



Следует отметить также, что выявленный для множества публикаций за 10 лет по базе РИНЦ индекс Хирша для сферы интеллектуальных систем равен 62, что свидетельствует, во-первых, о том, что и в дальнейшем в этой отрасли знаний следует ожидать стабильную публикационную активность учёных и специалистов, и, во-вторых, уровень научной активности российских исследователей в данной отрасли значительно превышает минимальный порог мирового уровня научной активности исследователей, равный 16 в соответствии с рекомендациями РИНЦ [8].

Литература

1. Рыбина Г.В. Основы построения интеллектуальных систем. М.: Финансы и статистика, ИНФРА-М, 2014. - 432 с.
2. Российский индекс научного цитирования. [Электронный ресурс] - Режим доступа: URL: <https://elibrary.ru/querybox.asp?scope=newquery> (дата обращения 20.01.2020 г.).
3. Арутюнов В.В. О некоторых результатах приоритетных исследований в области информационной безопасности // Научно-техническая информация. - Сер. 1. 2016. № 2. - С. 8-13.
4. Арутюнов В.В. Показатели эффективности эрготехнических систем // Научные и технические библиотеки. - 2014, № 6. - С. 5-14.
5. Арутюнов В.В. Особенности рейтинга цитируемости российских учёных по версии РИНЦ // Научные и технические библиотеки. - 2015, № 5. - С. 28-43.
6. Арутюнов В.В. Результативность научной деятельности опорных вузов России // Научные и технические библиотеки. - 2018, № 3. - С. 33-43.
7. Мазов Н.А., Гуреев В.Н., Эпов М.И. Российские публикации и журналы по наукам о земле в международных базах данных // Вестник Российской академии наук. – 2015. – Т. 85. – № 1. – С. 26–31.
8. Ершова С.К. Инструкция по использованию РИНЦ. [Электронный ресурс] - Режим доступа: URL: <https://eeip.ru/about-organization/rints.html> (дата обращения: 25.01.2020).

Л.В. Болотникова, И.В. Лёзина

РАСПОЗНАВАНИЕ РИМСКИХ ЦИФР С ПОМОЩЬЮ НЕЙРОННОЙ СЕТИ КОХОНЕНА

(Самарский университет)

Одной из наиболее важных задач, решаемых с применением нейронных сетей, является задача распознавания. Данный процесс может осуществляться для самых разнообразных и отличных друг от друга категорий объектов. Значительную долю из них занимают символьные значения.



В процессе изучения различных рукописных и печатных источников встаёт важный вопрос их оцифровки и перевода в другие форматы. Для этого необходимо иметь возможность как можно более точно распознавать текстовые символы.

При изучении текстовой информации из старинных источников, часто может возникнуть необходимость распознавания такой категории символов как римские цифры.

Для человеческого восприятия данная задача достаточно проста, но во время автоматизации этого процесса могут возникнуть сложности. Если обратиться к виду начертания римских цифр, то можно заметить, что числа даже в промежутке от 1 до 10 состоят из одних и тех же символов.

Для решения задачи распознавания римских цифр была выбрана нейронная сеть Кохонена.

Нейронные сети Кохонена — отдельный вид нейронных сетей, который используется для решения разнообразных задач классификации и связанных с ними [1].

Применение нейронных сетей Кохонена к задачам классификации требует некоторой формализации. Каждый из объектов, требующих классификации, представляется в виде некоторого вектора, идущего на вход нейронной сети. Размер входного вектора совпадает с количеством нейронов во входном слое. Размер выходного вектора равен количеству классов, на которые будут разделяться подаваемые на вход данные. Отсюда следует, что каждому нейрону в выходном слое соответствует свой класс. Выдаваемые на выходе сети значения показывают близость поданного на вход сети вектора к одному из классов [2].

Для обучения нейронной сети был выбран Алгоритм WTA (Winner Takes All - победитель получает все) [3].

Согласно алгоритму, расчёт активности каждого нейрона происходит после предъявления вектора x . Нейрон становится победителем, если он имеет самый сильный выходной сигнал, и его скалярное произведение наибольшее. В случае если вектора нормализованные, данное скалярное произведение равнозначно наименьшему эвклидову расстоянию между входным вектором и вектором весов нейронов. Нейрон, являющийся победителем, получает возможность уточнить свои веса в направлении вектора x согласно правилу [4].

$$w_w \leftarrow w_w + \eta(x - w_w)$$

Веса других нейронов не модифицируются.

Была смоделирована нейронная сеть Кохонена и выполнена её программная реализация на языке Java. В процессе разработки использовалась IDE IntelliJ IDEA.

Кроме реализации нейронной сети, был также создан генератор «зашумлённых» образов входных векторов, с предусмотренной в нём возможностью регулировать «количество» шумов. С его помощью можно создать из входных эталонных образов любое количество входных векторов для проведения обучения и последующего тестирования работы нейронной сети.



Для проведения экспериментов были созданы эталонные образы 10 римских цифр: I, II, III, IV, V, VI, VII, VIII, IX, X

Для обучений нейронной сети были сформированы несколько обучающих выборок:

Выборка №1: 100 входных векторов, образованных от каждой эталонной цифры. В сумме 1000.

Выборка №2: 500 входных векторов, образованных от каждой эталонной цифры. В сумме 5000.

Выборка №3: 1000 входных векторов, образованных от каждой эталонной цифры. В сумме 10000.

И также была сформирована одна тестовая выборка: 100 входных векторов, образованных от каждой эталонной цифры. В сумме 10000.

Результаты распознавания тестовой выборки представлены в таблице 1.

Исходя из результатов проведённых экспериментов, можно сделать вывод о том, что программа распознавания образов римских цифр, основанная на нейронной сети Кохонена эффективно справляется со своей задачей. Кроме того, была выявлена зависимость точности распознавания образов сетью от размера обучающей выборки.

Таблица 1. Результаты проведённых экспериментов.

Цифра	Выборка №1	Выборка №2	Выборка №3
I	80%	85%	95%
II	91%	95%	97%
III	80%	90%	94%
IV	82%	85%	96%
V	90%	91%	97%
VI	90%	97%	98%
VII	89%	90%	98%
VIII	90%	95%	96%
IX	93%	96%	97%
X	48%	59%	83%
Общий %	83,3%	88,3%	95,1%

Эксперимент также показал, что наиболее подвержены ошибкам распознавания римские цифры I и X, что возможно объясняется тем что большинство из рассматриваемой категории символов состоит из одних и тех же элементов.

Литература

1. <http://www.aiportal.ru/articles/neural-networks/network-kohonen.html> [Электронный ресурс].
2. Хайкин, С. Нейронные сети: полный курс [Текст]/С. Хайкин. – М.: ООО “И. Д. Вильямс”. 2006. – 1104 с



3. Осовский, С. Нейронные сети для обработки информации [Текст] / Осовский С.: Пер. с польского И.Д. Рудинского. – М.: Финансы и статистика, 2002. – 344 с..

4. Солдатова, О.П. Основы нейроинформатики [Текст] : учеб. пособие / О.П. Солдатова. – Самара: Изд-во Самар, гос. аэрокосм, ун-та, 2006. – 132 с. : ил.– ISBN 5-7883-0467-9.

Р. Р. Вафин, Р. В. Насыров

ОБОСНОВАНИЕ ВЫБОРА АЛГОРИТМА СИНТАКСИЧЕСКОГО АНАЛИЗА ПРЕДЛОЖЕНИЙ ДЛЯ ПОСТРОЕНИЯ НЕПРОЕКТИВНОГО ДЕРЕВА

(Уфимский государственный авиационный технический университет)

В настоящее время задача извлечения информации из текста на естественном языке является актуальной во многих сферах. Например, в медицине извлечение информации применяется для задач анализа текста медицинских выписок, историй болезни, формализации различных медицинских документов, например клинических рекомендаций.

С точки зрения лингвистического подхода извлечение информации состоит из нескольких этапов: лексический, морфологический, синтаксический и семантический анализ [1]. Для текстов на русском языке первые два этапа не представляют собой сложности, так как существующие инструменты позволяют с довольно большой точностью проводить лексический и морфологический анализ.

Существенная проблема встречается на этапе синтаксического анализа. Синтаксический анализ это построение грамматической структуры предложения и выявление синтаксических связей между словами в предложении [2]. Синтаксический анализатор должен построить синтаксическую структуру предложения. В автоматическом синтаксическом анализе, как правило, применяют синтаксические деревья, которые отражают структуру и/или связи в предложении.

В данной работе ставится задача разбора и сравнения методов синтаксического анализа предложений для определения наиболее подходящих из них в качестве инструмента анализа предложений на русском языке.

Существуют два наиболее распространенных подхода к отражению синтаксической структуры предложения: деревья непосредственно составляющих и деревья зависимостей.

Дерево непосредственно составляющих состоит из нетерминальных символов (различных групп, например именная, предложная, глагольная т.д.) и терминальных символов, самих слов предложения. Для построения деревьев такого типа преимущественно используются алгоритмы, основанные на формальной грамматике, в частности на контекстно-свободных грамматиках Хомского. Существенным ограничением является то, что дерево составляющих