

УДК 004.89

Н.М. Кораблев¹, Г.С. Иващенко²¹ ХНУРЭ, г. Харьков, Украина, korablev.nm@gmail.com,² gs2005@rambler.ru

ПРИМЕНЕНИЕ МОДЕЛИ КЛОНАЛЬНОГО ОТБОРА, ИСПОЛЬЗУЮЩЕЙ ВЫВОД ПО ПРЕЦЕДЕНТАМ, ДЛЯ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ

В статье предложен комбинированный метод прогнозирования временных рядов при помощи искусственных иммунных систем. Рассмотрена модель прогнозирования на основе модели клонального отбора и метода вывода по прецедентам (СВР). Оценка эффективности модели выполнена путем сравнительного анализа, представлены результаты экспериментальных исследований, демонстрирующие особенности предлагаемого подхода.

ПРОГНОЗИРОВАНИЕ, ВРЕМЕННОЙ РЯД, ВЫБОРКА, ВЫВОД ПО ПРЕЦЕДЕНТАМ, ИСКУССТВЕННАЯ ИММУННАЯ СИСТЕМА, МОДЕЛЬ КЛОНАЛЬНОГО ОТБОРА, АНТИТЕЛО, АНТИГЕН, АФФИННОСТЬ, КЛОНИРОВАНИЕ, МУТАЦИЯ, СУПРЕССИЯ

Введение

Задача прогнозирования временных рядов играет важную роль в процессе принятия решений в технических и экономических системах. Прогнозирование будущих значений временного ряда позволяет решить актуальную задачу определения будущего состояния на основе анализа уже имеющихся ретроспективных данных. Точность прогноза зависит от многих параметров, таких как объем собранной статистической информации, ее качество и достоверность, требуемая величина горизонта прогнозирования, учет возможности изменения среды, в которой протекает исследуемый процесс, число одновременно учтенных членов ряда, доступность необходимых вычислительных ресурсов.

Существует широкий спектр методов прогнозирования [1], и выбор среди них наиболее подходящего для заданных условий является одним из определяющих факторов получения достоверного прогноза. В настоящее время актуальным направлением в области разработки подходов прогнозирования является создание комбинированных моделей и методов, что позволяет компенсировать недостатки одних моделей при помощи особенностей других. В [2] предлагается применение группы предикторов, моделирующих разные участки временного ряда, что позволяет достичь более высокой точности прогноза по сравнению с применением по отдельности каждого предиктора из заданного набора.

Перспективным подходом является применение метода вывода на основе прецедентов (case based reasoning – СВР), в котором при рассмотрении новой задачи прогнозирования отыскивается подобный прецедент в предыстории в качестве аналога [3–5]. Метод, предложенный в [3], основывается на представлении всего временного ряда как последовательности выборок.

Прогнозирование при помощи СВР основывается на предположении, что путем определения выборки, максимально соответствующей последним известным значениям временного ряда, возможна

оценка будущих значений ряда. Проблема выбора подходящего прецедента среди имеющихся в хранилище прецедентов и его адаптация к текущим условиям является одной из самых актуальных в СВР-системах [4].

В настоящее время активно развиваются подходы к решению задачи краткосрочного прогнозирования на основе методов искусственного интеллекта, таких как искусственные нейронные сети и искусственные иммунные системы (ИИС) [6–9], которые могут интегрироваться с другими подходами и отличаются быстроедействием и адаптационными возможностями.

Существуют различные модели ИИС, основанные на принципах работы иммунной системы: модель клонального отбора, модель иммунной сети, модель отрицательного отбора и другие. В данной работе для решения поставленной задачи предлагается использование модели клонального отбора [6].

1. Постановка задачи

Рассматриваются дискретные временные ряды, значения которых получены в моменты времени $t_1, t_2, t_3, \dots, t_N$, которые могут быть неравноотстоящими [5]. Временной ряд $Z = z(t_1), z(t_2), z(t_3), \dots, z(t_N)$ длины N обозначим $Z_1^N = z_1, z_2, z_3, \dots, z_N$. Набор последовательных значений $Z_t^L = z_t, z_{t+1}, z_{t+2}, \dots, z_{t+L-1}$, лежащих внутри временного ряда Z_1^N , назовем выборкой из этого ряда длины L с моментом начала отсчета t , $L \in [1, N-1]$, $t \in [1, N-L]$. Задача прогнозирования временного ряда заключается в том, чтобы по его известному участку оценить будущие значения.

Согласно гипотезе, сформулированной в [3], если мера подобия между выборками Z_t^L и Z_{t-k}^L имеет значение, близкое к единице, то мера подобия между выборками длины P , следующими за ними, Z_{t+L}^P и Z_{t-k+L}^P , также близка к единице. Тогда путем определения выборки, максимально соответствующей последним известным значениям временного ряда, возможна оценка его будущих значений.

Цель проводимого исследования – разработка иммунного алгоритма краткосрочного прогнозирования временных рядов на основе модели клонального отбора, использующей метод вывода по прецедентам, и проведение экспериментальных исследований.

2. Модель клонального отбора, использующая метод вывода по прецедентам

Вывод, основанный на прецедентах, это метод анализа данных, который делает заключения относительно текущей ситуации по результатам поиска аналогий – прецедентов, хранящихся в базе. Данный процесс включает в себя такие этапы, как отбор множества прецедентов из имеющихся в базе на основе заданного отношения подобия, адаптация выбранных прецедентов в связи с текущими условиями, сравнение полученного и реального значений прогнозируемой величины, и сохранение в базе прецедентов для дальнейшего использования принятого решения и текущей ситуации в качестве нового прецедента, или соответствующее изменение ранее выбранного прецедента [4].

Возможность применения CBR для решения задачи прогнозирования определяется на основании предположений о природе временного ряда, предложенных в [2]: будущая динамика временного ряда зависит от его предыстории; существуют участки временного ряда, на которых действуют одинаковые или близкие зависимости (история повторяется).

Алгоритм клонального отбора оперирует строками данных фиксированной длины и часто используется для решения задач классификации, распознавания и оптимизации [6]. Каждая строка данных алгоритма, называемая антителом (в терминах CBR – прецедент), представляет собой набор параметров, описывающих поставленную задачу (набор известных значений ряда) и принятое решение (прогнозируемые значения ряда).

Для построения модели прогнозирования на основе ИИС необходимо сопоставить биологические объекты и процессы с их аналогами из предметной области.

Антитело включает в себя выборку известных значений ряда $Ab_i = ab_1, ab_2, ab_3, \dots, ab_L, \dots, ab_{L+f}$, где i – индекс в популяции, L – длина выборки известных значений ряда, f – величина горизонта прогнозирования, длина выборки прогнозируемых значений ряда. Антиген $Ag_j = ag_1, ag_2, ag_3, \dots, ag_L$, где j – индекс в популяции, представляет собой последовательность значений временного ряда, непосредственно предшествующих прогнозируемым.

Аффинность – скалярная величина, оценивающая меру близости между антителом и антигеном, основной критерий отбора антител – определяется как

$$Aff = \frac{\sum_{k=1}^L (1 + |ab_k - ag_k|)^{-1}}{L} \in (0, 1].$$

Антитело также включает в себя выборку длины f прогнозируемых им значений. Эта часть не участвует в определении аффинности. Число значений временного ряда, которые в нее входят, соответствует величине горизонта прогнозирования.

Таким образом, в терминах подхода CBR антитело исполняет роль прецедента – содержит описание текущей ситуации (последовательность известных значений ряда) и принятое в данный момент решение (соответствующий прогноз). Задача прогнозирования в таком случае сводится к нахождению антитела с наибольшей аффинностью, т.е. к поиску прецедента, наиболее соответствующего последнему участку значений временного ряда. Модель может быть расширена путем включения в антитело и антиген внешних факторов, представленных в виде других временных рядов.

Формально модель ИИС можно представить следующим образом:

$$AIS_CS = \langle Ab, n', n_c, Ab^C, Ab^M, Ag, D, \sigma_d, S, \sigma_s, \sigma_{age} \rangle,$$

где Ab – популяция антител размера n ; n' – число клеток, отобранных для клонирования и мутации; n_c – количество клонов, создаваемых одним антителом; Ab^C – популяция клонов; Ab^M – популяция клеток памяти, Ag – популяция антигенов размера m ; D – матрица Ab-Ag аффинностей; σ_d – пороговый коэффициент стимуляции клетки; S – матрица Ab-Ab аффинностей; σ_s – пороговый коэффициент сжатия; σ_{age} – пороговый возраст антител.

Модель ИИС включает в себя популяции антител и антигенов (отдельно выделяется популяция клеток памяти, подмножество популяции антител), параметры, определяющие супрессию и отбор – матрицы аффинностей между антителами и антигенами и антител между собой, а также пороговые значения, и параметры, регулирующие клонирование.

Каждое антитело обладает «возрастом» – числом запусков алгоритма (попыток получить прогноз) с момента появления антитела в популяции. Если «возраст» выше порогового значения, и антитело не входит в число клеток памяти, оно устраняется из популяции в ходе работы алгоритма. Клетка памяти – антитело, которое хотя бы один раз было отобрано для клонирования, т.е. обладавшее аффинностью выше пороговой.

Алгоритм получения прогноза включает в себя следующие этапы:

1. Создание начальной популяции антител. Для создания антител используется часть известных значений временного ряда. Неиспользуемые значения исполняют роль обучающей и контрольной выборок.

2. Формирование антигена, соответствующего текущей задаче, т.е. создание выборки из последовательности известных значений ряда, предшествующих прогнозируемым.

3. Определение Ab-Ag аффинностей и отбор антител, обладающих аффинностью выше порогового значения.

4. В качестве результата принимается прогноз антитела, имеющего наибольшую аффинность.

5. При наличии реального значения прогнозируемой величины определяется ошибка прогноза.

6. Клонирование антител с наибольшей аффинностью (отобранных на этапе 3), в процессе которого выполняется оператор мутации.

7. Определение Ab-Ab аффинностей, применение оператора супрессии с целью устранения избыточности. Оставшиеся антитела из числа ранее отобранных становятся клетками памяти.

8. Применение оператора старения и коррекция популяции.

После получения реального значения прогнозируемой величины (шаги 5-6), отобранные антитела клонируются, в процессе чего выполняется оператор мутации.

Мутация – случайное изменение элементов выборки известных значений и направленное изменение выборки прогнозируемых значений, входящих в состав антитела. Степень мутации определяется аффинностью данного антитела – наиболее подходящий прецедент подвергается более существенной коррекции.

После применения оператора старения возраст антител увеличивается на единицу. Антитела, обладающие возрастом выше порогового значения и не являющиеся клетками памяти, заменяются антителами с наибольшей аффинностью из популяции клонов.

Шаги 2-8 представляют собой обучение ИИС и повторяются для каждого антигена из обучающей выборки заданное число раз. В качестве критерия останова иммунного алгоритма используется останов по достижении определенного числа поколений или останов по достижении некоторого заданного значения ошибки.

Для оценки прогноза используется значение средней абсолютной ошибки (mean absolute error, MAE). На рис.1 показана зависимость ошибки прогноза антител, имеющих различную аффинность, для необученной (график 1) и обученной (график 2) ИИС. Таким образом, в процессе обучения происходит настройка системы на новые значения временного ряда.

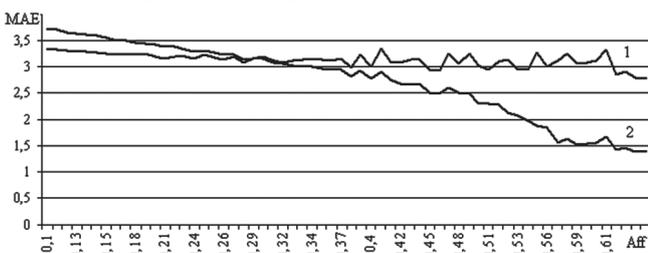


Рис. 1. Зависимость величины ошибки от аффинности

Выбор значений параметров иммунной системы, таких как максимальный размер популяции

антител и параметры отбора, клонирования и супрессии, для заданного ряда осуществляется методами вычислительного эксперимента. Определение длины выборки известных значений ряда, входящих в состав антител, это настройка на текущий ряд, которая выполняется на основе рекомендаций, изложенных в [5]. График на рис.2 демонстрирует влияние величины L на точность получаемого прогноза для необученной ИИС как построенной на основе модели клонального отбора, так и на основе модели иммунной сети.

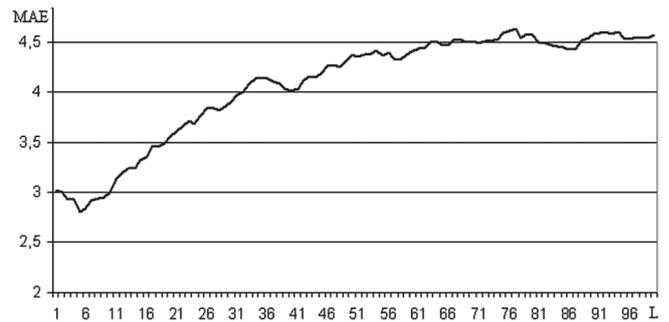


Рис. 2. Зависимость средней абсолютной ошибки прогнозирования от величины антитела

3. Результаты сравнительного анализа

В ходе экспериментальных исследований было проведено краткосрочное прогнозирование рядов, используемых в соревновании M3-Competition [10], а также ряда среднесуточных показаний температуры. Средние абсолютные ошибки прогнозирования разными методами приведены в табл. 1. Данные для методов, отличных от рассматриваемого в работе и AINet, получены из результатов соревнования в [10].

Сравнивая результаты предложенного подхода и модели на основе искусственной иммунной сети (AINet), можно отметить, что хотя для ряда температуры результаты близки, использование клонального отбора позволяет лучше прогнозировать ряды с меньшей предысторией, т.е. данный подход менее требователен к количеству значений временного ряда.

Результат прогнозирования ряда среднесуточных показаний температуры (Meteo) подтверждает преимущество использования подходов на основе ИИС – его размеры позволяют создать популяцию, соответствующую обширной базе прецедентов, и провести обучение.

Однако при малом количестве значений временного ряда при получении прогноза вероятно отсутствие в базе антител с высокой аффинностью (нет подходящего прецедента), а величина обучающей выборки не позволит в полной мере настроить (обучить) систему. Вследствие этого на некоторых рядах, представленных в таблице, традиционные методы имеют преимущества перед предложенным подходом.

Таблица 1

Средняя ошибка прогнозирования при использовании различных методов

| Метод | N2832 (104) | N2841 (104) | N2856 (92) | N2859 (79) | N2864 (79) | N2868 (79) | N2869 (79) | N2875 (79) | Meteo (21337) |
|-----------------|----------------|----------------|---------------|---------------|---------------|---------------|---------------|---------------|------------------|
| Exp.Smoothing | 3148,09 | 68,70 | 253,08 | 132,71 | 1028,13 | 775 | 497,5 | 350,625 | 2,92 |
| Holt–Winters | 3676,51 | 52,72 | 296,47 | 149,2 | 1146,32 | 947,2 | 580,31 | 327,585 | 2,9 |
| Box–Jenkins | 2814,27 | 68,7 | 260,21 | 125,09 | 896,62 | 1039,33 | 210,65 | 1061,77 | 2,99 |
| RBF | 3759,34 | 37,47 | 381,63 | 145,57 | 1091,78 | 1227,36 | 628,15 | 348 | – |
| ForecastPro | 3098,99 | 68,58 | 260,26 | 132,71 | 1028,13 | 775 | 185,79 | 1161,59 | – |
| SmartFcs | 3183,35 | 67,71 | 255,99 | 132,71 | 1077,06 | 1831,23 | 525,2 | 697,11 | – |
| Automat ANN | 1937,68 | 72,95 | 300,77 | 111,76 | 976,70 | 985,22 | 349,1 | 1072,96 | – |
| AINet | 2848,68 | 151,46 | 216,50 | 130,51 | 1463,12 | 1372,74 | 225,76 | 855,89 | 2,45 |
| ClonalSelection | 2917,45 | 126,1 | 216,50 | 128,3 | 812,5 | 1044,67 | 218,7 | 502,23 | 2,44 |

Выводы

В работе впервые предложена модель краткосрочного прогнозирования временных рядов на основе модели клонального отбора, использующей метод вывода по прецедентам. Основные особенности предложенной модели заключаются в следующем:

- использование популяции антител в качестве базы прецедентов;
- применение прямо пропорциональной направленной мутации;
- формирование множества паттернов (популяции клеток памяти) в процессе анализа временного ряда в ходе обучения ИИС;
- целесообразность применения для решения задачи краткосрочного прогнозирования;
- влияние размера антител на точность получаемого прогноза;
- необходимость обучающей выборки значительного объема, но меньшего, чем при использовании модели иммунной сети;
- низкий результат при наличии тренда, желательное постоянство математического ожидания;
- модель может быть расширена с целью решения задачи прогнозирования временного ряда с учетом влияния внешних факторов, представленных в виде других временных рядов.

Представленные в работе результаты подтверждают эффективность использования предложенного подхода для краткосрочного прогнозирования временных рядов, однако использование модели для прогнозирования ряда в условиях недостатка исходной информации требует дополнительных исследований и доработки модели прогнозирования.

Список литературы 1. Лукашин, Ю. П. Адаптивные методы краткосрочного прогнозирования временных рядов [Текст] : Учебное пособие / Ю. П. Лукашин. – М. : Финансы и статистика, 2003. – 416 с. ISBN 5-279-02740-5. **2.** Батуро, А. П. Финансовые временные ряды: кусочное прогнозирование и проблема обнаружения предвестников существующего изменения закономерности [Текст] / А. П. Батуро, Н. М. Еременко // Банковские технологии. – 2001. – № 12. – С. 70–77. **3.** Singh, S. Pattern Modelling in Time-Series Forecasting [Текст] / S. Singh // Cybernetics and Systems – An International Journal. – 2000. – № 1(31). –

Р. 49–65. **4.** Черный, С. Г. Применение case based reasoning для поддержки принятия решений [Текст] / С. Г. Черный // Вестник ХНТУ. – 2010. – № 2(38). – С. 336–342. **5.** Чучуева, И. А. Модель экстраполяции временных рядов по выборке максимального подобия [Текст] / И. А. Чучуева // Информационные технологии. – 2010. – № 12. – С. 43–47. **6.** Дасгупта, Д. Искусственные иммунные системы и их применение [Текст] : пер. с англ. / под ред. А. А. Романюхи; М.: ФИЗМАТЛИТ, 2006. – 344 с. **7.** Timmis, J. I. An Overview of Artificial Immune Systems [Текст] / J. I. Timmis, T. Knight, L. N. De Castro, E. H. Art // Computation in Cells and Tissues. – Springer. – 2004. – Р. 51–86. **8.** Бидюк, П. И. Алгоритм клонального отбора для прогнозирования нестационарных динамических систем [Текст] / П. И. Бидюк, В. И. Литвиненко, И. В. Баклан, А. А. Фелелов // Искусственный интеллект. – 2004. – № 4. – С. 89–99. **9.** Самигулина, Г. А. Разработка интеллектуальных экспертных систем прогнозирования и управления на основе искусственных иммунных систем [Текст] / Г. А. Самигулина // Проблемы информатики. – Новосибирск, 2010. – № 1. – С. 15–22. **10.** Makridakis, S. The M-3 Competition: Results, Conclusions and Implications [Текст] / S. Makridakis, M. Hibon // International Journal of Forecasting. – 2000. – № 16. – Р. 451–476.

Поступила в редколлегию 19.02.2013

УДК 004.89

Застосування моделі клонального відбору з використанням висновку за прецедентами для прогнозування часових рядів / М. М. Корабльов, Г. С. Івашенко // Біоніка інтелекту: наук.-техн. журнал. – 2013. – № 1 (80). – С. 108–111.

У статті розглядається комбінований метод короткострокового прогнозування часових рядів за допомогою штучних імунних систем. Запропоновано підхід на основі моделі клонального відбору та методу виведення з прецедентів. Був проведений порівняльний аналіз ефективності застосування запропонованої моделі.

Табл. 1. Іл. 2. Бібліогр.: 10 найм.

UDK 004.89

Application of the clonal selection model using the case based reasoning for time series forecasting. / N. M. Korablyov, G. S. Ivaschenko // Bionics of Intelligense: Sci. Mag. – 2013. – № 1 (80). – P. 108–111.

This paper proposes the combined method of short-term time series forecasting using artificial immune systems. A model of the prediction based on the model of clonal selection and the case based reasoning method. Was performed a comparative analysis of the effectiveness of the proposed model, the approach based on the immune network model and the traditional methods of time series prediction.

Tab. 1. Fig. 2. Ref.: 10 items.