

УДК 004.94

ПОИСК КОЛЛОКАЦИЙ В ТЕХНИЧЕСКОЙ ДОКУМЕНТАЦИИ НА ЭЛЕКТРОННЫЕ СРЕДСТВА НА ОСНОВЕ НЕЙРОННОЙ СЕТИ С ГЕНЕТИЧЕСКИМ АЛГОРИТМОМ ОБУЧЕНИЯ

© Яковлев Д.Р.

e-mail: denis21.97@mail.ru

*Казанский национальный исследовательский технический университет
имени А.Н. Туполева – КАИ, г. Казань, Республика Татарстан, Российская Федерация*

В настоящее время актуальной проблемой является автоматическое выявление необходимых словесных конструкций из текстов. Хотя в данное время и существует национальный корпус русского языка, где коллокации можно выявить с помощью статистических методов, но все же некоторые коллокации могут быть обнаружены только с помощью эвристических методов. В данной работе для решения этой задачи используется нейронная сеть, которая обучается с использованием генетического алгоритма.

Коллокацией принято называть сочетание двух или более слов, которые имеют тенденцию к совместной встречаемости [1]. В данной задаче мы будем рассматривать такие коллокации, где одно из слов – это омоним, а другое слово в паре «разрешает» эту омонимию. К примеру, рассмотрим два словосочетания: красный угол; красное пятно. Видно, что слово «красный» может иметь различный смысл, но именно по второму слову мы понимаем, какой именно смысл слово приобретает в той или иной ситуации.

Задача поиска коллокаций остается актуальной, поскольку на сегодняшний день отсутствуют абсолютно точные способы их выявления, поэтому зачастую используются эвристические методы. В данном докладе рассматривается нахождение коллокаций с помощью нейронной сети, которая обучается с помощью генетического алгоритма.

Нейронная сеть – это математическая модель, которая построена по принципу организации и функционирования биологических нейронных сетей, которая представляет собой систему из соединенных между собой и взаимодействующих нейронов [2]. На вход нейронной сети поступают выбранные заранее характеристики «претендента» на коллокацию. На выходе нейронной сети принимается ее решение о том, является ли данное словосочетание коллокацией или нет.

Обучение нейронной сети производится с помощью генетического алгоритма. Он представляет собой адаптивный метод поиска, который очень часто используют при решении задачи функциональной оптимизации. Он основан на генетических процессах биологических организмов.

Для того чтобы нейронная сеть обучилась, необходимо сформировать обучающую выборку. Для этого нейронная сеть в ее текущем состоянии производит поиск «претендентов» на коллокации, а затем эксперт решает, что из полученного списка является коллокациями. Затем эти данные отправляются на обучение нейронной сети, после чего она адаптируется и меняет свои внутренние параметры в соответствии с обучающей выборкой.

Таким образом, рассмотрен эвристический метод поиска коллокаций с помощью нейронной сети и генетическим алгоритмом ее обучения.

Библиографический список

1. Павельева Т. Ю. Изучение коллокаций на основе лингвистических корпусов текстов // Вестник ТГУ. 2016. №3–4. С. 155–156.
2. Беркинблит М. Б. Нейронные сети. М.: МИРОС и ВЗМШ РАО. 1993. 96 с.