

## Проблемы создания интеллектуальных средств поиска, анализа и обработки биомедицинской информации

В.О. Дунин, В.А. Егоров

Пензенская государственная технологическая академия

Целью работы является создание специализированной единой базы знаний актуальной информации, итогов последних исследований и разработок в сфере медицины, а так же возможность подключения и группирования вокруг единого информационного ядра разнонаправленных ИС не только поддержки, но и оптимизации работы по предоставлению медицинских услуг, и, как следствие, повышения качества исследований и методик лечения различных заболеваний.

Схематично, систему можно представить схемой, изображённой на рисунке 1. Центральной частью является узел агрегации информации, состоящий из группы серверов индексирования, аналитики и высокопроизводительная система управления базами данных, основной функцией которых является автоматизированный поиск, сбор, структурирование, анализ и хранение полученной полезной информации. Под полезной информацией понимается найденная индексатором информация, имеющая информативную ценность для дальнейшей обработки системой, её другими компонентами, а так же содержащую в себе информационную ценность для пользователей. Кроме, полученной от индексатора и анализатора информации, СУБД хранит анонсы подписанных исследовательских учреждений и фармацевтических компаний, получаемые через сервер веб-приложений. Так же, на серверах СУБД хранится информация по посетителям учреждений здравоохранения, формируя карточки пациента, автоматически рекомендуя лечащим врачам различные варианты медикаментов, учитывая индивидуальные особенности нуждающихся в медицинской помощи посетителей.

Связка серверов индексирования и аналитики является специализированной поисковой системой, реализация которой сводится к созданию поискового робота последовательно собирающего информацию со страниц сети Интернет, составляя копию особого формата, облегчающего анализ накопленных данных. Собранные данные должны регулярно обновляться согласно расписанию[1]. Сохранённая версия передаётся на обработку серверам аналитики, которые, из общего потока выделяют полезную информацию, и распределяют её по категориям. Для решения такой задачи предлагается использования метода кластерного. Рассматривая более широко, кластерным анализом является многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов, и затем упорядочивающая объекты в сравнительно однородные группы[2-4](кластеры). Кластерный анализ может быть осуществлён различными подходами, например, такими как вероятностный подход, логический подход на основе дерева решений, теоретико-графовый подход[5], иерархический подход, а так же подходы на основе систем искусственного интеллекта, включающих в себя метод нечёткой кластеризации, генетический алгоритм, а так же нейронную сеть[6], реализовать которую можно представленным ниже образом.

Поскольку все искусственные нейронные сети [6] базируются на концепции нейронов, соединений и передаточных функций, существует сходство между разными структурами или архитектурами нейронных сетей. Большинство отличий зависит разных правил обучения. При обучении используются эталонные данные, для получения которых требуется информация, на сбор которой необходимо достаточно много времени, поэтому, рассмотрим пример распознавания нейросетью простейшего синусоидального сигнала, представляющий собой поток данных – имеющий определенную последовательность в изменениях.

Сигнал, представленный на рисунке 2, который можно описать формулой:  $SIGNAL=x*(SIN(y)*z)$ , где  $SIN$  – это функция, описывающая синусоидальное изменение величины сигнала, а  $x,y,z$  – некоторые коэффициенты которые меняются случайным образом.

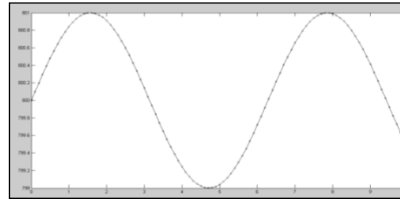
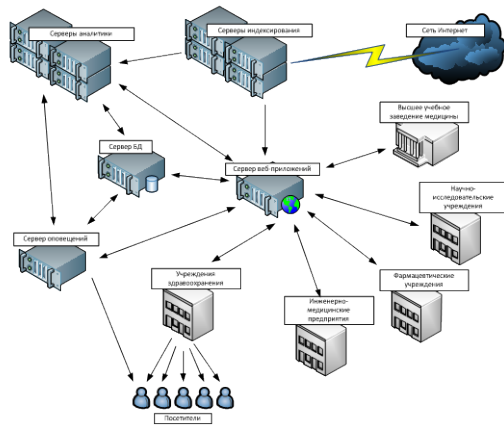


Рисунок 1 – общая схема системы Рисунок 2 – эталонный сигнал

Далее в среде Simulink необходимо спроектировать схему обработки сигнала, представленную на рисунке 2

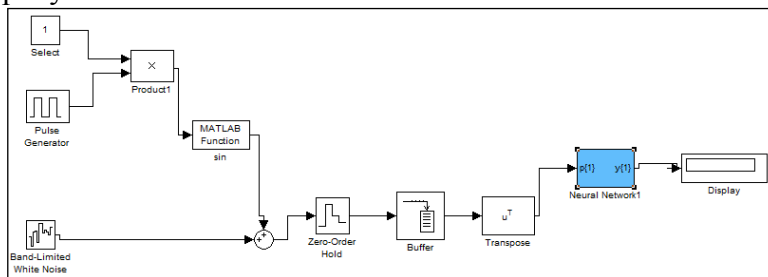


Рисунок 3 – подсистемы идентификации в среде симулинк

Для распознавания сгенерируем сеть с архитектурой перцептрона (рис. 4).

С помощью инструмента NeuralNetworkTraining анализируем работу нейронных сетей и корректируем параметры для достижения оптимального результата работы. Процесс обучения представлен на рисунке 5.

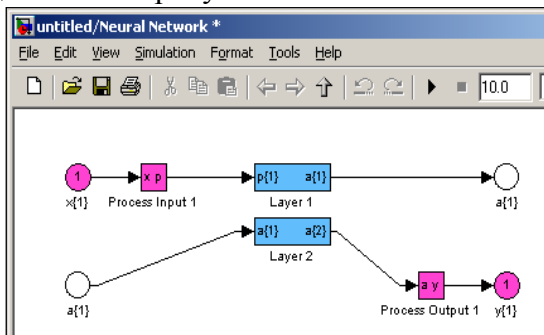


Рисунок 4 – схема слоев нейронной сети

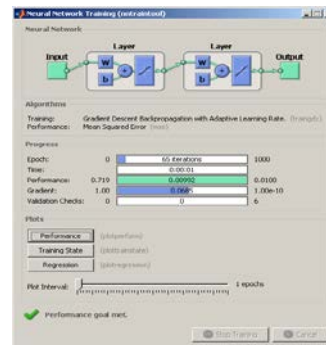


Рисунок 5 – инструмент управления ходом тренировки нейронной сети.

Производительность нейронной сети представлена на рисунке 6. Показатель характеризует производительность тренировки за определенное количество циклов. На рисунке 7 представлены результаты работы алгоритма определения ошибок нейронной сети.

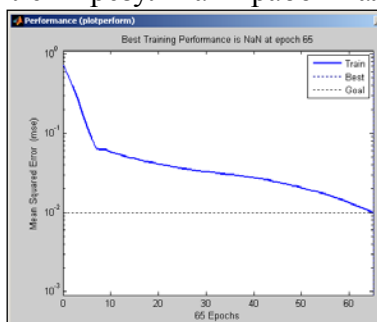


Рисунок 6 – диаграмма производительности нейронной сети

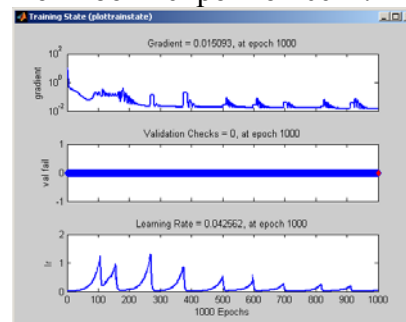


Рисунок 7 – результаты тренировки нейронной сети

Полученную модель необходимо оптимизировать методом планирования эксперимента, в результате которого получим график, представленный на рисунке 8.

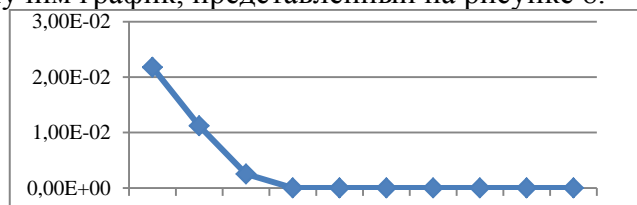


Рисунок 8 – вероятность ошибки нейронной сети при распознавании обучающей выборки сигнала

Как видно из графика на рисунке 8, вероятность ошибки достигает оптимума на 4 шаге оптимизации, но, начиная с 13 шага, происходит перетренировка сети, которая приводит к потере точности распознавания.

Итогом проведённого моделирования является выделение оптимального значения вероятности безошибочного распознавания нейронной сетью сигнала: количество слоев 2, количество нейронов 18, количество эпох 60.

Таким образом, разработанная нейронная сеть позволяет реализовать спроектированный механизм анализа данных и избежать логических ошибок и ложных срабатываний анализе потока данных различной природы, в том числе и собранных системой индексирования, применяемой для сбора информации из ресурсов сети Интернет, используемой в работе предлагаемой системы.

Полученная ИС позволит создать специализированную единую базу знаний о существующих и перспективных исследованиях в сфере медицины, а так же даст возможность подключения и группирования вокруг единого информационного узла различно-направленных ИС не только поддержки, но и оптимизации работы по предоставлению услуг здравоохранения, и, как следствие, повышения качества методов лечения различных заболеваний.

### Список литературы:

1. Компания Яндекс – Индексирование интернета  
<http://company.yandex.ru/technologies/searchindex/index.xml> (08.04.2012)
2. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Классификация и снижение размерности. – М.: Финансы и статистика, 1989. – 607 с.
3. Мандель И.Д. Кластерный анализ. – М.: Финансы и статистика, 1988. – 176 с.
4. Классификация и кластер. Под ред. Дж. Вэн Райзина. М.: Мир, 1980. 390 с.
5. А.Б. Щербань, В.А. Егоров. Об одном алгоритме поиска изоморфных отображений обобщённых структурных моделей. // Современные информационные технологии-2010: Тез. Доклмеждунар. Науч-технич. Конф., г. Пенза, ПГТА, 2010 г.- Пенза: Изд-во Пензгос технолог. Акад., 2010 – с. 17-20.
6. Круглов Владимир Васильевич, Борисов Вадим Владимирович Искусственные нейронные сети. Теория и практика. – 1-е. – М.: Горячая линия - Телеком, 2001. – с. 382.

