Красноярск),
доктор технических наук, профессор



Сергей Николаевич ЕФИМОВ,
профессор кафедры социально-экономических
наук и информатики Сибирского юридического
института МВД России (г. Красноярск),
кандидат технических наук, доцент



Антон Александрович МЕЛЬНИКОВ,
преподаватель-методист отдела технических
средств обучения Сибирского юридического
института МВД России (г. Красноярск)



Дмитрий Сергеевич СТЕЛЬМАШ,
начальник сектора безопасности и защиты
безопасности Емельяновского отделения №6190
Сбербанка России (п. Емельяново
Красноярского края)

ВЫБОР МЕТОДА АВТОМАТИЗАЦИИ ПРИНЯТИЯ РЕШЕНИЙ ПРИ КРЕДИТОВАНИИ ФИЗИЧЕСКИХ ЛИЦ

THE CHOICE OF A METHOD OF AUTOMATION OF DECISION-MAKING AT CREDITING PHYSICAL PERSONS

Статья посвящена проблеме выбора метода автоматизации принятия решений при кредитовании физических лиц в банковской сфере. Рассмотрены различные технологии интеллектуального анализа данных, приведены практические результаты решения задач.

The article is devoted to the problem of a choice of a method of automation of decision-making at crediting physical persons in bank sphere. Various technologies of the intellectual analysis of data are considered, practical results of problem solving are described.

Ключевые слова: интеллектуальный анализ данных, искусственные нейронные сети, генетическое программирование.

Keywords: intellectual analysis of data, artificial neural networks, genetic programming.



Системы интеллектуального анализа данных (ИАД) – одно из наиболее перспективных направлений развития информационных технологий. Разработка систем ИАД является актуальной научно-технической задачей для таких областей, как управление сложными технологическими процессами, прогнозирование развития природных и антропогенных катастроф, управление безэкипажными боевыми машинами, разработка систем глобальной безопасности, обработка экспериментальных данных и др.

ИАД предъявляет чрезвычайно высокие требования к квалификации разработчиков и применяемым вычислительным средствам.

Проблема недостатка квалифицированных специалистов может быть в значительной степени ослаблена с помощью автоматизации процесса разработки систем интеллектуального анализа данных для конкретных приложений. В этом случае основные алгоритмы будут генерироваться автоматически и адаптивно настраиваться на решаемую задачу. Основную роль будут играть специалисты в прикладной области, а не разработчики информационных систем. Тем самым разработка систем интеллектуального анализа данных значительно ускорится, а затраты на ее разработку снизятся.

Информационные технологии интеллектуального анализа данных

Основная задача интеллектуального анализа данных – извлечение скрытой информации из баз данных, накопленных в конкретной предметной области. Закономерности процессов предметной области могут быть извлечены из базы данных, содержащей результаты экспериментов или наблюдений за поведением изучаемого объекта, с помощью специальных информационных технологий. Основные интеллектуальные информационные технологии, применяемые при таком подходе, – это искусственные нейронные сети (ИНС), системы на нечеткой логике (НЛС), метод генетического программирования (ГП).

ИНС являются самообучающимися системами, которые при выбранной структуре (число слоев нейронов, число нейронов на каждом слое, типы функций активации, виды синаптических связей между нейронами) и имеющемся «задачнике» в состоянии автоматически настроить весовые коэффициенты синапсов таким образом, чтобы в дальнейшем при предъявлении новых входов генерировать релевантные выходы. Это дает возможность решать задачи аппроксимации, классификации, распознавания, управления и др. Положительной чертой ИНС является их способность к самообучению. Недостатком с точки зрения интеллектуального анализа данных является то, что ИНС являются «черным ящиком», то есть их работа не может быть объяснена в явном виде (вся сила ИНС скрыта в структуре и наборе весовых коэффициентов). С точки зрения удобства разработки ИНС их недостатком является то, что выбор подходящей структуры требует высокой квалификации конечного пользователя, чего трудно ожидать на практике от специалистов конкретных предметных областей. С точки зрения аппаратной реализации недостатком ИНС является то, что они обычно проектируются избыточными (многослойными с послойной организацией связей «от каждого к каждому»), а значит, излишне громоздкими. Таким образом, для эффективного применения ИНС необходимо разрабатывать методы, позволяющие автоматизировать проектирование ИНС произвольной конфигурации (нейроны, функции активации, конфигурация связей и др.), минимальные по сложности и оптимальные по точности решения задач, с обеспечением возможности извлечения знаний в явном виде.

В данном направлении получены следующие результаты:

– разработан, реализован и проверен на классических тестовых и реальных задачах подход, позволяющий извлекать из обученной нейронной сети произвольной конфигурации знания в виде правил «если... то ...иначе»¹;



– разработаны, реализованы и проверены на тестовых и реальных задачах системы автоматизации нейросетевого моделирования² на основе генетических алгоритмов, автоматически генерирующих оптимальную структуру ИНС с произвольной организацией функционирования и выбором функций активации для каждого нейрона, а также настраивающих весовые коэффициенты синаптических связей;

– разработана, реализована и проверена на тестовых и реальных задачах система автоматического проектирования и настройки нейронных сетей на основе алгоритма генетического программирования, генерирующего структуру ИНС, и генетического алгоритма, настраивающего весовые коэффициенты.³

Системы на нечеткой логике являются интеллектуальными информационными системами, содержащими знания об объекте или явлении в явном виде. Эти знания сформулированы в базе знаний в виде правил-продукций на основе лингвистических переменных и нечеткой логики. Явный вид знаний, предъявляемых пользователю НЛС, является положительной чертой данного подхода. С точки зрения интеллектуального анализа данных недостатком НЛС является необходимость взаимодействия между экспертом предметной области и разработчиком, в ходе которого создается база знаний и формализуются правила принятия решений. Основное время разработки НЛС уходит на то, чтобы эти специалисты обучили друг друга своей профессии. К концу разработки инженер по знаниям становится «квази-экспертом» в предметной области, а эксперт становится почти профессионалом в интеллектуальных информационных технологиях. Создание НЛС – это не только дорогостоящая, но и чрезмерно длительная процедура. Автоматизация настройки НЛС затруднительна, если взаимодействуют люди-специалисты. С точки зрения аппаратной реализации НЛС также имеет недостаток избыточности, а значит, и громоздкости получаемых устройств. В связи с этим для эффективного использования

НЛС необходимо разрабатывать методы автоматизации проектирования НЛС, обеспечивающие генерирование релевантной и оптимальной по составу базы знаний и тонкую настройку лингвистических переменных.

Результатами, полученными в данной области, являются:

– система автоматизированного проектирования баз знаний НЛС, содержащая алгоритм генерирования базы правил, методику оценки эффективности базы правил и алгоритм оптимизации базы правил⁴, прошедшая апробацию на реальных практических задачах;

– программная система автоматического формирования нечеткого логического контроллера⁵, автоматически генерирующая базу правил с помощью нейронных сетей и генетического алгоритма и настраивающая семантику лингвистических переменных с помощью алгоритма оптимизации;

– система эволюционного формирования экспертных систем на нечеткой логике, генерирующая обобщенные правила с помощью алгоритма генетического программирования и настраивающая семантику с помощью генетического алгоритма оптимизации⁶, прошедшая апробацию на тестовых и реальных практических задачах.

Алгоритм генетического программирования является интеллектуальной информационной технологией, способной автоматически строить математические модели в аналитическом виде, решая задачу символьной регрессии. Положительной для ИАД чертой алгоритмов ГП является то, что они в состоянии генерировать истинные функциональные зависимости (при правильной методике применения), неизвестные до начала работы. Недостатком ГП является его трудоемкость, а также то, что его эффективность существенно зависит от выбора настроек, который трудно сделать даже специалисту. Итак, для эффективного применения ГП необходимо установить правильную методику его применения, позволяющую находить истинные или близкие к истинным функ-



циональные зависимости и дополнить ГП механизмом самонастройки.

В данной области получены следующие результаты:

– системы решения задачи символьной регрессии алгоритмом генетического программирования с адаптивной настройкой коэффициентов функциональной зависимости стандартным и модифицированным генетическим алгоритмом⁷, позволяющие за счет точного выбора коэффициентов получать с большой вероятностью точные математические модели;

– коэволюционный алгоритм генетического программирования⁸, обладающий свойством самонастройки параметров, позволяющим его применять конечному пользователю, не имеющему глубоких знаний теории эволюционного моделирования, прошедший апробацию на тестовых и реальных практических задачах.

Описанные подходы к разработке инструментария для интеллектуального анализа данных и поддержки принятия решений при управлении сложными системами (по отдельности и в различных сочетаниях) успешно применялись при выполнении конкретных прикладных разработок в области авиационной и космической техники⁹, управления технологическими процессами¹⁰ и в других технических областях¹¹, а также в областях психологии и медицины¹² и управления инновационной и инвестиционной деятельностью¹³.

Решение задачи автоматизации принятия решений при кредитовании физических лиц

Основной долей доходов банка являются доходы, полученные от кредитования. Выделяют два вида кредитования: кредитование частных клиентов и кредитование юридических лиц. Если при кредитовании юридических лиц довольно просто оценить возможные кредитные риски, связанные с невозвратом заемных средств банку, то при кредитовании физических лиц данную оценку провести без

помощи автоматизированной системы довольно сложно. В каждом отделении Сберегательного банка РФ для принятия решения о выдаче кредита физическому лицу созывается кредитный комитет и коллегиально принимается положительное или отрицательное решение о выдаче кредита.

В условиях современного экономического кризиса не только Сбербанку РФ, но и любому другому банку принципиально важно минимизировать любые возможные риски, связанные с кредитованием, а также уменьшить материальные и производственные затраты, связанные с судебными тяжбами по возврату заемных средств.

Чтобы уменьшить нагрузку на сотрудников, увеличить скорость рассмотрения кредитных заявок и минимизировать невозврат заемных средств при выдаче потребительских кредитов, разработана автоматизированная система поддержки принятия решений (СППР) кредитного скоринга.¹⁴

Скоринговая система помогает на этапе приема документов от потенциального заемщика на основе полученных о нем данных оценивать вероятность возврата заемных средств и тем самым улучшает качество кредитного портфеля банка.

В СППР используется метод генетического программирования для построения математической модели в аналитическом виде. Математическая модель позволяет вычислить прогноз в виде вероятности возвращения кредита. Исходными данными для системы являются: положительная/отрицательная кредитная история в Сбербанке РФ, задолженности в Сбербанке РФ по кредитам (поручительства), задолженность перед другими кредитными организациями, положительная/отрицательная кредитная история в других кредитных организациях, задолженность по налогам и прочим сборам и штрафам (данные поступают от службы судебных приставов и ГИБДД), наличие/отсутствие задолженности организации по заработной плате, налогам и прочим сборам, средняя зара-



ботная плата в данной организации, прочая негативная информация о заемщике (приводы в милицию, наличие судебных решений и т.п.), наличие и количество иждивенцев, наличие имущества.

Конечно, окончательное решение о выдаче кредита остается за кредитным комитетом банка, но результат, который выдает скоринговая система, занимает не последнее место в принятии решения.

В настоящий момент система проходит тестирование и находится на стадии доработки специалистами управления автоматизации Восточно-Сибирского банка Сбербанка РФ.

После завершения тестирования системы проблемную задолженность физических лиц будет гораздо проще спрогнозировать на начальном этапе рассмотрения кредитной заявки и выдачи кредита, чем когда она появляется практически внезапно и приходится прилагать большие усилия, чтобы вернуть заемные средства. В перспективе возможно более широкое использование данной системы при рассмотрении заявок по другим видам кредитования.

1 Ильин Е.С., Семенкин Е.С. Интеллектуальная система анализа данных на основе нейронных сетей. М.: ВНТИЦ, 2004. Гос. рег. №50200400996.

2 Семенкин Е.С., Тынченко В.В., Тынченко В.С. Генетические алгоритмы выбора структуры и настройки искусственных нейронных сетей. М.: ВНТИЦ, 2006. Гос. рег. №50200601955.

3 Липинский Л.В., Семенкин Е.С. Система эволюционного формирования нейросетевых моделей сложных систем. М.: ВНТИЦ, 2006. Гос. рег. №50200602025.

4 Галыгин А.Н., Семенкин Е.С. Автоматическое формирование и оптимизация базы правил нечеткой логики. М.: ВНТИЦ, 2004. Гос. рег. №50200401213.

5 Липинский Л.В., Малько В.А., Семенкин Е.С. Интеллектуальные технологии автоматизации проектирования системы управления на нечеткой логике. М.: ВНТИЦ, 2006. Гос. рег. №50200600369.

6 Липинский Л.В., Семенкин Е.С. Система эволюционного формирования интеллектуальных систем на нечеткой логике. М.: ВНТИЦ, 2006. Гос. рег. №50200602024.

7 Жуков В.Г. Программная реализация метода генетического программирования с адаптивной настройкой коэффициентов для решения задачи символьной регрессии. М.: ВНТИЦ, 2006. Гос. рег. №50200601392.

8 Жуков В.Г., Семенкин Е.С. Программная система моделирования сложных систем с помощью коэволюционного алгоритма генетического программирования. М.: ВНТИЦ, 2006. Гос. рег. №50200601391.

9 Гречкосеева О.А., Процыков Г.В., Семенкин Е.С. Оптимизация размещения наземных станций слежения за орбитальными группировками космических аппаратов. М.: ВНТИЦ, 2006. Гос. рег. №50200601025.

10 Корнейчук Е.О., Семенкин Е.С., Ефимов С.Н. Коэволюционный алгоритм оптимизации конструкции зубчатых передач. М.: ВНТИЦ, 2006. Гос. рег. №50200600586.

11 Тынченко В.С., Семенкин Е.С., Ефимов С.Н. Автоматизированная система поддержки принятия решений «Формирование эффективной конфигурации GRID-системы» [Текст]. М.: ВНТИЦ, 2008. Гос. рег. №50200802059.

12 Прогнозирование успешности учебной и служебной деятельности образовательных учреждений МВД России / В.А.Терсков [и др.]. Красноярск: СибЮИ МВД России, 2004.

13 Ворожейкин А.Ю., Медведев А.В., Семенкин Е.С. Автоматизированное рабочее место инвестиционного аналитика. М.: ВНТИЦ, 2006. Гос. рег. №50200600629.

14 Ефимов С.Н., Мельников А.А., Стельмаш Д.С. Решение задачи символьной регрессии при разработке автоматизированной системы кредитного скоринга // Математические методы и информационно-технические средства : материалы VI Всероссийской научно-практической конференции. Краснодар: Краснодарский университет МВД России, 2010. С.46–53.