

ИСПОЛЬЗОВАНИЕ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ КЛАССИФИКАЦИИ И АНАЛИЗА ТОНАЛЬНОСТИ ТЕКСТОВ

Нестеренков С. Н., Федоров П. А., Денисов В. А.

Кафедра программного обеспечения информационных технологий, кафедра информатики, Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь

E-mail: nsn@bsuir.by, pavl.fedorov@gmail.com, q1munys@gmail.com

В последние несколько лет применение сверточных нейронных сетей становится всё более широким. В данной работе приведены основные задачи и методы классификации текста с использованием сверточных нейронных сетей. Рассмотрена структура сверточных нейронных сетей и их принцип работы.

ВВЕДЕНИЕ

Отнесение документов к определённой категории на основании его содержимого называется классификацией документа. Данная задача является одной из задач информационного поиска. Процесс классификации может осуществляться как полностью в ручную так и с помощью применения методов машинного обучения в частности сверточных нейронных сетей. Также следует отличать классификацию от кластеризации, где в последнем случае тексты тоже группируются по категориям, которые заранее определены. Задача классификации текстов становится все более востребованной в связи с постоянным ростом информации в интернете и необходимостью в ней ориентироваться. Например, задача классификации текстов применима к решению следующих задач:

1. Персонафикация рекламы;
2. Разделение сайтов по тематическим каталогам;
3. Борьба со массовой рассылкой корреспонденции рекламного характера;
4. Распознавание тональности текстов.

В тоже время задача интеллектуального анализа текстовой информации, которая способна определять автора и пол автора текста, возраст, уровень образования, эмоциональное состояние автора в момент написания текста также является актуальной задачей. Под тональностью будем понимать эмоционально окрашенную лексику и эмоциональную оценку, выраженную автором относительно чего-либо. Анализ тональности имеет важное практическое применение:

1. Оценка качества товаров и услуг на основании отзывов пользователей интернет-ресурсов;
2. Противодействие экстремизму и терроризму;
3. Анализ ситуации на фондовых рынках и прогнозирование волатильности финансовых активов;
4. Составление текстов с заранее заданными эмоциональными характеристиками.

I. ПОСТАНОВКА ЗАДАЧИ

Задача классификации текстовой информации определяется следующим образом. Пусть существует конечное множество категорий $C = \{c_1, c_2, \dots, c_m\}$, конечное множество документов $D = \{d_1, d_2, \dots, d_m\}$ и неизвестная целевая функция Φ , определяющая соответствие для каждой пары <документ, категория> $\Phi : D \times C \rightarrow \{0,1\}$. Задача состоит в нахождении функции Φ' , которая является максимально близкой к целевой функции Φ . Эта функция называется классификатором.

Существует различные виды классификации текстов:

1. Субъективность/объективность
2. Классификация по многополосной шкале;
3. Классификация по бинарной шкале;
4. Классификация по многополосной шкале.

Основной задачей анализа тональности является выявление мнений в тексте и определить их свойств. Мнения делятся на два типа:

1. сравнение;
2. непосредственное мнение.

Непосредственное мнение содержит высказывания автора об одном объекте. Более формальное определение мнения выглядит как кортеж из пяти элементов (op, e, f, t, h), где:

1. orientation или polarity — тональная оценка;
2. (entity, feature) — объект тональности e или его свойства f;
3. момент времени time, когда было оставлено мнение;
4. holder — субъект тональности (автор).

Примеры тональных оценок:

1. нейтральная;
2. негативная;
3. позитивная.

Нейтральная оценка подразумевает отсутствие какого-либо эмоционального окраса. Также могут существовать и другие оценки.

Чаще всего в современных системах автоматизированного определения тональности текста применяется одномерное эмотивное пространство: позитив или негатив (хорошо или плохо).

II. НЕЙРОННЫЕ СЕТИ

Нейронная сеть представляет собой систему соединённых и взаимодействующих между собой простых процессоров. На рисунке 1 изображена схема нейронной сети, где голубым — скрытые нейроны, зелёным цветом обозначены входные нейроны, жёлтым — выходной нейрон.

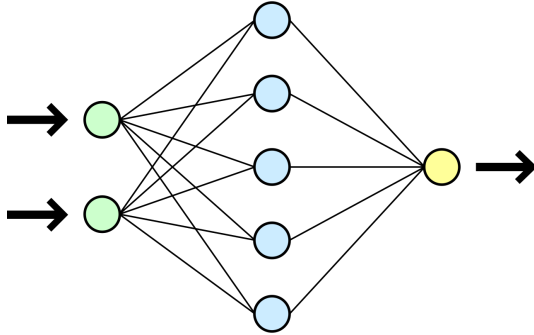


Рис. 1 – Схема простой нейронной сети.

По характеру обучения нейронные сети бывают:

1. использующие обучение без учителя;
2. использующие обучение с учителем.

Многослойная нейронная сеть состоит из слоев:

1. Входной - определение входных признаков;
2. Скрытые - расчет промежуточных результатов;
3. Выходной - вывод окончательного значения, вычисленного по гипотезе.

III. СВЕРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ

Применение нейронных сетей становится всё более широким. Особую популярность получили сверточные нейронные сети. Типовая архитектура сверточной нейронной сети изображена на рисунке 2.

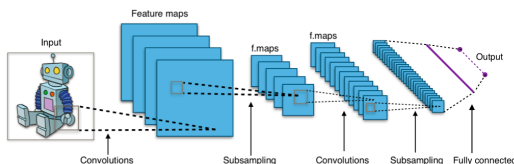


Рис. 2 – Типовая архитектура сверточной нейронной сети

Сверточные нейронные сети - специальная архитектура искусственных нейронных сетей.

Название сеть получила из-за наличия операции свертки, которая изображена на рисунке 3.

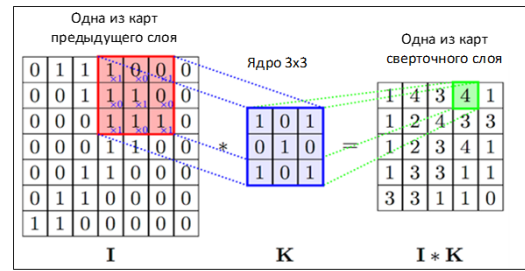


Рис. 3 – Операция свертки и получение значений сверточной карты

Особое внимание свёрточные нейронные сети получили после конкурса ImageNet, посвящённого распознаванию объектов на фотографиях. Победитель данного конкурса Алекс Крижевский, используя сверточную нейронную сеть, значительно превзошёл остальных участников. После успеха применения свёрточных нейронных сетей к классификации изображений привело к попыткам использования данного метода и к другим задачам. В последнее время их стали активно использоваться для задачи классификации текстов.

IV. ЗАКЛЮЧЕНИЕ

На сегодняшний день применение сверточных нейронных сетей не ограничивается анализом изображений. Их применяют для решения различных типов задач: анализа тональности, классификации текста, распознавания речи и многое другое. Благодаря применению сверточных нейронных сетей можно значительно сократить количество обучаемых параметров и получать высокое качество классификации.

СПИСОК ЛИТЕРАТУРЫ

1. Нестеренков, С. Н. Модифицированный генетический алгоритм для обучения нейронной сети / С. Н. Нестеренков, К. П. Белов // Информационные технологии и системы 2017 (ИТС 2017) : материалы междунар. науч. конф., Минск, 25 окт. 2017 г. / Белорус. гос. ун-т информатики и радиоэлектроники ; редкол.: Л. Ю. Шилин [и др.]. - Минск, 2017. - С. 204-205.
2. Нестеренков, С.Н. Применение больших данных в электронном образовании / С.Н. Нестеренков, М.И. Макаров, Н.В. Ющенко, А.Д. Радкевич // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня : сб. материалов V Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 13-14 марта 2019 года). В 2 ч. Ч. 2 / редкол. : В. А. Богущ [и др.]. - Минск : ВГУИР, 2019. - С. 242-245.
3. Нестеренков, С.Н. Использование генетического алгоритма для нахождения коэффициентов и структуры нейронной сети / С.Н. Нестеренков, К.П. Белов // Информационные технологии и системы 2018 (ИТС 2018) : материалы междунар. науч. конф., Минск, 25 окт. 2018 г. / Белорус. гос. ун-т информатики и радиоэлектроники ; редкол.: Л. Ю. Шилин [и др.]. - Минск, 2018. - С. 124-125.
4. Yanyan W., Qun C., Jiquan S., Boyi H., Murtadha A, Zhanhuai Li G. Machine Learning for Aspect-level Sentiment Analysis // arXiv:1906.02502 – 2019.