

МЕТОД АВТОМАТИЗИРОВАННОЙ УНИВЕРСАЛЬНОЙ ДЕСЯТИЧНОЙ КЛАССИФИКАЦИИ В СФЕРЕ ДИСТАНЦИОННОГО ОБУЧЕНИЯ

Ф. И. Третьяков, Л. В. Серебряная

*Белорусский государственный университет информатики и радиоэлектроники
Минск, Беларусь, Fiodor.Tretyakov@gmail.com, l_silver@mail.ru*

Abstract. There are presented new method of automatic decimal classification. It is based on an artificial neural network that is learning with a teacher. The algorithm should be wrapped into a secured service, which will accept data in some well-recognizable text format. The offered method will allow to receive more exact classification of papers, and also to cut down expenses on performance of the considered procedure.

Дистанционное обучение является ярким примером того, как различные образовательные процессы переносятся в виртуальное информационное пространство. Обязательным атрибутом любой печатной научной работы, в том числе и выложенной в Интернет, является универсальный десятичный классификатор (УДК). С его помощью выполняется классификация информации, необходимая во всем мире для систематизации произведений науки, литературы и искусства, периодической печати, различных видов документов и организации картотек [1].

В настоящее время УДК назначается вручную на основе специальных справочников. Данная работа посвящена методам и средствам, позволяющим решить задачу автоматически присваивать работе УДК, не привлекая к этому человека. Одним из методов ее решения является построение искусственной нейронной сети (ИНС), обучаемой с учителем. Рассмотрим алгоритм обучения ИНС.

1. Инициализировать случайным образом небольшими ненулевыми значениями весовые коэффициенты нейронов и параметры функции. В качестве функции активации будет использоваться сигмоидальная передаточная функция [2].

2. Подать на вход сети обучающий образ и рассчитать выходное значение.

3. Вычислить ошибку классификации и сравнить ее с пороговым значением.

4. Если ошибка классификации превышает допустимое пороговое значение, то изменить весовые коэффициенты так, чтобы ошибка классификации уменьшилась. Иначе обучение завершено. Повторить шаги 2–4.

Рассмотрим алгоритм работы сервиса с применением ИНС.

1. Пользователь регистрируется в сервисе и отправляет на него свою работу.

2. Текст подается на вход ИНС с целью классификации работы.

3. Классификация базируется на всех предыдущих результатах обучения ИНС. УДК удобно представить в виде дерева, в котором классификация будет выполняться посредством перехода от ветви к ветви до достижения листа. На каждом шаге выбирается наиболее подходящая ветвь. Поскольку УДК включает в себя чрезвычайно много областей науки, то нейронами в ИНС являются целые множества ключевых слов, а не отдельные ключевые слова. Если переход на любую другую ветвь либо лист осуществляется с вероятностью ниже заданного порога, то итогом будет данная ветвь.

4. В результате классификации пользователь получает УДК, который будет закреплен за его работой в базе.

Предложенный метод позволит получить более точную классификацию работ, а также сократить расходы на выполнение рассмотренной процедуры.

Литература

1. Дрезен, Э. К. За всеобщим языком: три века исканий / Э. К. Дрезен.—3-е изд.— Москва: Едиториал УРСС, 2012.— 272 с.

2. Серебряная, Л. В. Нейросетевой подход к распознаванию образов / Л. В. Серебряная, О. А. Шушина // Доклады БГУИР.— Минск, 2009.— № 8 (46).— С. 85–92.