

**РЕКУРРЕНТНЫЕ НЕЙРОННЫЕ СЕТИ С
МЕХАНИЗМОМ ВНИМАНИЯ ДЛЯ АНАЛИЗА
ТОНАЛЬНОСТИ РУССКИХ ТЕКСТОВ**

**Иванов Илья Сергеевич¹
Ботвиновский Евгений Александрович²**

1: студент, факультет ФИВТ МФТИ (ГУ), Долгопрудный, Россия

2: к.ф.-м.н., DeepHackLab, Долгопрудный, Россия

E-mail: ilya.ivanov.se@gmail.com, botvinovski@gmail.com

Анализ сообщений в социальных медиа ресурсах представляет огромный практический интерес со стороны бизнеса, поскольку мнение пользователей влияет на покупательскую способность. В частности, одной из актуальных и практически важных задач для бизнеса является анализ тональности [1].

Для тестирования алгоритмов определения тональности существуют специальные конкурсы в рамках ежегодных соревнований (например, SemEval, конференция «Диалог»), а также открытые наборы данных [1,2,3].

В данной работе рассматривается задача классификации русских текстов по тональности. В качестве классификаторов используются такие модели как двунаправленная рекуррентная нейронная сеть и двунаправленная рекуррентная нейронная сеть с механизмом внимания [3,4]. Целью является сравнение данных моделей. Рассматриваемые в работе классификаторы экспериментально проверялись на наборе русских сообщений из Twitter, представленных на соревновании Dialogue Evaluate 2016.

Предварительно текст обрабатывается при помощи Python библиотеки `py morphology2`, позволяющей проводить лемматизацию слов. Для получения векторных представлений слов используется алгоритм Word2Vec. Далее последовательность векторов, кодирующая одно сообщение, подается на вход рекуррентной сети.

В Таблице 1 представлены результаты 5-фолд кросс-валидации различных моделей на обучающей выборке и результаты на тестовой выборке. Используемая метрика - макро-усреднённая F1-мера по классам положительной и отрицательной тональностей. Помимо наших экспериментов в таблице представлены результаты победителя [5] и бейзлайны соревновательной дорожки по анализу тональности Dialogue Evaluate 2016. Здесь стоит отметить, что решения участников соревнования содержали различные дополнительные методы

Таблица 1: F1-мера различных моделей на кросс-валидации (CV) и на тестовой выборке

	Банки		Телекоммуникационные компании	
	CV(mean, std)	test	CV(mean, std)	test
Bi-GRU	0.74, 0.02	0.48	0.62, 0.01	0.52
Bi-GRU+Attention	0.74, 0.02	0.51	0.60, 0.02	0.49
2-layer GRU, reversed sequences (Arhipenko)	0.62, -	0.55	0.66, -	0.56
CNN (Arhipenko)	-	0.48	-	0.47
SVM baseline	-	0.46	-	0.46

обработки данных, помимо лемматизации, которые мы не использовали.

Видно, что лишь на одном из двух доменов алгоритм с исследуемым механизмом внимания превзошёл аналогичный алгоритм без механизма внимания. Весь код решения и значения гиперпараметров доступны по ссылке www.github.com/ilivans/tf-rnn-attention.

Также из таблицы видно, что значения F1-меры на обучающей и тестовой выборке существенно отличаются. Мы провели ряд экспериментов, для того чтобы найти гиперпараметры, при которых бы отсутствовало переобучение алгоритмов. Однако эта разница наблюдалась во всех наших экспериментах, как при уменьшении размера сети, так и с увеличением параметра dropout [6]. Чтобы исследовать причины этого расхождения, мы провели эксперимент со смешиванием обучающей и тестовой выборок и последующей кросс-валидацией моделей на смешанной выборке. Результаты данного эксперимента приведены в Таблице 2. Судя по тому, что кросс-валидация на смешанной выборке показала результаты очень близкие к кросс-валидации на обучающей выборке, можно предположить, что между обучающей и тестовой выборками есть существенные различия. Однако для тщательной проверки этой гипотезы требуется провести детальный сравнительный анализ данных, что авторы планируют сделать в будущем.

Таким образом, исследована применимость модели на основе двунаправленной рекуррентной нейронной сети с механизмом внимания в задаче классификации тональности русскоязычных текстов. Про-

Таблица 2: Результаты эксперимента со смешиванием обучающей и тестовой выборки

	Банки		Телекоммуникационные компании	
	CV(mean,std)		CV(mean,std)	
	train	train+test	train	train+test
Bi-GRU	0.74, 0.02	0.71, 0.02	0.62, 0.01	0.62, 0.01
Bi-GRU+Attention	0.74, 0.02	0.72, 0.01	0.60, 0.02	0.62, 0.01

ведено сравнение данной модели с её ранее изученными аналогами. В будущем планируется провести эксперименты на других, более сбалансированных и крупных, наборах данных.

Литература

1. Рубцова Ю. Построение корпуса текстов для настройки тонового классификатора // Программные продукты и системы, 2015, С. 72–78.
2. Maas A., Daly R., Pham P., Huang D., Ng A., Potts C. Learning Word Vectors for Sentiment Analysis // Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Stroudsburg, USA, 2011, Vol.1, P. 142–150.
3. Yang Z., Yang D., Dyer C., He X., Smola A., Hovy E. Hierarchical attention networks for document classification // Proceedings of the 15th Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, USA, 2016, P. 1480–1489.
4. Bahdanau D., Cho K., Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate // Computing Research Repository, 2014, Vol. abs/1409.0473
5. Arkhipenko K., Kozlov I., Trofimovich J., Skorniakov K., Gomzin A., Turdakov D. Comparison of Neural Network Architectures for Sentiment Analysis of Russian Tweets // Proceedings of the International Conference «Dialogue 2016», 2016, Vol.15, P. 50–58.
6. Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting // The Journal of Machine Learning Research, 2014, Vol. 15, № 1, P. 1929–1958.