



СЕКЦИЯ 2

**КОМПЬЮТЕРНЫЙ
МОНИТОРИНГ И МЕТОДЫ
ИНТЕЛЛЕКТУАЛЬНОГО
АНАЛИЗА ДАННЫХ**

УДК 004.838.3

**МЕТОД ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ ПРЕЦЕДЕНТОВ
ДЛЯ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ****Поминчук Е.В., Иващенко А.Б.**

Донецкий национальный технический университет г. Донецк

Кафедра компьютерных систем мониторинга

E-mail: jeka275@ukr.net

Аннотация

Поминчук Е.В., Иващенко А.Б. Метод интеллектуального анализа данных прецедентов для прогнозирования временных рядов. В данной статье рассмотрен метод интеллектуального анализа данных прецедентов, обсуждаются идеи алгоритма, представлена схема разрабатываемого программного обеспечения.

Общая постановка проблемы

При современном уровне развития информационных технологий и, более конкретно, систем поддержки принятия решений различают два направления в развитии логического вывода знаний:

- развитие систем логического вывода, основанного на правилах;
- развитие систем логического вывода, основанного на прецедентах.

Практически все ранние экспертные системы моделировали ход принятия решения экспертом как чисто дедуктивный процесс с использованием логического вывода, основанного на правилах. Это означало, что в систему закладывалась совокупность правил вида "если...то...", согласно которым на основании входных данных генерировалось то или иное заключение по интересующей проблеме. Выбранная модель являлась основой для создания экспертных систем первых поколений, которые были достаточно удобны как для разработчиков, так и для пользователей-экспертов. Однако с течением времени было осознано, что дедуктивная модель моделирует один из наиболее редких подходов, которому следует эксперт при решении проблемы.

На самом деле, вместо того, чтобы решать каждую задачу, исходя из первичных принципов, эксперт часто анализирует ситуацию в целом и вспоминает, какие решения принимались ранее в подобных ситуациях. Затем он использует эти решения или же адаптирует их к обстоятельствам, изменившимся для конкретной проблемы.

Моделирование такого подхода к решению проблем, основанного на опыте прошлых ситуаций, привело к появлению технологии логического вывода, основанного на прецедентах (по-английски – Case-Based Reasoning, или CBR), и в дальнейшем – к созданию программных продуктов, реализующих эту технологию.

В ряде ситуаций метод вывода по прецедентам имеет серьезные преимущества по сравнению с выводом, основанным на правилах, и особенно эффективен, когда:

- основным источником знаний о задаче является опыт, а не теория,
- решения не уникальны для конкретной ситуации и могут быть использованы в других случаях;
- целью является не гарантированное верное решение, а лучшее из возможных.

Таким образом, вывод, основанный на прецедентах, представляет собой метод построения экспертных систем, которые делают заключения относительно данной проблемы или ситуации по результатам поиска аналогий, хранящихся в базе прецедентов.

Системы вывода по прецедентам показывают очень хорошие результаты в самых разнообразных задачах, но обладают рядом существенных недостатков.

Во-первых, они вообще не создают каких-либо моделей или правил, обобщающих предыдущий опыт, – в выборе решения они основываются на всем массиве доступных исторических данных, поэтому невозможно сказать, на основе каких конкретно факторов системы вывода по прецедентам строят свои конкретные ответы.

Можно выделить две основные проблемы, с которыми сталкиваются подобные системы: поиск наиболее подходящих прецедентов и последующая адаптация решения.

В основе всех подходов к отбору прецедентов лежит тот или иной способ измерения степени близости прецедента и текущего случая. При таких измерениях вычисляется численное значение некоторой меры, определяющей состав множества прецедентов, которые нужно обработать для достижения удовлетворительной классификации или прогноза. Основным недостатком таких систем является произвол, который допускают системы при выборе меры близости. Кроме того, безосновательным выглядит распространение общей меры близости на выборку данных в целом.

Еще один недостаток метода связан с конструкцией прецедентов и назначения уместных весов их атрибутам, что уменьшает их применимость (универсальность).

В большинстве случаев методы поиска прецедентов сводятся к индукции деревьев решений или к алгоритму "ближайшего соседа", дополненному, может быть, использованием знаний о предметной области. Что касается адаптации и использования найденного решения, эта задача до сих пор остается недостаточно формализованной и сильно зависящей от предметной области.

Обе проблемы – поиск прецедентов и адаптация выбранного решения – решаются (полностью или частично) с привлечением фонового знания, иными словами, знания о предметной области (domain knowledge). Существуют разные способы получения информации о предметной области:

- привлечение экспертного знания. Оно может выражаться, например, в ограничениях, накладываемых на диапазоны изменений признаков объектов, или же в формулировании набора правил для разбиения базы прецедентов на классы.

- получение необходимых знаний из набора имеющихся данных методами добычи данных (по-английски – Data Mining). Сюда относятся все методы выявления отношений в данных, в частности, кластеризация, регрессия, поиск ассоциаций. Использование методов добычи данных может выделить узкую группу показателей, от которых зависит интересующая исследователя характеристика, и представить обнаруженную закономерность в аналитической форме.

- формирование знаний на основе обучающей выборки, представленной экспертом (обучение с учителем). Этот способ включает в себя оба первых.

Изначально в системах вывода по прецедентам в качестве источников фонового знания выступали эксперты – высококвалифицированные специалисты предметных областей, а также текстовые материалы – от учебников до протоколов, и, разумеется, базы данных. Роль эксперта заключалась в вербализации, то есть переводе таких источников в эксплицитную форму. Учитывая, что важнейшей задачей в процессе формализации извлечения знаний является минимизация роли эксперта, его роль должны взять на себя средства добычи данных.

Среди извлекаемых закономерностей на практике чаще всего встречаются отношения эквивалентности и порядка. Первые присущи, в частности, задачам классификации, диагностики и распознавания образов. С другой стороны, отношения порядка свойственны задачам шкалирования, прогнозирования и т.п.

Актуальность

Необходимость предвидения вероятного развития событий на будущее в Украине, никогда ранее не была такой важной как сейчас. Решения, принимающиеся сегодня, опираются на признаки развития явлений. В свою очередь, они более или менее влияют на

это в будущем. Именно поэтому, исследования моделей прогнозирования временных рядов в условиях недостаточной информации поможет избежать принципиальных ошибок при принятии каких либо решений. Изучение этой проблемы является актуальным как для теории, так и для практики.

Идея алгоритма

Условно, алгоритм включает следующие этапы:

1. Ввод временного ряда для некоторой переменной;
2. Выбор «текущей даты» и числа «к» ближайших соседей;
3. Определение расстояний от значения текущей даты до значений предыдущих дат;
4. Сортировка расстояний по возрастанию;
5. Отбор «к» первых расстояний после сортировки;
6. Формирование массива «следующих значений»;
7. Адаптация и анализ значений, например, поиск арифметическое среднего, минимального или максимального значения;
8. Вывод полученного значения как результат прогноза;

На рисунке 1 представлена диаграмма модулей разрабатываемой системы.



Рисунок 1 – Диаграмма модулей разрабатываемой системы

Цель алгоритма

Спрогнозировать значение временного ряда для будущего временного периода на основе закономерностей и связей выявленных в БД содержащей значения (измерения) данного ряда по прошедшему периоду.

Вывод

Преимущества разрабатываемого алгоритма является возможность его использования для прогнозирования динамических показателей и факторов из любой области знаний и сферы деятельности человека. Например, для прогнозирования метеорологических параметров и состояния погоды, динамика курсов акций и валют, Прогнозирование потребительского спроса, прогнозирование объемов кредитования на следующий отчетный период, прогнозирование урожайности и многое другое.

Литература

1. История развития методов интеллектуального анализа данных – Data Mining. Интернет-ресурс. - Режим доступа: <http://azfor.ucoz.ru/publ/3-1-0-3>.
2. Обучение на примерах. Интернет-ресурс. - Режим доступа: http://ru.wikipedia.org/wiki/Обучение_по_прецедентам.
3. Машинное обучение. Интернет-ресурс. - Режим доступа: http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение.