

данных;

– «Составление структуры БД КИС» – бизнес-процессы производственного процесса влияют на логику построения структуры данных;

– «Проектирование модели обработки банков данных» – сведения о метаданных БД являются источником набора сущностей и связей между ними, которые должна обработать модель для построения алгоритма сбора необходимых данных;

– «Формирование общих требований к разработке ПС» – выходной информацией контекстной диаграммы является техническое задание на разработку ПС обработки банков данных.

Каждый процесс отвечает за определенную область функционирования модели. Управляющими стрелками диаграммы являются: устав предприятия, модель угроз безопасности КИС, а также стандарты управления данными – ГОСТ 7.70-2003, ГОСТ 34.321-96 [3, 4].

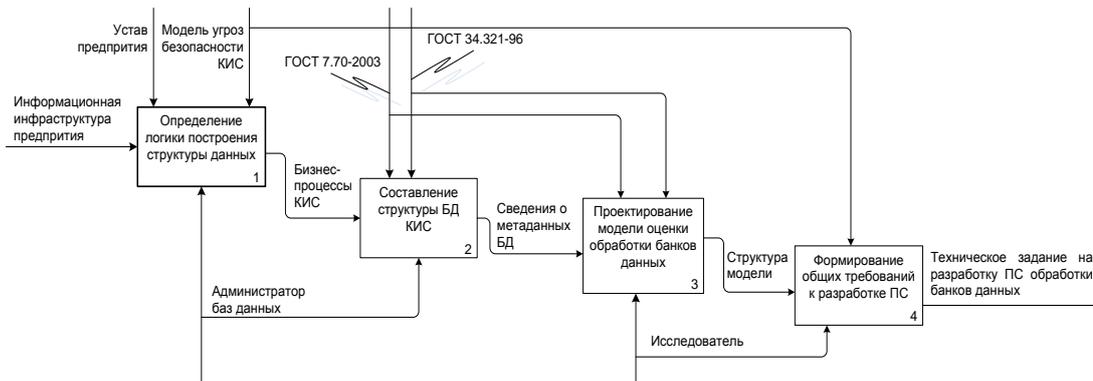


Рис. 1 – Функциональная модель сбора общих требований обработки банков данных

Вышеобозначенные процессы влияют на модель оценки обработки банков данных, и, следовательно, влияют на процессы: оценка структуры КИС, оценка структуры БД, определение набора обрабатываемых таблиц, определение набора данных для обезличивания.

Построение функциональной модели позволило формализовать общие требования к разработке программного средства обработки банков данных. Одной из актуальных и востребованных задач становится построение математической модели оценки обработки банков данных различных архитектур КИС, которая в дальнейшем будет внедрена в логику работы программного средства.

Список использованных источников:

1. Моженкова, Е. В. Проблемы и задачи при работе с корпоративными вычислительными сетями / Е. В. Моженкова, А.И. Парамонов // Компьютерные системы и сети: материалы 53-й научной конференции аспирантов, магистрантов и студентов (Минск, 2 – 6 мая 2017 г.). – Минск: БГУИР, 2017. – С. 101 – 103.
2. Моженкова, Е. В. Актуальность проблем и задач сопровождения корпоративных информационных систем / Е. В. Моженкова, А. И. Парамонов // Информационные технологии и системы 2017 (ИТС 2017) = Information Technologies and Systems 2017 (ITS 2017): материалы междунар. науч. конф. (Республика Беларусь, Минск, 25 октября 2017 года) / редкол.: Л. Ю. Шилин [и др.]. – Минск: БГУИР, 2017. – С. 270 - 271.
3. ГОСТ 7.70-2003 Описание баз данных и машиночитаемых информационных массивов.
4. ГОСТ 34.321-96 Информационные технологии. Система стандартов по базам данных. Эталонная модель управления данными.

ПОСТРОЕНИЕ ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ ДЛЯ РАСПОЗНАВАНИЯ ОБРАЗОВ

Белорусский государственный университет информатики и радиоэлектроники г. Минск, Республика Беларусь

Моисеенко И.В.

Серебряная Л.В. – к.т.н., доцент

Задача распознавания образов играет важнейшую роль в области машинного зрения, прогнозирования, управления и принятия решений в различных сферах человеческой деятельности. Одним из направлений применения алгоритмов распознавания образов является анализ временных рядов, цель которого - прогнозирование их дальнейшего развития.

Успешно зарекомендовавшим себя инструментом распознавания является искусственная нейронная сеть (ИНС). ИНС возникли на основе знаний о функционировании нервной системы живых существ. Они представляют собой попытку использования процессов, происходящих в нервных системах, для выработки новых технологических решений [1].

Существует много видов искусственных нейронных сетей с разными архитектурами (рис. 1), подходящих для распознавания образов, однако алгоритм их построения будет схож.

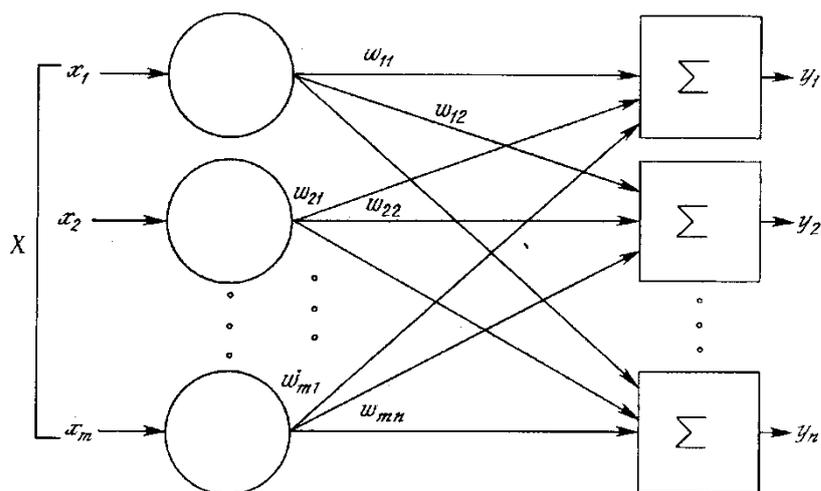


Рис. 1 - Модель искусственной нейронной сети

Построение ИНС подразделяется на следующие этапы:

1) Сбор данных для обучения:

Набор данных для обучения должен быть репрезентативным (данные должны иллюстрировать истинное положение вещей в предметной области) и непротиворечивым (противоречивые данные в обучающей выборке приведут к плохому качеству обучения сети). При необходимости применяются методы искусственного расширения обучающей выборки [3].

2) Выбор топологии сети:

Постановка задачи и имеющиеся данные для обучения определяют тип сети. Для обучения с учителем необходимо наличие для каждого элемента выборки "экспертной оценки". Если такой оценки нет (или получить ее невозможно), выбором будет ИНС, обучающаяся без учителя.

3) Подбор параметров обучения и характеристик сети:

На данном этапе происходит подбор количества слоев ИНС, кол-во нейронов в каждом слое. Чем больше суммарное число связей между нейронами, тем выше способности ИНС к обобщению. Количество нейронов в выходном слое, как правило, равно количеству определяемых классов [4]. При этом устанавливается соответствие между выходом нейронной сети и классом, который он представляет. Когда сети предъявляется некий образ, на одном из её выходов должен появиться признак того, что образ принадлежит этому классу. В то же время на других выходах должен быть признак того, что образ данному классу не принадлежит.

4) Обучение ИНС:

В процессе обучения сеть просматривает обучающую выборку. При обучении с учителем выборку делят на следующие части: обучающая (60%), валидационная (20%) и тестовая (20%). Обучающая выборка подается сети для обучения, а валидационная используется для расчета ошибки сети. При обучении сети предлагаются различные образцы образов с указанием того, к какому классу они относятся. Образец, как правило, представляется как вектор значений признаков. При этом совокупность всех признаков должна однозначно определять класс, к которому относится образец. В случае, если признаков недостаточно, сеть может соотнести один и тот же образец с несколькими классами, что неверно.

5) Тестирование качества обучения:

Тестирование качества обучения нейросети необходимо проводить на примерах, которые не участвовали в её обучении.

Наиболее подходящие существующие архитектуры ИНС для решения задачи распознавания поведенческих шаблонов на графиках временных рядов [5]:

1. Сеть Хопфилда:

Структура сети Хопфилда представляет собой однослойную рекуррентную сеть, замкнутую обратными связями [6]. Внутренняя организация сети Хопфилда, обрабатывая зашумленный и искаженный образ, подстраивает собственные параметры до тех пор, пока не приходит к некоторому стационарному состоянию (одному из эталонов) и, следовательно, позволяет решать задачи классификации, фильтрации и восстановления образов.

Преимущество - для обучения сети Хопфилда нужны только эталонные значения, на основе которых однократно подбирается матрица весовых коэффициентов.

Недостатки:

- относительно небольшой объем памяти;
- достижение устойчивого состояния не гарантирует правильный ответ сети.

2. Сеть Кохонена:

Сеть Кохонена состоит из двух слоев, причем работа сети организована таким образом, что для каждого входного образа только один из нейронов выходного слоя примет единичное значение [6]. Алгоритм обучения сети основывается на вычислении разностей между вектором синаптических связей и входным вектором и на пересчете весовых коэффициентов для нейрона, соответствующего наименьшей из

вычисленных разностей. Из этого следует, что сеть Кохонена позволяет разделить множество входных образов на несколько групп в соответствии с характерными признаками.

Преимущество - позволяет разделить множество входных образов на кластеры, обладающие некоторыми общими свойствами.

3. Многослойный персептрон:

Многослойный персептрон является одной из базовых и самых распространенных моделей нейронных сетей, позволяющих решать широкий спектр задач, одна из которых - задача распознавания образов. Многослойный персептрон относится к сетям, обучающимся на основе алгоритма обратного распространения ошибки, весовые коэффициенты подстраиваются итерационно.

Преимущество - может формировать на выходе произвольную многомерную функцию при соответствующем выборе количества слоев, диапазона изменения сигналов и параметров нейронов [1].

Недостатки:

- требуется обучающая выборка значительного объема, содержащая множество образцов каждого типа;

- для каждого из обучающих примеров в процессе обучения многократно происходит пересчет значений всех нейронов сети и связей между ними.

Выбор конкретной архитектуры ИНС для распознавания временных рядов должен быть основан на имеющемся наборе обучающих данных, требованиях к времени обучения, а также ограничений объема памяти необходимого для работы и обучения ИНС. Исходя из вышеизложенного, целесообразной является разработка такой архитектуры ИНС, которая удовлетворит требованиям задачи распознавания и использует преимущества существующих ИНС.

Список использованных источников:

1. Серебряная, Л. В. Методы и алгоритмы принятия решений: учебно-методическое пособие / Л. В. Серебряная, Ф. И. Третьяков. – Минск: БГУИР, 2016. – 64 с.
2. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume II-3/W3, 2013 CMRT13 - City Models, Roads and Traffic 2013, 12 – 13 November 2013, Antalya, Turkey, 13 - 16 с.
3. Dictionary of Philosophy of Mind [Электронныйресурс] / McCulloch, Warren Sturgis - May 2004. - Режимдоступа: <https://sites.google.com/site/minddict/mcculloch-warren-sturgis>. - Датадоступа: 17.03.2018
4. Artificial Neural Network [Электронныйресурс] / Wikipedia. – Режимдоступа: https://en.wikipedia.org/wiki/Artificial_neural_network. – Дата доступа: 18.03.2018.
5. Найман, Э. Малая энциклопедия трейдера / Э. Найман. – М.: Альпина Паблишер, 2003. – 378 с.
6. Хайкин, С. Нейронные сети: полный курс / С. Хайкин. – М.: Вильямс, 2006. – 104 с.

АНАЛИЗ ПОБОЧНЫХ ДАННЫХ СОСТОЯНИЯ СИСТЕМЫ И ДАННЫХ РЕЗУЛЬТИРУЮЩИХ ТОЧЕК СИСТЕМЫ

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Мялик И.С.

Куликов С.С. – к.т.н., доцент

Уже сейчас большинство программных системы обладают теми или иными возможностями мониторинга собственной работоспособности. Одним из видов такого мониторинга является сохранение данных состояния системы или данных результата выполнения каких-либо ключевых функций, т.е. протоколирование определенных, чаще всего, скрытых от пользователя, сведений о системе. Как правило, протокол работы сохраняется в файлы на сервере или в базы данных с определенной моделью для будущей обработки. В данный момент не до конца решенным остается вопрос обработки большого объема протоколов работы системы в реальном времени.

Допустим, перед нами стоит задача обработки большого объема протоколов работы системы размером в несколько ТБ (например, системы показа рекламы).

Одна из возможных стратегий может быть описана следующим образом:

1. Простейшей эвристикой определяются «нужные пользователи», например, по числу заходов на сайт за неделю.
2. Теперь для каждого сайта считается его посещаемость «нужными пользователями» и всеми пользователями, определяется доля «правильных пользователей» от числа всех посетителей сайта.
3. Далее сортируются сайты по убыванию рассчитанного показателя, выбираются, например, 300 сайтов, начиная с начала, на которых находится целевая аудитория.

Вероятнее всего результатом решения этой задачи будут веб-сайты содержащие один и тот же тип информации. Можно предположить, что результатом будут веб-сайты по аренде и продаже жилья. Таким образом, определяется новый этап задачи, а именно: каким образом разместить рекламные баннеры. На каких сайтах следует разместить аренду домов в Минске, а на каких продажу квартир в Ингушетии. Для выполнения этой задачи, необходимо работать с инструментами Big Data и иметь доступ к данным для их анализа.

Инструмент Big Data