

Практическая работа №3. Рекуррентные модели для обработки последовательностей.

В работе необходимо провести экспериментальное сравнение следующих моделей для задачи обработки последовательностей (текстов):

1. Модель MLP с обучаемым слоем Embedding
2. Модель MLP с предобученным слоем Embedding (word2Vec, glove или fastText в зависимости от варианта).
3. Рекуррентная модель со слоем SimpleRnn
4. Рекуррентная модель со слоем LSTM\GRU
5. Модель по варианту: сверточная модель, двунаправленная рекуррентная сеть, два рекуррентных слоя.

В экспериментальном исследовании оценить влияние следующих факторов: число узлов в нейронной сети; длина вектора слова; длина последовательности (количество слов в сообщении).

В п.3-5 использовать вариант Embedding-слоя, который оказался наиболее эффективным в п.1-2.

Провести анализ сходства слов в векторном пространстве. Для этого отобрать слова, специфичные для каждого класса (не менее 5 слов на класс), и построить матрицу корреляций по векторным представлениям этих слов, используя Embedding-слои (предобученный и обучаемый).

На курсе необходимо выбрать набор данных (можно предложить свой вариант кроме IMDB). В рамках одного выбранного набора данных необходимо выбрать вариант для пунктов №2 и №5.

Вариант	Предобученная модель Embedding для задания №2	Модель для задания №5
1	Word2Vec	CNN
2	Glove	CNN
3	fastText	CNN
4	Word2Vec	Bidirectional
5	Glove	Bidirectional
6	fastText	Bidirectional
7	Word2Vec	Два рекуррентных слоя
8	Glove	Два рекуррентных слоя
9	fastText	Два рекуррентных слоя

Наборы данных

	Набор данных	Ссылка на исходные данные
1.	Twitter Sentiment Analysis	https://www.kaggle.com/datasets/jp797498e/twitter-entity-sentiment-analysis
2.	Sentiment Analysis on Financial Tweets	https://www.kaggle.com/datasets/vivekrathi055/sentiment-analysis-on-financial-tweets
3.	News Headlines Dataset For Sarcasm Detection	https://www.kaggle.com/datasets/rmisra/news-headlines-dataset-for-sarcasm-detection
4.	Fake News Classification	https://www.kaggle.com/datasets/saurabhshahane/fake-news-classification
5.	Twitter Financial News	https://www.kaggle.com/datasets/sulphatet/twitter-financial-news