



ОТКРЫТАЯ НАУКА
издательство

Международный журнал информационных технологий и
энергоэффективности

Сайт журнала:

<http://www.openaccessscience.ru/index.php/ijcse/>



УДК 519

АВТОМАТИЗАЦИЯ НАПИСАНИЯ ДОКУМЕНТАЦИИ С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

Цепелев Ю.А., Раскатова М.В.

Федеральное государственное бюджетное образовательное учреждение высшего образования «Национальный исследовательский университет «МЭИ», Россия (111250, г.Москва, ул. Красноказарменная, д. 14); e-mail: foxsidark@yandex.ru

В статье рассмотрен способ наполнения словаря для написания документации новыми словами с использованием искусственных нейронных сетей.

Ключевые слова: искусственные нейронные сети, алгоритм, словарь.

AUTOMATING WRITING DOCUMENTATION USING ARTIFICIAL NEURAL NETWORKS

Tsepelev Y. A., Raskatova, M. V.

National Research University "Moscow Power Engineering Institute", Russia (111250, Moscow, Krasnokazarmennaya street, 14); e-mail: foxsidark@yandex.ru

The paper describes the method of filling the dictionary for writing documentation with new words using artificial neural networks.

Keywords: artificial neural networks, algorithm, dictionary.

В современном мире огромную роль играет документооборот и разнообразная отчетность. Для её написания требуются значительные временные затраты. Как правило, многие из отчетных документов имеют примерно одинаковый шаблон, меняются лишь небольшие части этих документов. Но изменение составных частей влияет на остальной текст, следовательно, возникает необходимость в автоматизации написания таких документов поскольку обычные шаблоны недостаточно эффективны. Для решения данной задачи интересным является применение искусственных нейронных сетей. В данной статье исследуется возможность применения алгоритма, разработанного на базе искусственных нейронных сетей, для добавления слов в словарь. Выявлены недостатки и возможные пути развития представленного алгоритма.

Идея создания искусственных нейронных сетей пришла нам из биологии, так человек может легко справляться с достаточно широким пластом задач связанными с распознаванием

образов и многими другими задачами, на которые традиционные алгоритмы тратят огромное количество вычислительных ресурсов. Так желание перенести эту способность человека привело к созданию математической модели искусственных нейронных сетей.

«Нейронная сеть – это громадный распределенный параллельный процессор, состоящий из элементарных единиц обработки информации, накапливающих экспериментальные знания и предоставляющих их для последующей обработки. Нейронная сеть сходна с мозгом с двух точек зрения.

Знания поступают в нейронную сеть из окружающей среды и используются в процессе обучения.

Для накопления знаний применяются связи между нейронами, называемые синаптическими весами.»

В свою очередь искусственная нейронная сеть состоит из нейронов [1], рисунок 1.

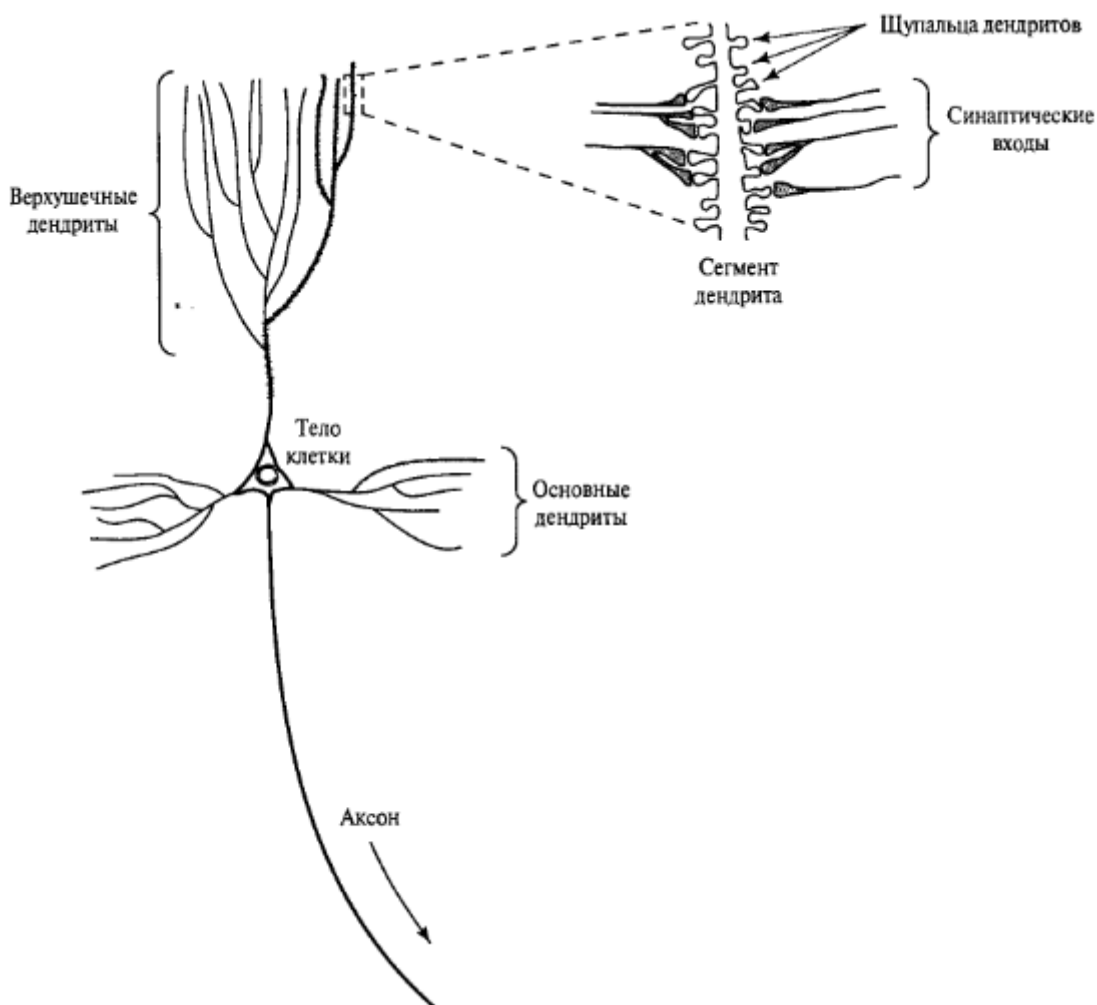


Рисунок 1 – Биологическое представление нейрона

Для нас же интерес представляет математическая модель нейрона [2] (рисунок 2), включает в себя следующее:

- 1) набор синапсов(связей), каждый синапс имеет собственный вес;
 - 2) сумматор, который складывает входные сигналы;
- функция активации, которая нормализует выходной сигнал.

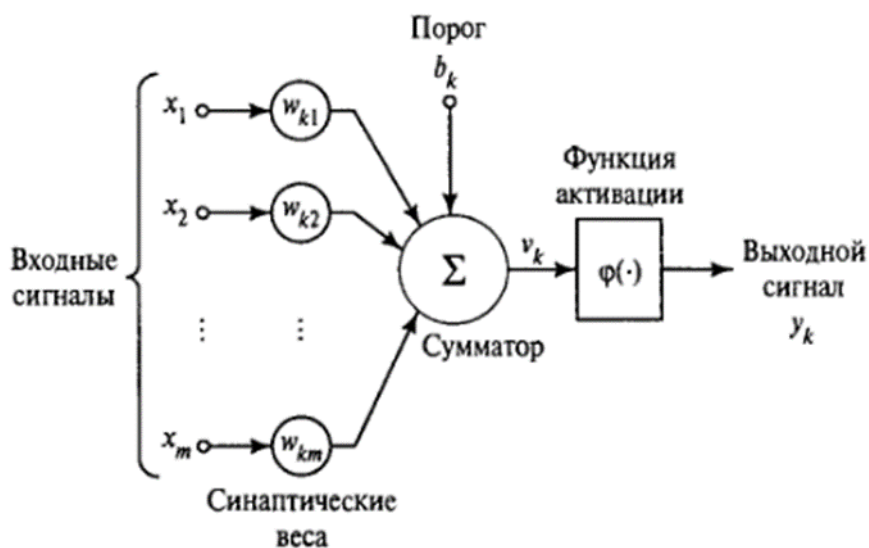


Рисунок 2 – Математическая модель нейрона

Существуют различные алгоритмы, которые помогают в написании текста самым известным является Т9. Данный алгоритм позволяет предсказывать слова, которые следуют друг за другом, а также дописывать слова, написанные частично. Однако, у данного алгоритма имеется недостаток. Нет возможности добавлять свои слова в словарь.

При написании различной документации часто используются характерные сокращения, которых нет в исходном словаре. Но эту проблему можно решить, используя алгоритмы с возможностью самообучения. Такие алгоритмы позволяют адаптироваться к написанию различного рода документам, а также добавлять новые слова в словарь.

В данной работе рассмотрим алгоритм, который позволяет добавлять новые слова в словарь.

Для демонстрации работы алгоритма разработана специальная программа. В качестве среды программирования выбрана Visual Studio и язык С# [4]. Поскольку в дальнейшем данный алгоритм планируется использовать в кроссплатформенном приложении, разработанном с использованием Unity. В этом случае Visual Studio предоставляет наиболее удобный инструментарий для разработки. Также поскольку для языка С#, отсутствуют завершённые рабочие библиотеки для работы с нейронными сетями, а у имеющихся тестовых библиотек функционал слишком избыточен.

Поэтому была разработана специальная библиотека для построения искусственных нейронных сетей с простой топологией.

Библиотека включает в себя два класса для работы с искусственной нейронной сетью:

- 1) класс Dendrite – позволяет соединять между собой нейроны;
- 2) класс Neuron – включает себя функционал для работы нейрона и его обучения.

Для обучения использовался алгоритм обратного распространения ошибки. Данный алгоритм является разновидностью алгоритмов, которые базируются на принципе обучения с учителем.

Алгоритм базируется на искусственной нейронной сети с топологией [3] (рисунок 3), которая позволяет проверять на совпадения слова. Сам алгоритм состоит из использования большого количества искусственных нейронных сетей для распознавания слов на вход подается вектор из текущих значений того на сколько текущий символ близок к тому, что

содержится в слове на данной позиции. Для создания такого вектора существует большое количество алгоритмов.

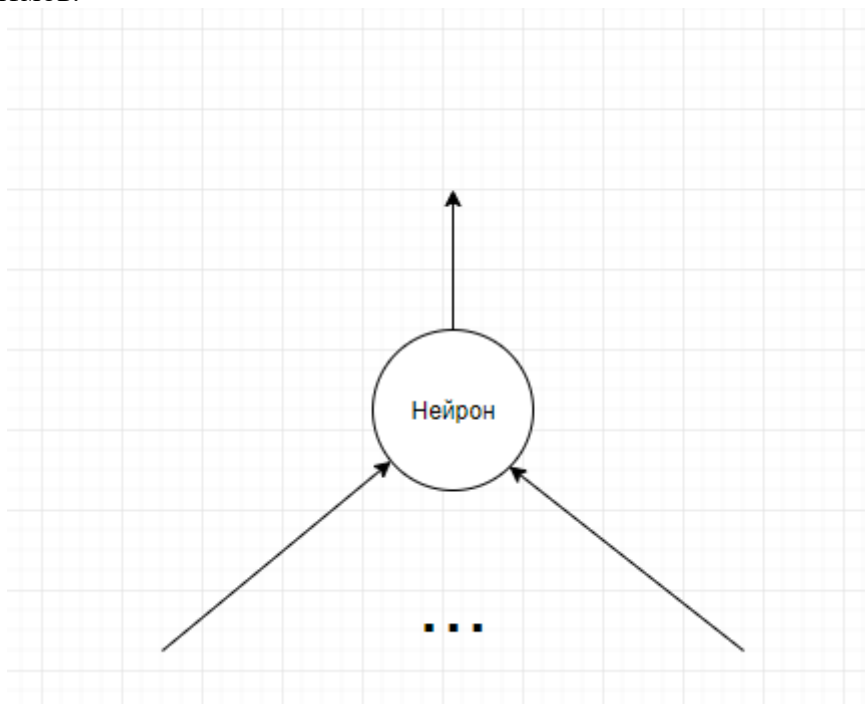


Рисунок 3 – Однослойная искусственная нейронная сеть прямого распространения

В данной работе мы используем простой алгоритм [1], которого вполне достаточно для тестового варианта.

$$\frac{C_i^p}{|C_i^p + |C_i^p - C_i^{in}||} \quad (1)$$

Где C_i^p – числовое представление символа в слове. C_i^{in} – числовое представление входного символа.

Проверка работы алгоритма проведена на следующих тестах (рисунок 4).

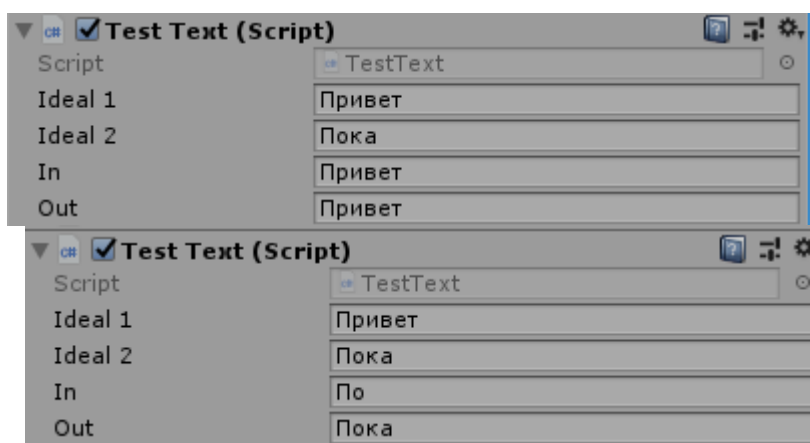


Рисунок 4 – Тесты

В результате тестирования были выявлены следующие ограничения алгоритма. Поскольку все символы имеют одинаковый вес в независимости от их позиций, а размер входного вектора зависит от размера слова более короткие слова имеют больший приоритет на срабатывания.

Вывод: данный алгоритм нельзя в полной мере использовать для добавления слов в словарь. Однако, в результате было выявлено несколько путей развития алгоритма, для того, чтобы начать его полноценно использовать. Самый очевидный путь добавление нормализации, чтобы слова с меньшим количеством букв имели одинаковый приоритет срабатывания с более длинными словами.

Список литературы

1. Хайкин Саймон Нейронные сети: полный курс, 2-е издание, ; Пер. с англ. – М. : Издательский дом “Вильямс”, 2006. – 1104 с. : ил. – Парал. тит. англ.
2. Фрэнк Розенблатт: Принципы нейродинамики. Перцептроны и теория механизмов мозга.
3. Рашид Т. Создаем нейронную сеть. : Пер. с англ. Санкт-Петербург. : ООО ”Альфа-книга”, 2017 – 272 с.
4. Шилдт Герберт C# 4.0: полное руководство.: Пер. с англ. – М.: ООО “И.Д. Вильямс”, 2011. – 1056 с.: ил. – Парал. тит. англ.
5. Мартин Р., Мартин М. Принципы, паттерны и методики гибкой разработки на языке C#. – Пер. с англ. – СПб.: Символ-Плюс, 2011. – 768 с., ил.

References

1. Simon Haykin Neural networks: a complete course, 2nd edition ; TRANS. from English. – Moscow : Publishing house " Williams", 2006. – 1104 p.: Il. – Paral. Titus. English.Volkova V. N, Denisov A.A.Theor of systems: - M: the Higher school, 2006. (in Russian)
 2. Frank Rosenblatt: Principles of neurodynamics. Perceptrons and theory of brain mechanisms.
 3. Rashid T. Creating a neural network. : Per. with English. Saint-Petersburg. : Alfa-kniga LLC, 2017 – 272 p.
 4. Shield Herbert C# 4.0: a complete guide.: Per. with English. – M.: LLC “I. D. Williams”, 2011. – 1056с.: Il. – Paral. Titus. English.
 5. Martin R., Martin M. Principles, patterns and techniques of agile development in language C#. – Per. with English. – SPb.: Plus Symbol, 2011. – 768 p., Il.
-