

# Рекуррентные нейронные сети с механизмом внимания для анализа тональности русских текстов

Иванов Илья Сергеевич  
студент  
ФИВТ МФТИ

## 1 Введение

С появлением в Интернете социальных медиа ресурсов, таких как социальные сети, форумы, блоги и др. наблюдается рост как количества пользователей этих ресурсов, так и объема данных, создаваемых пользователями. Люди активно делятся своим мнением в комментариях, обзорах, отзывах, обсуждениях. Эта активность представляет огромный практический интерес со стороны бизнеса, поскольку влияет на покупательскую способность [wang]. В частности, одной из актуальных и практически важных задач для бизнеса является анализ тональности [mokoron2012]. Интерес к данной области стал активно проявляться в последние годы. Существует несколько ежегодных соревнований, такие как, например, соревнование по созданию систем автоматического анализа тональности на английском языке SemEval [semEval2010], которое проходит с 2010 года.

Для тестирования алгоритмов определения тональности существует также несколько открытых наборов данных. Для английского языка, например, это обзоры фильмов <sup>1</sup> [imdb, socher-rdm] и товаров <sup>2</sup> [amazon]. Наилучшие алгоритмы в данной области, основаны на рекурсивных тензорных нейронных сетях [socher-rdm].

Для русского языка в рамках соревнований Семинара РОМИП в 2011 году [romip] был сформирован корпус текстов, содержащий отзывы о книгах, фильмах и товарах. В 2014 году Рубцова Ю.В. создала корпус коротких сообщений Twitter [mokoron2015]. На соревновательной дорожке по анализу тональности Dialogue Evaluate 2016 [senti-ru-eval] был представлен корпус сообщений Twitter, посвященный банкам и телекоммуникационным компаниям. Победителем последнего соревнования в 2016 году стало решение, основанное на рекуррентной нейронной сети [arhipenko].

В данной работе рассматривается задача классификации русских текстов по тональности. В качестве классификаторов используются такие модели как двунаправленная рекуррентная нейронная сеть [schuster] и двунаправленная

---

<sup>1</sup><http://nlp.stanford.edu/sentiment/treebank.html>

<sup>2</sup><http://jmcauley.ucsd.edu/data/amazon/>

рекуррентная нейронная сеть с механизмом внимания [bahdanau]. Целью является сравнение данных моделей. Ранее для анализа тональности русского текста не применялись модели с механизмом внимания, хотя они успели показать впечатляющие результаты в задачах классификации англоязычных текстов [yang-att-2016]. Представленные в работе классификаторы экспериментально проверяются на указанном выше наборе русских сообщений из Twitter, представленных на Dialogue Evaluate 2016.

## 2 Задача

Дана обучающая выборка из русскоязычных сообщений с метками трёх классов: -1 (негативная тональность), 0 (нейтральное сообщение), 1 (позитивная тональность). Используя данную выборку, требуется построить классификатор, наиболее точно распознающий тональность сообщений. Целевой метрикой является макро-усреднённая F1-мера по классам позитивной и негативной тональностей.

## 3 Алгоритм

Предварительно текст обрабатывается при помощи Python библиотеки `rumorphy2`, позволяющей проводить лемматизацию слов. Для получения векторных представлений слов используется алгоритм Word2Vec. Далее последовательность векторов, кодирующая одно сообщение, подается на вход рекуррентной сети.

### 3.1 Рекуррентная нейронная сеть

Ключевой компонентой алгоритма решения поставленной задачи является однослойная двунаправленная рекуррентная нейронная сеть [schuster] типа GRU [cho]. GRU (Gated Recurrent Unit) может быть описан следующими уравнениями:

$$z_t = \sigma_g(W_z x_t + U_z h_{t-1}) \quad (1)$$

$$r_t = \sigma_g(W_r x_t + U_r h_{t-1}) \quad (2)$$

$$\tilde{h}_t = \tanh(W x_t + U(r_t \circ h_{t-1})) \quad (3)$$

$$h_t = (1 - z_t) \circ \tilde{h}_t + z_t \circ h_{t-1} \quad (4)$$

где  $x_t$  –  $t$ -ый элемент последовательности, а  $h_t$  – внутреннее состояние сети в  $t$ -ый момент времени (после обработки  $x_t$ ).

Входная последовательность подается на вход одной рекуррентной сети прямым порядком и другой сети - обратным. После чего выходы этих двух слоёв конкатенируются, образуя выходную последовательность  $y_t = [\vec{h}_t, \overleftarrow{h}_t]$ . Таким образом строится двунаправленная рекуррентная сеть, архитектура которой изображена на Рис. ??.

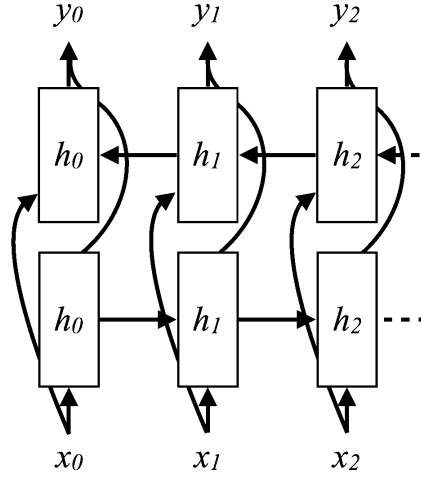


Рис. 1: Двухнаправленная рекуррентная нейронная сеть.

### 3.2 Механизм внимания

Классическим подходом к работе с выходом нейронной сети является рассмотрение лишь последнего вектора  $y_T$ , где  $T$  – длина входной (и выходной) последовательности, т.к. он аккумулирует в себе извлеченную информацию из всей входной последовательности. Рассматриваемый же нами подход – механизм внимания [bahdanau] – утилизирует все выходные векторы  $y_t$ , вычисляя их линейную комбинацию с коэффициентами  $\alpha_t$ , которые обучаются вместе с остальной сетью. Уравнения механизма внимания выглядят следующим образом:

$$v_t = \tanh(W_\omega y_t + b_\omega) \quad (5)$$

$$\alpha_t = \frac{\exp(v_t^T u_\omega)}{\sum_{j=1}^T \exp(v_j^T u_\omega)} \quad (6)$$

$$v = \sum_{t=1}^T \alpha_t y_t \quad (7)$$

Архитектура двухнаправленной рекуррентной нейронной сети с механизмом внимания представлена на Рис. ?? . Во время обучения к выходу слоя механизма внимания  $v$  применяется регуляризатор Dropout [srivastava] для уменьшения переобучения.

### 3.3 Полносвязанный слой

После слоя механизма внимания следует полносвязный слой с функцией активации softmax, переводящий вектор  $v$  в трёхмерное пространство, где каждое из чисел обозначает вероятность принадлежности объекта к соответствующему классу:

$$\hat{s} = (\hat{s}^{(-1)}, \hat{s}^{(0)}, \hat{s}^{(1)}) = \text{softmax}(Wv + b) \quad (8)$$

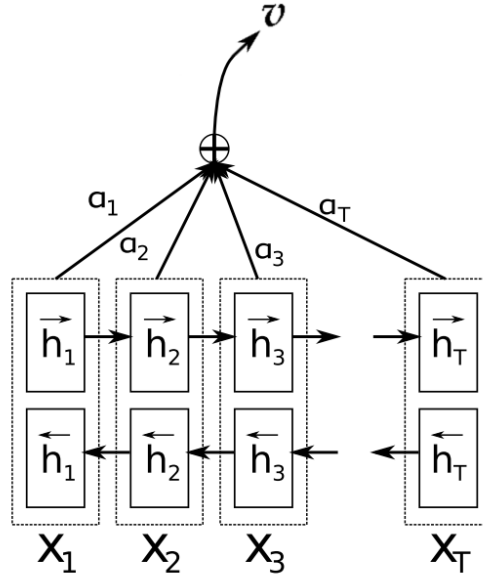


Рис. 2: Двухнаправленная рекуррентная нейронная сеть с механизмом внимания.

## 4 Обучение

Классификатор обучается при помощи алгоритма оптимизации Adam, минимизирующего перекрёстную энтропию между выходным распределением  $\hat{s}$  и истинным  $s$ :

$$L(W) = - \sum_{i=1}^n \sum_{c \in \{-1,0,1\}} s_i^{(c)} \log \hat{s}_i^{(c)}, \quad (9)$$

Параметры модели подбираются при помощи перекрёстной валидации.

## 5 Данные

Организаторами соревнования Dialogue Evaluate 2016 было предложено 2 набора данных. Один содержал сообщения, упоминающие банковские компании, и состоял приблизительно из 12700 экземпляров (9400 в обучающей выборке и 3300 - в тестовой). Второй набор был посвящён телекоммуникационным компаниям и состоял приблизительно из 10800 экземпляров (8600 в обучающей выборке и 2200 - в тестовой).

## 6 Эксперименты

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 [arhipenko] и бейзлайны соревновательной дорожки по анализу тональности Dialogue Evaluate 2016 [senti-ru-eval]. Здесь стоит

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

	Banks		Telecommunication companies	
	5-fold CV (mean, std)	test	5-fold 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
Bi-GRU (Arhipenko)	0.62, -	-	0.65, -	-
LSTM (Arhipenko)	0.60, -	-	0.64, -	-
CNN (Arhipenko)	-	0.48	-	0.47
SVM baseline	-	0.46	-	0.46
Majority baseline	-	0.31	-	0.19

отметить, что решения участников соревнования содержали различные дополнительные методы 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](https://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 регуляризации dropout [srivastava]. 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. Стоит отметить, что размеры обучающей и тестовой выборок сравнимы (5:2). 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. Проведено сравнение данной модели с её ранее изученными аналогами. В будущем планируется провести эксперименты на других, более сбалансированных и крупных, наборах данных.

Таблица 2: Результаты эксперимента со смешиванием обучающей и тестовой выборок

	Banks			Telecommunication companies		
	cross-validation		test	cross-validation		test
	train	train+test		train	train+test	
Bi-GRU	0.74, 0.02	0.71, 0.02	0.48	0.62, 0.01	0.62, 0.01	0.52
Bi-GRU + Attention	0.74, 0.02	0.72, 0.01	0.51	0.60, 0.02	0.62, 0.01	0.49