

УДК 004.8

ПРОГНОЗИРОВАНИЕ ЗАБОЛЕВАЕМОСТИ С ИСПОЛЬЗОВАНИЕМ НЕЙРО-НЕЧЕТКИХ СЕТЕЙ

*Дмитриев А.Н., студент
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Медико-технические информационные технологии»*

*Научный руководитель: Котин В.В., к.ф.-м.н., доцент
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана
pvi@bmstu.ru*

Введение

Искусственные нейронные сети (ИНС) — математические модели, а также их программные или аппаратные реализации, построенные по принципу организации и функционирования биологических нейронных сетей — сетей нервных клеток живого организма [1],[2]. Нейро-нечеткая сеть является искусственной нейронной сетью с одним из разновидностей нечеткого вывода - вывода Такаги-Сугэно.

Назначение нейро-нечётких сетей – извлечение знаний. Они предназначены для реализации нечётких правил на базе нейронных сетей. Такой подход позволяет компенсировать один из главных недостатков нейронных сетей, который состоит в том, что ответ нейронных сетей является не «прозрачным».

Существует ряд классических методов прогнозирования, базирующихся на аппарате математической статистики, среди которых выделяются методы анализа и моделирования временных рядов, методы многомерного регрессионного анализа. Особенностью указанных методов является необходимость четкой спецификации конструируемых моделей, кроме того, дополнительные трудности для использования данных методов создает не стационарность исследуемых процессов.

Выбор структуры нейро-нечеткой сети

Для анализа возможностей нейросетевого прогноза заболеваемости был выбран временной ряд заболеваемости скарлатиной в 1996-2008 гг в г. Москве среди детей в возрасте от 3 до 6 лет, изображенный на рисунке 1.

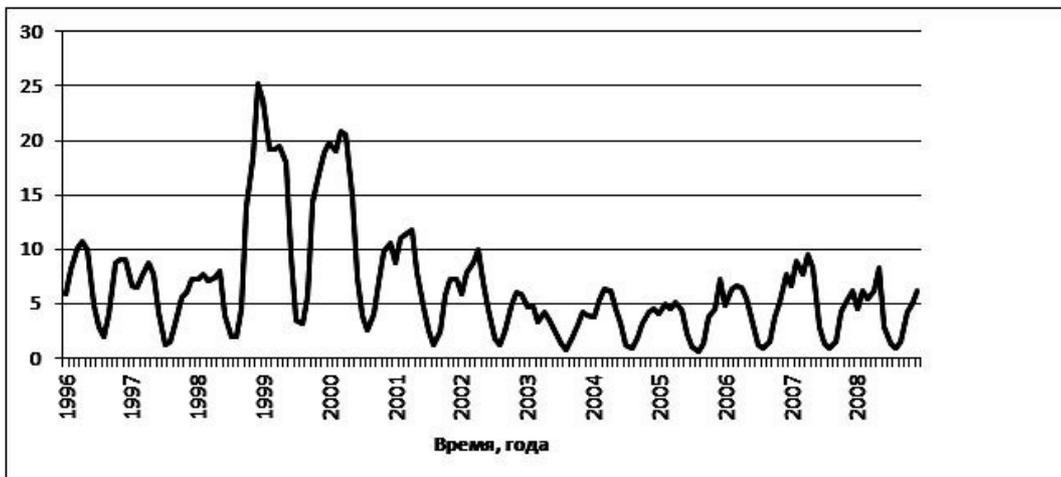


Рис. 1. Временной ряд заболеваемости скарлатиной

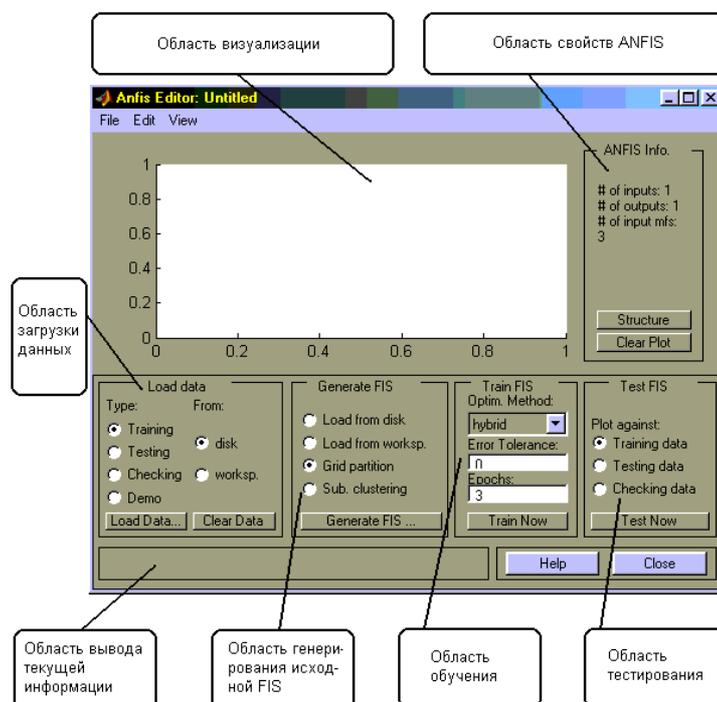


Рис. 2. Основное окно ANFIS-редактора

В качестве базового варианта была выбрана адаптивная нейро-нечеткая система (ANFIS). ANFIS-редактор, изображенный на рисунке 2, позволяет автоматически синтезировать из экспериментальных данных нейро-нечеткие сети [6]. Нейро-нечеткую сеть можно рассматривать как одну из разновидностей систем нечеткого логического вывода типа Сугэно. При этом функции принадлежности синтезированных систем настроены (обучены) так, чтобы минимизировать отклонения между результатами нечеткого моделирования и экспериментальными данными.

Для демонстрации работы ANFIS-редактора рассмотрены несколько этапов обучения нейро-нечеткой сети: ввод данных, выбор структуры сети, выбор метода обучения, обучение, проверка обучения.

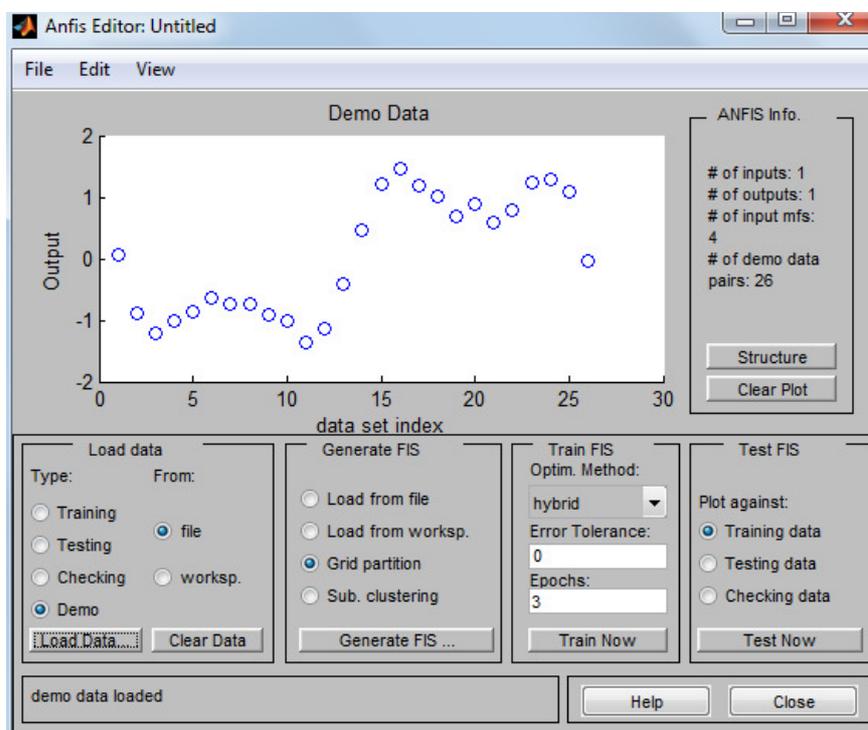


Рис. 3. Окно работы программы Anfis (Fuzzy Logic Toolbox)

Режим ввода демонстрационных данных. Для ввода демонстрационных данных в область загрузки данных выбирают Load data.

Как видно на рисунке 3, в области загрузки данных (**Load data**) расположены:

- меню выбора типа данных (**Type**), содержащее альтернативы:
 - **Training** - обучающая выборка;
 - **Testing** - тестирующая выборка;
 - **Checking** - контрольная выборка;
 - **Demo** - демонстрационный пример;
- меню выбора источника данных (**From**), содержащее альтернативы:
 - **disk** – диск;
 - **worksp.** - рабочая область MatLab;
- кнопка загрузки данных **Load Data...**, по нажатию которой появляется диалоговое окно выбора файла, если загрузка данных происходит с диска, или окно ввода идентификатора выборки, если загрузка данных происходит из рабочей области;
- кнопка очистки данных **Clear Data**.

Следующий шаг это выбор количества и типа лингвистических переменных нейро-нечеткой сети. Как видно на рисунке 4 выбраны 4 лингвистических термина вида гаусса **gaussmf** для каждого входного нейрона (в данном случае имеется один входной нейрон). Выходные лингвистические термины заданы функцией **linear**.

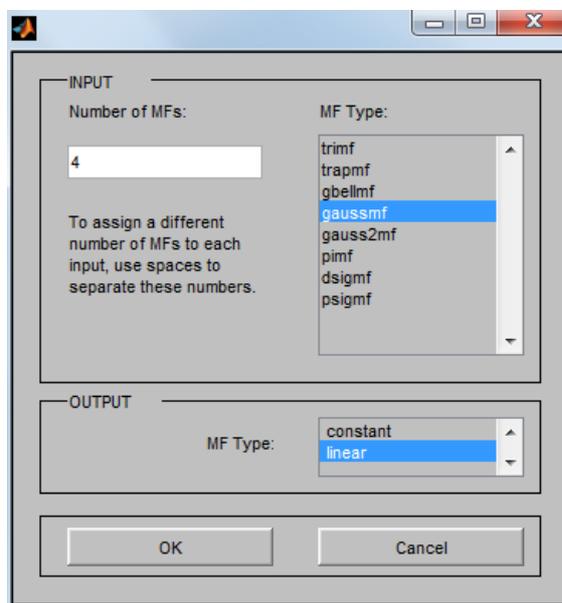


Рис. 4. Окно создания нейро-нечеткой сети

На рисунке 5 представлена структурная схема нейронной сети, параметры которой были выбраны в окне создания нейро-нечеткой сети (см. рис.4), она имеет довольно простую структуру.

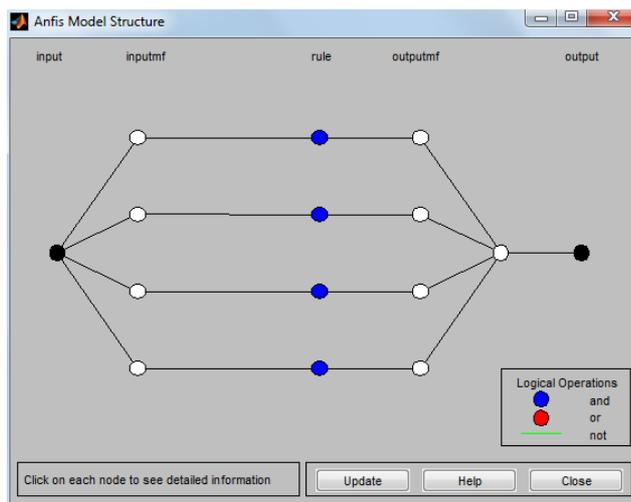


Рис. 5. Структура нейро-нечеткой сети

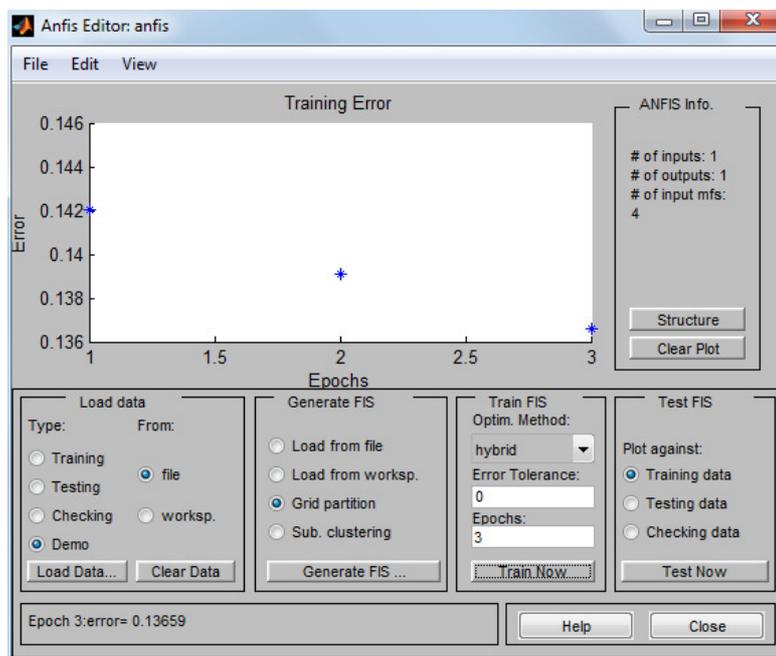


Рис. 6. Обучения нейро-нечеткой сети

В области обучения, изображенной на рисунке 6 (**Train FIS**) расположены меню выбора метода оптимизации (**Optim. method**), поле задания требуемой точности обучения (**Error tolerance**), поле задания количества итераций обучения (**Epochs**) и кнопка **Train Now**, нажатие которой запускает режим обучение. Промежуточные результаты обучения выводятся в область визуализации и в рабочую область MatLab. В ANFIS-редакторе реализованы два метода обучения:

- **backpropa** – метод обратного распространения ошибки, основанный на идеях метода наискорейшего спуска;
- **hybrid** – гибридный метод, объединяющий метод обратного распространения ошибки с методом наименьших квадратов.

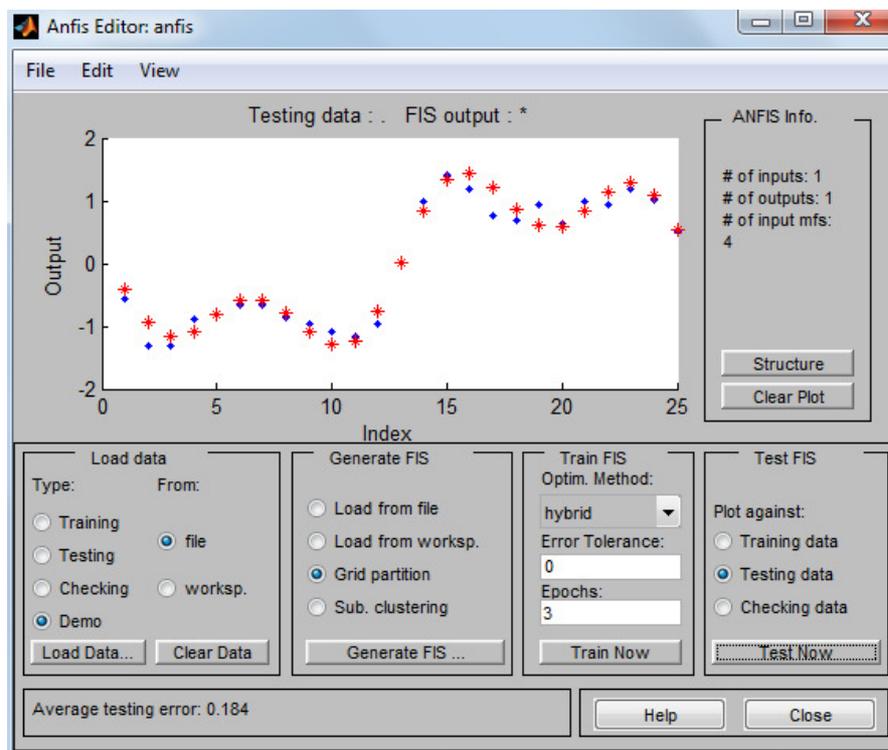


Рис. 7. Тестирование нейронной сети. Визуализация результатов обучения

Прогнозирование по полугодиям

Исходный временной ряд (см. рис.1) обладает свойствами сезонности. Также видно наличие двух пиков заболеваемости 1999-го и 2000-го гг. Как показано в работе [4] наличие этих пиков сильно влияет на прогноз заболеваемости с применение искусственных нейронных сетей (ИНС).

Важно отметить, что при моделировании динамики заболеваемости традиционно принятая в естествознании и технике методология прямой верификации теоретических выкладок по имеющемуся эмпирическому материалу сталкивается с недоступностью сколь либо обоснованной процедуры оценки погрешности входных данных. Действительно, получение каждой «экспериментальной точки» временного ряда заболеваемости, обусловлено многоступенчатой и плохо формализуемой цепочкой медицинского документооборота. Не следует также забывать, что, как правило, используются косвенные данные о численностях групп риска, переболевших и приобретших иммунитет и т.д. Таким образом, на эмпирические данные накладывается цепочка неопределённостей, специфических для медико-биологических задач, при этом множество экспериментальных точек размывается, формируя так называемые *нечёткие временные ряды* [8].

В случае редко встречающихся пропущенных значений во временном ряду может быть предложен подход сокращения временного ряда, путем пренебрежения пропущенных данных[5].

Соответственно, рассматривается временной ряд инцидентности скарлатиной без учета пиков (отсутствуют значения инцидентности с 31 месяц по 55).

Была использована выборка на основе 36-ти значений инцидентности скарлатины. Проведен анализ краткосрочного прогноза на основе малой выборки. Рассмотрены две структуры нейро-нечеткой сети: 1) структура **A** с 3 лингвистическими термами, 2) структура **B** с 4 лингвистическими термами.

Можно заметить, что тестирование сети, показано на рисунке 8, со структурой **A** с меньшим количеством лингвистических термов, дало лучшие результаты, чем структура **B**, результаты тестирования сети показаны на рисунке 11. Обучение сети **B** показывает на рисунке 9, что средняя ошибка между выходом сети и истинным значением временного ряда, крайне мала. Однако в недостатках моделей было указано, что при использовании ANFIS-редактора возникают сложности при построении модели с большим числом входным переменных (более 5-6) [7]. Как видно на рисунке 10, система, состоящая из пяти входных переменных и 3 лингвистических термов, достаточно сложна, что может повлиять на качество прогноза.

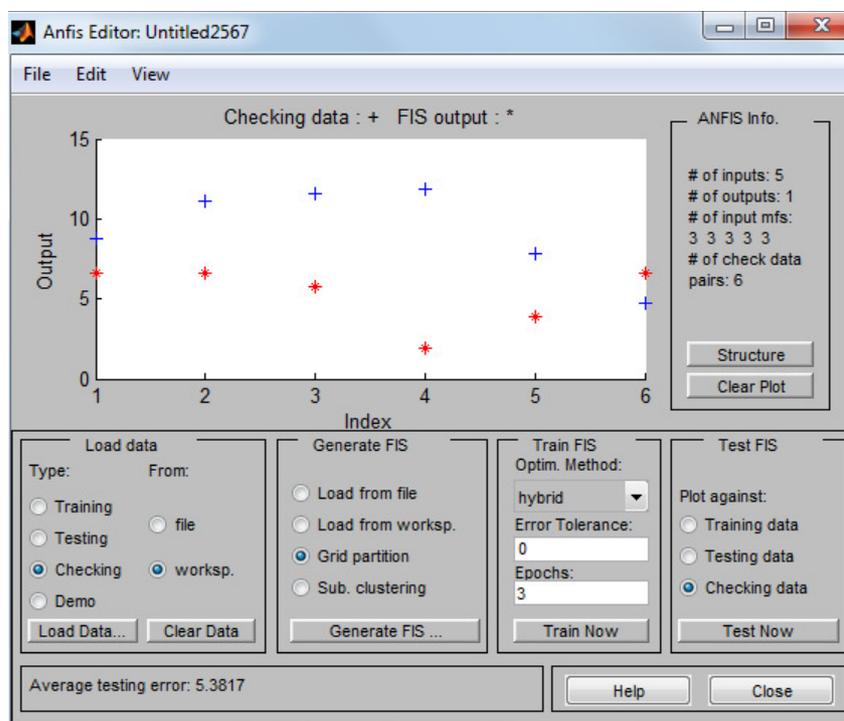


Рис. 8. Окно тестирование сети A

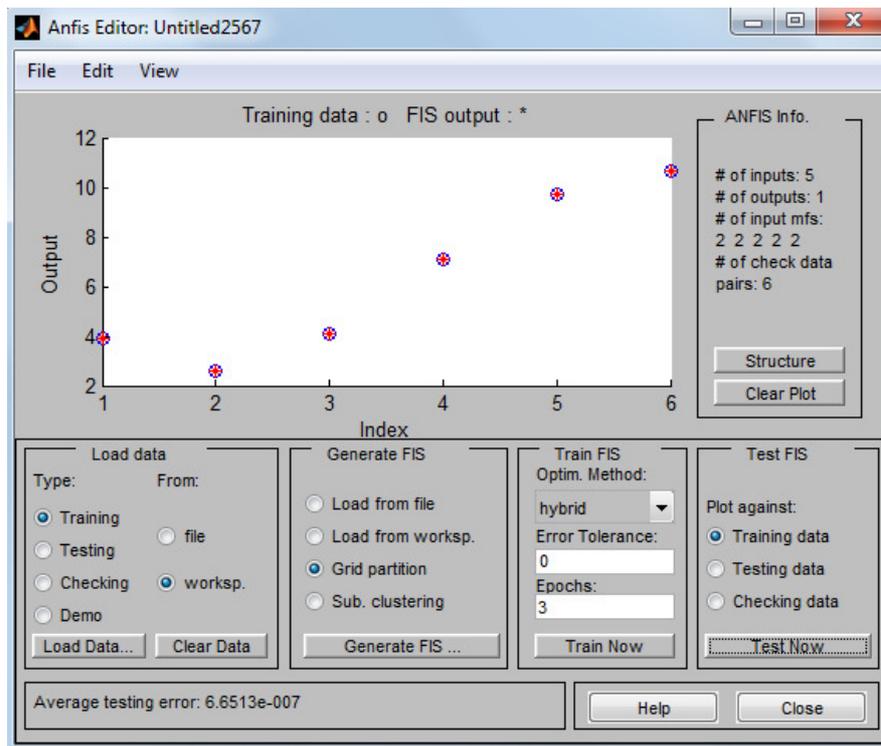


Рис. 9. Окно обучение сети В

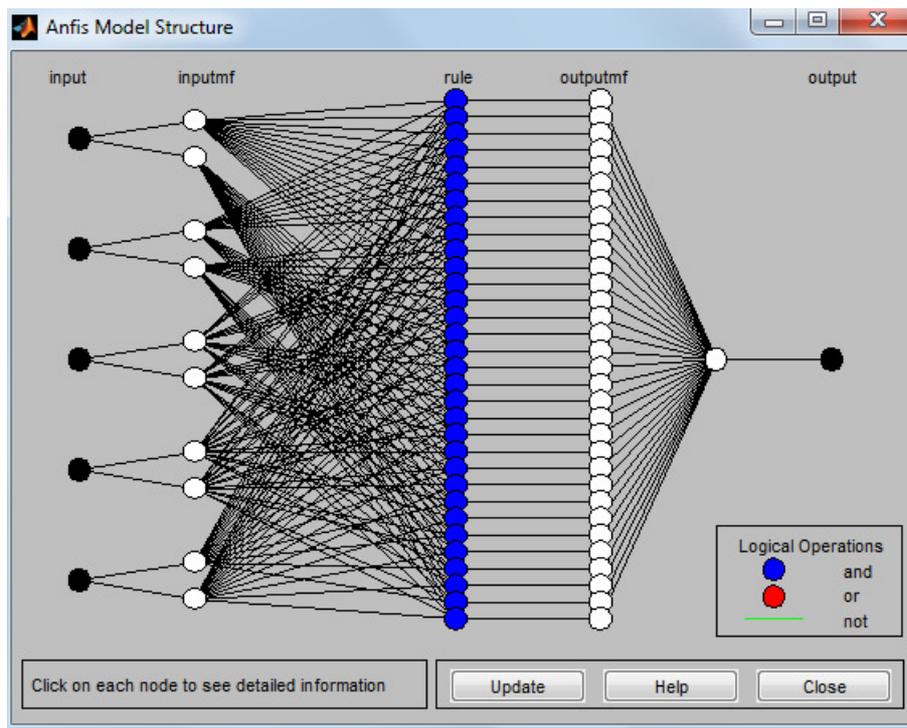


Рис. 10. Структура сети В

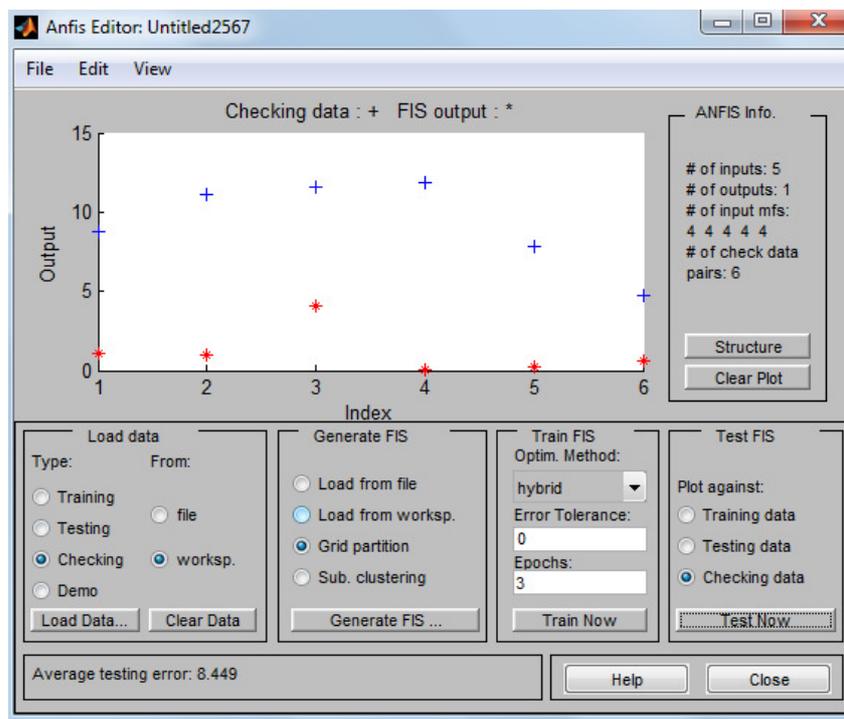


Рис. 11. Тестирование сети В

Выводы.

Адаптивно нейро-нечеткая система при прогнозировании на малой выборке, при малом количестве лингвистических термов дает лучшие результаты, чем при использовании большого количества термов. При этом средняя погрешность сети относительно велика. Некоторые значение инцидентности(около спадов) прогнозировались достаточно точно.

При долговременном прогнозе и большем количестве элементов выборки, возможно улучшение результата прогнозирования сети если, учитывать ограничения количества входных переменных.

Список литературы

1. Агупов, М.А., Эфендиева, Э.М., Гермашев, И.В. Формирование обучающей выборки для компьютерной системы анализа химических структур // Современные проблемы информатизации в экономике и обеспечении безопасности. – Воронеж: Научная книга, 2012. – С. 40–42.
2. Мак-Каллок У. С., Питтс В., Логическое исчисление идей, относящихся к нервной активности // В сб.: «Автоматы» под ред. К. Э. Шеннона и Дж. Маккарти. — М.: Изд-во иностр. лит., 1956. — с.363-384.

3. В.И. Покровский, Н.И. Брико. Общая эпидемиология с основами доказательной медицины. Руководство к практическим занятиям. –М. Геотар-Медиа, 2008 – 400с.
4. Дмитриев А. Н., Котин В. В. Моделирование временных рядов заболеваемости с использованием искусственных нейронных сетей // Медицинская техника. - 2013. - № 1. -35-38с.
5. Корчак. Т.В. Методы анализа и прогнозирования временных рядов с пропущенными данными при использовании нейро-нечетких моделей типа Такаги-Сугено и адаптивных моделей//Системы управления, навигации и связи, - 2009. - № 4
6. С.Д.Штовба "Введение в теорию нечетких множеств и нечеткую логику"
URL.<http://matlab.exponenta.ru/fuzzylogic/book1/index.php>(Дата обращения: 12.03.13)
7. Дьяконов В., Круглов В. Математические пакеты расширения MATLAB. Специальный справочник. – СПб.: Питер,201 – 480 с.
8. Н. Г. Ярушкина, Т. В. Афанасьева, И. Г. Перфильева. Интеллектуальный анализ нечетких временных рядов: учебное пособие – Ульяновск: УлГТУ, 2010. – 320 с.