

УДК 004.89

МОДЕЛЬ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ ИСКУССТВЕННЫХ ИММУННЫХ СЕТЕЙ

Иващенко Г.С., Кораблев Н.М.

Харьковский национальный университет радиоэлектроники, Украина

В работе рассмотрены вопросы прогнозирования временных рядов с помощью искусственных иммунных систем, предложена модель на основе иммунной сети, использующая метод вывода по прецедентам (CBR). Оценка эффективности модели выполнена путем сравнительного анализа, представлены результаты экспериментальных исследований, демонстрирующие особенности предлагаемого подхода.

Введение

Прогнозирование будущих значений временного ряда остается важной научно-технической проблемой, так как позволяет решить всегда актуальную задачу определения будущего состояния различных систем (экономических, социальных, экологических и других) на основе анализа уже имеющихся ретроспективных данных. Результат прогнозирования зависит от многих параметров, таких как объем необходимой статистической информации (величины предыстории), ее качество и достоверность, учет возможности изменения среды, в которой протекает процесс, число одновременно учтенных членов ряда, доступность необходимых вычислительных средств. Одним из определяющих факторов является обоснованный выбор необходимого метода прогнозирования.

Несмотря на широкое распространение в экономической практике статистических методов прогнозирования [1, 2], в настоящее время активно развиваются подходы на основе методов искусственного интеллекта, таких как искусственные нейронные сети и искусственные иммунные системы (ИИС) [3-6], которые могут интегрироваться с другими подходами.

Перспективным является применение метода вывода на основе прецедентов, в котором при рассмотрении новой задачи отыскивается подобный прецедент в предыстории в качестве аналога [7-8]. Проблема выбора подходящего прецедента и его адаптация к текущим условиям является одной из самых важных в таких системах [9].

Цель проводимого исследования – разработка модели прогнозирования временных рядов на основе искусственной иммунной сети, использующей метод вывода по прецедентам, программная реализация разработанного подхода и проведение экспериментальных исследований.

Постановка задачи

В данной работе рассматриваются дискретные временные ряды, значения которых получены в моменты времени $t_1, t_2, t_3, \dots, t_N$. Временной ряд $Z = z(t_1), z(t_2), \dots, z(t_N)$ обозначим $Z_1^N = z_1, z_2, z_3, \dots, z_N$. Набор последовательных

значений $Z_t^M = z_t, z_{t+1}, z_{t+2}, \dots, z_{t+M-1}$, лежащих внутри временного ряда Z_1^N , назовем выборкой из этого ряда длины M , с моментом начала отсчета t , $M \in [1, N-1]$, $t \in [1, N-M]$. Задача прогнозирования временного ряда заключается в том, чтобы по его известному участку оценить будущие значения.

Целесообразность применения методов вывода по прецедентам определяется на основании следующих фундаментальных предположений о природе временного ряда [2]:

- будущая динамика временного ряда зависит от его предыстории;
- зависимость временного ряда может со временем меняться, но на некоторых участках она сохраняет определенное постоянство;
- существуют участки временного ряда, на которых действуют одинаковые или близкие зависимости (история повторяется);
- существуют участки временного ряда, для которых существует принципиальная возможность построения предикторов.

Согласно гипотезе, сформулированной в [9], если мера подобия между выборками Z_t^M и Z_{t-k}^M имеет значение, близкое к единице, то мера подобия между выборками длины P , следующими за ними, Z_{t+M}^P и Z_{t-k+M}^P , также близка к единице. Тогда путем определения выборки, максимально соответствующей последним известным значениям временного ряда, возможна оценка будущих значений ряда. В предложенной работе для решения поставленной задачи предлагается использование модели искусственной иммунной сети.

Модель иммунной сети, использующей CBR

Вывод, основанный на прецедентах, представляет собой такой метод анализа данных, который делает заключения относительно данной ситуации по результатам поиска аналогий, хранящихся в базе прецедентов. Данный процесс включает в себя следующие этапы [7]:

- выбор из хранилища прецедентов наиболее уместного прецедента (или множества прецедентов), обычно на основе заданного отношения подобия;
- использование выбранных прецедентов для решения поставленной задачи;
- адаптация выбранного решения в связи с текущими условиями, если это необходимо;
- проверка корректности полученного решения;
- сохранение в хранилище для дальнейшего использования принятого решения и текущей ситуации в качестве нового прецедента, или соответствующее изменение выбранного прецедента.

Для построения модели прогнозирования на основе ИИС необходимо сопоставить биологические объекты и процессы с их аналогами из предметной области. Антитела представляют собой известные значения временного ряда. Антигены – значения временного ряда, непосредственно предшествующие прогнозируемым. Аффинность антител – главный критерий отбора в алгоритме иммунной сети – скалярная величина оценки, показывающая близость между антителом и антигеном. В рассматриваемом

подходе аффинность определяется не для отдельной пары антиген-антитело, а между кортежами антител и антигенов – мультиантителом mAb и мультиантигеном mAg : $mAb = ab_1, ab_2, ab_3, \dots, ab_M$, $mAg = ag_1, ag_2, ag_3, \dots, ag_M$.

Аффинность определяется по формуле:

$$Aff = \frac{\sum (1 + d_i)^{-1}}{M}, i \in [1, M] \quad (1)$$

где d_i – евклидово расстояние между парами значений mAb и mAg .

Мультиантитело включает в себя прогнозируемое им значение – эта часть не участвует в определении аффинности. Число антител, которые в нее входят, соответствует величине горизонта прогнозирования. В терминах подхода CBR – мультиантитело исполняет роль прецедента. Задача прогнозирования в таком случае сводится к нахождению мультиантитела с наибольшей аффинностью.

Идентификация модели (определение величины мультиантитела) выполняется на основе рекомендаций, изложенных в [9], в качестве критерия используется средняя абсолютная ошибка (MAE). График на рис. 1 демонстрирует влияние величины мультиантитела M на точность получаемого прогноза.

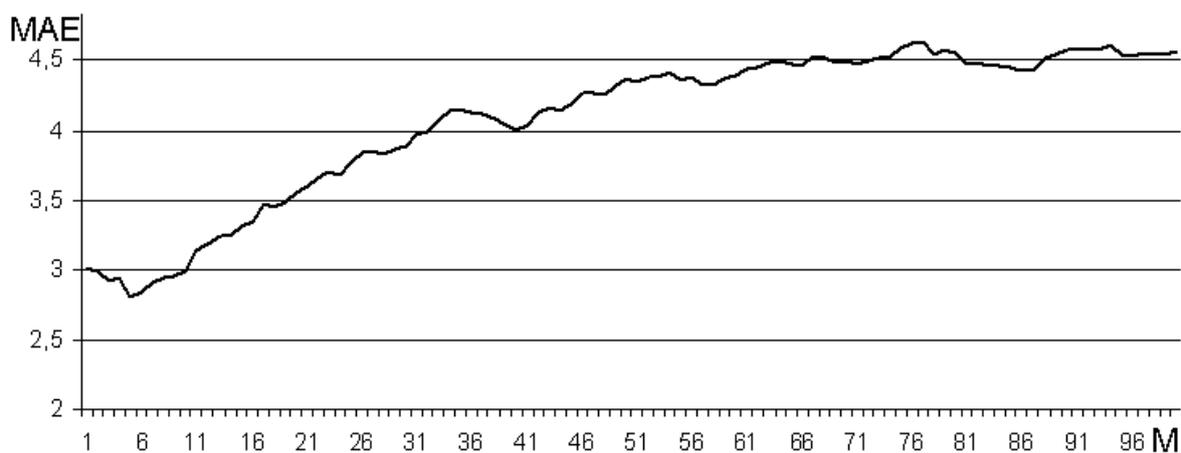


Рисунок 1 – Зависимость средней ошибки прогноза от величины мультиантитела

Алгоритм получения прогноза при помощи ИИС следующий:

1. Для создания иммунной сети используется часть известных значений временного ряда. Неиспользуемые значения выполняют роль обучающей и контрольной выборки.
2. Если позволяет размер ряда, проводится обучение сети.
3. При значительном размере сети целесообразно применение оператора супрессии – определение аффинности между мультиантителами и последующее сжатие сети.
4. На основании поставленной задачи формируется мультиантиген.
5. Определяется набор мультиантител, аффинность которых больше некоторого порогового значения. Прогнозом иммунной сети является результат мультиантитела с наибольшей аффинностью.
6. После получения реального значения, происходит коррекция сети. Отобранные мультиантитела подвергаются клонированию, в процессе которого выполняется

оператор мутации. Клоны замещают мультиантитела, их породившие. Мутации подвергается только та часть мультиантитела, которая определяет его прогноз, и не участвует в определении аффинности. Степень мутации определяется аффинностью мультиантитела.

7. Если остались нерешенные задачи прогнозирования, возврат к пункту 4.

Шаги 4-7 представляют собой обучение иммунной сети. Шаг 6 выполняется при наличии реального значения прогнозируемой величины. На рис. 2 показана зависимость точности прогноза от аффинности для необученной (график 1) и обученной ИИС (график 2). Таким образом, цель обучения ИИС заключается в настройке системы в соответствии с новыми значениями временного ряда.

Настройка параметров ИИС (таких как пороговая аффинность, величина сети и коэффициент сжатия) на конкретное приложение в подавляющем количестве применений осуществляется методами вычислительного эксперимента.

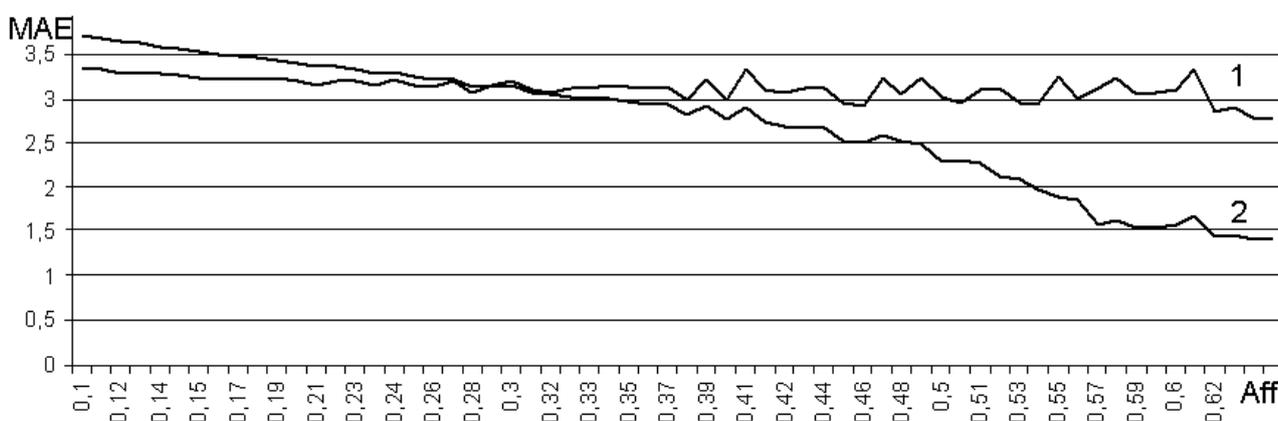


Рисунок 2. Зависимость величины ошибки от аффинности мультиантител

Результаты сравнительного анализа

В ходе экспериментальных исследований было проведено краткосрочное прогнозирование на примере рядов, используемых в M3-Competition [10], а также ряда среднесуточных показаний температуры. Для оценки прогноза использовалось значение средней абсолютной ошибки (MAE), результаты прогнозирования методами, отличными от рассматриваемого в работе, взяты из [10]. Результаты приведены в таблице 1.

Таблица 1. Ошибка прогнозирования при использовании различных методов

Метод	N2832 (104)	N2841 (104)	N2856 (92)	N2859 (79)	N2864 (79)	N2869 (79)	Meteo (21337)
Exp.Smoothing	3148,09	68,70	253,08	132,71	1028,13	497,5	2,92
Holt-Winters	3676,51	52,72	296,47	149,2	1146,32	580,31	2,9
Box-Jenkins	2814,27	68,7	260,21	125,09	896,62	210,65	2,99
ForecastPro	3098,99	68,58	260,26	132,71	1028,13	185,79	–
SmartFes	3183,35	67,71	255,99	132,71	1077,06	525,2	–
Automat ANN	1937,68	72,95	300,77	111,76	976,70	349,1	–
AINet	2848,68	151,46	216,50	130,51	1463,12	225,76	2,45

Низкий результат на рядах M3-Competition объясняется размером иммунной сети (малое количество значений временного ряда), вследствие чего вероятно отсутствие мультиантител с высокой аффинностью, при этом величина обучающей выборки не позволяет в полной мере настроить систему. Результат прогнозирования ряда значений температуры подтверждает преимущество использования обученной иммунной сети.

Выводы

В работе предложена модель прогнозирования временных рядов на основе искусственной иммунной сети, использующей метод вывода по прецедентам.

Основные особенности модели заключаются в следующем:

- решение задачи краткосрочного прогнозирования;
- необходимость обучающей выборки значительного объема;
- низкий результат при наличии тренда, желательно постоянство математического ожидания;
- чувствительность к выбросам на ранних этапах обучения сети.

Представленные в работе результаты подтверждают эффективность использования рассмотренного подхода для краткосрочного прогнозирования временных рядов, величина которых позволяет сформировать обучающую выборку достаточного объема. Использование модели для прогнозирования ряда в условиях недостатка исходной информации требует дополнительных исследований и доработки модели.

Литература

- [1] Дуброва Т.А. Статистические методы прогнозирования в экономике:
- [2] Учебное пособие, практикум, тесты, программа курса / Дуброва Т.А., Архипова М. Ю. – Московский государственный университет экономики, статистики и информатики, 2004. – 136 с. – ISBN 5-7764-0453-3.
- [3] Батуро А.П. Финансовые временные ряды: кусочное прогнозирование и проблема обнаружения предвестников существующего изменения закономерности / А. П. Батуро, Н. М. Еременко // Банковские технологии. – 2001. – № 12. – С. 70–77.
- [4] Дасгупта Д. Искусственные иммунные системы и их применение / Дасгупта Д. ; [пер. с англ. под ред. А. А. Романюхи]. – М. : ФИЗМАТЛИТ, 2006. – 344 с. – ISBN 5-9221-0706-2.
- [5] Прасолов С.В. Постановка задачи прогнозирования, основанная на применении упрощенной модели искусственной иммунной сети / С.В. Прасолов, Д.В. Шойтов // Ученые записки. Электронный научный журнал Курского государственного университета. – 2009. – Т. 2. – С. 1–3.
- [6] Бидюк П. И. Алгоритм клонального отбора для прогнозирования нестационарных динамических систем / П. И. Бидюк, В. И. Литвиненко, И. В. Баклан, А. А. Фефелов // Искусственный интеллект. – 2004. – № 4. – С. 89–99.
- [7] Самигулина Г.А. Разработка интеллектуальных экспертных систем

- прогнозирования и управления на основе искусственных иммунных систем / Г.А. Самигулина // Проблемы информатики. – Новосибирск, 2010. – № 1. – С. 15–22.
- [8] Черный С.Г. Применение case based reasoning для поддержки принятия решений / С.Г. Черный // ВЕСТНИК ХНТУ. – 2010. – № 2(38). – С. 336–342.
- [9] Поминчук Е.В. Метод интеллектуального анализа данных прецедентов для прогнозирования временных рядов / Е. В. Поминчук, А. Б. Иващенко // Материалы 4-й международной научно-технической конференции «Моделирование и компьютерная графика - 2011». – Донецк, 2011. – С. 193–197.
- [10] Чучуева И.А. Модель экстраполяции временных рядов по выборке максимального подобия / И. А. Чучуева // Информационные технологии. – 2010. – № 12. – С. 43–47.
- [11] Makridakis S. The M-3 Competition: Results, Conclusions and Implications / S. Makridakis, M. Hibon // International of Forecasting. – 2000. – № 16, С. 451–476.