КОМПЬЮТЕРНАЯ МАТЕМАТИКА

УДК 519.68

Н.А. ГУПАЛ, Ю.П. МАТУСОВ

ПРИМЕНЕНИЕ БАЙЕСОВСКОЙ ПРОЦЕДУРЫ РАСПОЗНАВАНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ДИНАМИКИ КУРСОВ АКЦИЙ

Экспертные системы, методы индуктивного вывода

На основе информации из Интернета относительно ежедневного изменения курсов акций ценных бумаг 570 предприятий решалась задача прогнозирования динамики курса акции на следующий день в зависимости от динамики нескольких предыдущих дней. Использовались байесовские процедуры распознавания на цепях Маркова разных порядков.

Введение. На практике часто встречаются задачи, которые связаны с наблюдениями случайных величин, например, прогнозированием курсов акций, валют, инвестиционной деятельности банков. Для подобных задач не удается построить детерминированные модели, поэтому используются статистические методы, построенные на основе информации относительно наблюдений предыдущих значений данных.

Цепь Маркова – один из простых случаев последовательности случайных событий, но несмотря на свою простоту она может быть использована для описания довольно сложных явлений. В задачах распознавания вторичной структуры белков используются байесовские процедуры на цепях Маркова [1]. Развиваемый подход мы применили для прогнозирования курсов акций.

В задаче прогнозирования курса акций можно считать, что завтрашний курс акций зависит только от значений курсов акций на протяжении последних 5–10 дней. Если это верно, то наблюдение значений курсов акций в течение нескольких месяцев позволит достаточно точно спрогнозировать изменение колебаний курсов в будущем.

Из Интернета получена информация относительно курсов акций ценных бумаг 570 предприятий. Акции ценных бумаг имеют собственный идентификатор. Для каждого момента времени (суток) указано четыре цены акции в действительных числах. Например:

AAPL 7,205; 7,455; 7,175; 7,331; 18913200.

Последнее число задает количество проведенных операций в течение дня. Для большинства акций общее количество записей составляет 1057 дней, т. е. примерно три года .Полученные данные составляют для нас обучающую выборку. На основе этой информации построим ряд процедур прогнозирования изменения цены конкретной акции в следующий момент времени. Без ограничения общности прогноз строится для одной цены акции, например, для первой.

Для решения этой задачи удобно использовать метод k -ближайших соседей [2]. Пусть известен курс $x(t)$ акции AAPL в момент времени t. Для значения курса $x(t)$ определим $k=10$ ближайших значений курсов в моменты времени $t_1, t_2, ..., t_{10}$, не превышающих t; пусть они будут $x(t_1), x(t_2), ..., x(t_{10})$. Значение курса акции AAPL в момент времени $t+1$ определим как среднее арифметическое значений $x(t_1), x(t_2), ..., x(t_{10})$:

$$x(t+1) = \frac{x(t_1) + ... + x(t_{10})}{10}.$$

Поскольку известны значения курсов акции в течение 1057 моментов времени, то можно проверить построенный прогноз с реальными данными. Оказалось, что в 607 случаях мы правильно спрогнозировали повышение или понижение курса $x(t)$ в момент времени $t+1$ по сравнению с известным курсом $x(t)$ в момент t. Этот результат составляет примерно 57 % угаданных изменений курсов для акции AAPL. Отметим, что число ближайших соседей $k=10$ определено опытным путем, т. е. для других значений k были получены результаты хуже.

Сравним этот результат с результатами байесовских процедур, построенных с помощью цепей Маркова разных порядков. Было доказано, что байесовские процедуры распознавания оптимальны на таких структурах, как цепи Маркова и независимые признаки. Они имеют полиномиальные оценки погрешности в зависимости от размеров обучающих выборок и количества признаков [3–5]. Показано, что байесовская процедура распознавания для независимых признаков имеет полиномиальную верхнюю и нижнюю оценки погрешности от входа задачи:

$$\upsilon(Q_Б, C) \sim a\sqrt{\frac{n}{\min(m_0, m_1)} + \frac{1}{m_2}}.$$

Решение получается за один шаг вычислений, и оценка погрешности стремится к нулю с увеличением размеров обучающей выборки. Байесовские процедуры распознавания, построенные на нестационарных цепях Маркова, имеют для больших выборок такие же оценки погрешности, что и процедуры для независимых признаков [5]. При выборе модели цепи Маркова и построении процедур распознавания необходимо решить вопрос относительно стационарности и нестационарности переходных вероятностей. В работе [6] обосновано применение критерия χ^2 относительно стационарности переходных вероятностей для цепей Маркова.

Сначала сделаем расчеты для самого простого случая в предположении, что курсы акций в разные моменты времени являются независимыми случайными величинами (модель цепи Маркова порядка 0). Известно, что цепи Маркова выгодно использовать в том случае, когда случайная величина $x(t)$ принимает конечное число значений. В данном случае мы следим за повышением или понижением значения курса акции, т.е. величина $x(t)$ принимает два значения: «+», если значение курса акции $x(t)$ в момент времени t превышает значение курса $x(t-1)$ в момент времени $t-1$; и «–», если значение курса $x(t)$ меньше по сравнению с $x(t-1)$.

В моделях цепей Маркова предполагается, что повышение или понижение завтрашнего курса акции зависит от значений курсов акции в течение нескольких последних k дней. Эксперименты показали, что наилучшие результаты достигаются при $k=5$. При построении байесовских процедур прогнозирования необходимо подсчитать оценки вероятностей коротких цепочек $x(t-4), x(t-3), x(t-2), x(t-1), x(t)$, для которых отмеченные величины принимают два значения: «+» или «–».

Естественно, мы считаем, что оценка вероятности цепочки $x(t-4)$, $x(t-3)$, $x(t-2), x(t-1), x(t)$ зависит от того, какое состояние имеет случайная величина $x(t)$ в интересующий нас момент времени $t+1$. Таким образом, оценки вероятностей указанных цепочек подсчитываем для двух классов объектов: обучающая выборка класса «+» состоит из всех цепочек $x(t-4), x(t-3)$, $x(t-2)$, $x(t-1), x(t)$, для которых состояние $x(t+1)$ равно «+». Наоборот, обучающая выборка класса «–» состоит из всех цепочек $x(t-4), x(t-3), x(t-2)$, $x(t-1), x(t)$, для которых состояние $x(t+1)$ равно «–». Для независимых случайных величин $x(t)$ подсчеты оценок вероятностей цепочек $x(t-4)$, $x(t-3), x(t-2), x(t-1), x(t)$ проводятся очевидным образом. Вычисления показали, что число спрогнозированных изменений значений курсов акции AAPL составляет 598, т. е. -51 %.

Для моделей цепей Маркова первого порядка вероятность цепочки $x(t-4)$, $x(t-3)$, $x(t-2)$, $x(t-1)$, $x(t)$ вычисляется по формуле

$$P(x(t-4), x(t-3), ..., x(t) | f) =$$

$$= P(x(t-4) | f) \times P(x(t-3) | x(t-4), f) \times ... \times P(x(t) | x(t-1), f), \qquad (1)$$

где $P(x(k) | x(k-1), f)$, $k = t-3, ..., t$ – нестационарные переходные вероятности; $P(x(t-4) | f)$ – начальное распределение вероятностей; $f \in \{+, -\}$. Случайные величины $x(t)$ принимают два значения : «+» или «–».

В численных расчетах оценки переходных вероятностей определяются на основе частот

$$\hat{P}(x(k)=j|x(k-1)=i,f)=\frac{m(x(k-1)=i,x(k)=j,f)}{m(x(k-1)=i,f)}, \tag{2}$$

где $m(x(k-1)=i,x(k)=j,f)$ – число цепочек $x(t-4), x(t-3), x(t-2), x(t-1)$, $x(t)$, принадлежащих заданному классу $f\in\{+,-\}$ в обучающей выборке, для которых величина $x(k-1)$ принимает значение i, а величина $x(k)$ – значение j, $i,j\in\{+,-\}$; $m(x(k-1)=i,f)$ – число указанных цепочек, для которых величина $x(k-1)$ принимает значение i.

Байесовская процедура прогнозирования на цепях Маркова первого порядка задает состояние величины $x(t+1)=+$, если оценка вероятности цепочки $x(t-4), x(t-3), x(t-2), x(t-1), x(t)$, которая подсчитана для обучающей выборки класса «+», превышает оценку вероятности этой цепочки, подсчитанной для обучающей выборки класса «–». Вычисления показали, что число спрогнозированных изменений курсов акции AAPL составляет 637, или – 61 %.

Аналогично строится байесовская процедура прогнозирования на цепях Маркова второго порядка. В отличие от формулы (2) оценки переходных вероятностей в числителе вычисляются на основе количества троек состояний величин $x(t)$ для трех последовательных моментов времени, а в знаменателе (2) соответственно присутствует количество двоек состояний величин $x(t)$. Расчеты показали, что число спрогнозированных изменений курсов акции AAPL для этой процедуры составляет 616, или – 58 %.

Объем обучающей выборки позволяет построить байесовскую процедуру прогнозирования на цепях Маркова третьего порядка, которая определяет зависимость состояний случайных величин от состояний трех предыдущих моментов времени. Численные расчеты показали, что число спрогнозированных изменений курсов акции AAPL для этой процедуры составляет 621, или – 59 %.

Приведем небольшой фрагмент таблицы численных расчетов, в колонках которой содержатся показатели правильно спрогнозированных изменений курса для процедуры на цепи Маркова:

1 – идентификатор акции;
2 – число порядка 0, $k=5$;
3 – число порядка 1, $k=5$;
4 – число порядка 2, $k=5$;
5 – число порядка 3, $k=5$;
6 – число для метода k – ближайших соседей, $k=10$;
7 – общее число записей в обучающей выборке.

ТАБЛИЦА. Результаты прогнозирования курсов акций

1	2	3	4	5	6	7
AAPL.csv	598	637	616	621	607	1057
AATI.csv	254	268	245	247	250	405
ABAX.csv	569	585	572	558	558	1057
ABBI.csv	647	653	640	617	593	1057
ABFS.csv	591	612	577	578	578	1057
ACAS.csv	611	634	639	623	605	1057
ACGY.csv	565	587	576	564	564	1057
ACLI.csv	208	224	210	199	193	358
ACXM.csv	578	621	610	601	573	1057
ADBE.csv	600	612	603	601	578	1057
ADCT.csv	580	609	599	596	572	1057
ADSK.csv	632	636	591	587	603	1057
ADTN.csv	601	635	630	602	589	1057

Результаты экспериментов для других акций показали, что байесовская процедура прогнозирования на цепях Маркова первого порядка дает наилучший прогноз изменения курсов акций по сравнению с описанными методами прогнозирования. Полученные результаты дают представление о байесовской процедуре индуктивного вывода как об универсальной процедуре, которая перерабатывает информацию оптимальным образом.

М.А. Гупал, *Ю.П. Матусов*

ЗАСТОСУВАННЯ БАЙЄСІВСЬКОЇ ПРОЦЕДУРИ РОЗПІЗНАВАННЯ
ДЛЯ ПРОГНОЗУВАННЯ ДИНАМІКИ КУРСІВ АКЦІЙ

На основі інформації з Інтернету стосовно щоденної зміни курсів акцій цінних паперів 570 підприємств розв’язувалась задача прогнозування динаміки курсів акцій на наступний день в залежності від динаміки декількох попередніх днів. Було застосовано байєсівські процедури розпізнавання на ланцюгах Маркова різних порядків.

N.A. Gupal, *Yu.P. Matusov*

APPLICATION OF BAYESIAN RECOGNITION PROCEDURES FOR PREDICTION
OF DYNAMICS OF SHARE PRICES

On the basis of information regarding daily change of share prices of securities of 570 enterprises obtained from the Internet, the problem of prediction of dynamics of share prices for the next day depending on the one of few previous days was solved. Bayesian recognition procedures on the Markov chains of different orders were used.

1. *Сергиенко И.В., Белецкий Б.А., Васильев С.В., Гупал А.М.* Предсказание вторичной структуры белков на основе байесовской процедуры распознавания на цепях Маркова // Кибернетика и системный анализ. – 2007. – № 2. – С. 59–64.

2. *Журавлев Ю.И., Рязанов В.В., Сенько О.В.* Распознавание. Математические методы. Программная система. Практические применения. – М.: Фазис, 2006. – 176 с.

3. *Гупал А.М., Сергиенко И.В.* Оптимальные процедуры распознавания // Там же. – 2003. – № 1. – С. 33–39.

4. *Вагис А.А.* Принципы построения процедур индуктивного вывода // Компьютерная математика. – 2006. – № 1. – С. 132–139.

5. *Белецкий Б.А., Вагис А.А., Васильев С.В., Гупал Н.А.* Сложность байесовской процедуры индуктивного вывода. Дискретный случай // Проблемы управления и информатики. – 2006. – № 6. – С. 55–70.

6. *Белецкий Б.А., Вагис А.А., Васильев С.В.* Распознавание гипотез в моделях цепей Маркова // Компьютерная математика. – 2006. – № 1. – С. 104–112.

Получено 09.12.2008

Об авторах :

Гупал Никита Анатольевич,

студент факультета менеджмента и маркетинга
Национального технического университета Украины «КПИ»,

Матусов Юрий Петрович,

старший преподаватель факультета менеджмента и маркетинга
Национального технического университета Украины «КПИ».