

СЕКЦИЯ: «КОМПЬЮТЕРНЫЕ НАУКИ»

РАЗРАБОТКА МОДЕЛИ ПРЕДСКАЗАНИЯ ИНТЕНЦИИ СООБЩЕНИЙ СОЦИАЛЬНОЙ СЕТИ

А.В.Демидовский

Нижний Новгород, Россия

Национальный исследовательский университет

«Высшая школа экономики»

Образовательная программа «Бизнес-информатика», курс 6

monadv@yandex.ru

В работе представлен метод анализа диалогов, с помощью глубокого обучения, собранных в социальной сети. Была использована выборка диалогов из социальных сетей, которая была размечена в соответствии с интенциями в каждом тексте. Были размечены направление интенции и тип интенции. Используя эту выборку и заранее натренированные вектора, представляющих слова, авторы оценили работу двух нейронных архитектур. В результате, авторами был произведен выбор модели для автоматического предсказания типа интенции и ее направленности для произвольного текста.

Ключевые слова: обработка естественного языка, интент-анализ, глубокое обучение.

Наблюдается растущий интерес к анализу социальных сетей в силу той роли, которую они играют в жизни современного человека. Фундаментальным трендом является использование социальных сетей в качестве площадки для общения и коллективной работы. Подавляющее большинство людей зарегистрированы хотя бы в одной социальной сети и люди общаются посредством обмена текстовыми сообщениями. Именно поэтому так важно уметь анализировать эти тексты и уметь выделять их важные свойства.

Одной из наиболее популярных форм общения в социальной сети является понятие «пост». Это произвольное сообщение, выражающее в произвольной форме мысли и позицию автора. Такой пост обычно вызывает у аудитории всплеск эмоций и читатели начинают активно обсуждать это сообщение, добавляя в так называемый "диалог" больше сообщений. Зачастую тема обсуждения в таких диалогах быстро меняется и, наконец, к моменту завершения диалога тема перестает быть связан с темой оригинального сообщения.

Почему так важно уметь предсказывать интенцию заданного текста? Идея манипуляции диалогом и его центральным посылом является ключевой в некоторых аспектах политической лингвистики. Действительно, крайне важно убедиться в том, что диалог кандидата в президенты с аудиторией несет правильную мысль и правильно побуждает ее к действию. Принимая во внимание обсуждавшуюся выше тенденцию к быстрой смене темы диалога, важно оставаться уверенным в том, что диалог сохраняет свой изначальный посыл. С точки зрения авторов, первым шагом к созданию таких манипулятивных инструментов является умение автоматически предсказывать интенцию текста.

В то же время, возрастающее число публикаций в сфере анализа Интернет текстов раскрывает некоторые интересные особенности современных текстов. В частности, каждый речевой акт имеет собственную интенцию - намерение или желание выразить идею. Кроме того, интенция характеризуется своей направленностью, что означает, что любая фраза пользователя социальной сети может быть направлена на автора поста, на самого пользователя и т.д. Более того, в связи с большим количеством таких текстов, достаточно легко собрать большую выборку текстов, что критично для точной работы современных инструментов анализа текстов.

В то время как дебаты об интент-анализе набирают все большую популярность, по-прежнему существует пробел, который заключается в отсутствии математических моде-

лей и прикладных инструментов для автоматического предсказания интенций. Несмотря на существующие работы, посвященные разработке текстовых классификаторов, задача автоматического предсказания интенций остается нерешенной.

Работа имеет следующую структуру. В Разделе 2 будет дано детальный обзор существующих подходов и важнейших теоретических трендов в анализе текстов. Затем, в Разделе 3 будет описана экспериментальная методология, которая была разработана авторами для проведения экспериментов. Детальное описание нейронных архитектуры будет дано в Разделе 4. Результаты предлагаемого подхода будут представлены в Разделе 5. Наконец, в Разделе 6 будет произведен анализ полученных результатов и будут сформулированы направления дальнейшего развития.

Работа выполнена на стыке двух предметных областей. С одной стороны, это психология коммуникации. С другой стороны, современные методы анализа естественного языка. Начнем с психологии.

Самые ранние работы в этом направлении датируются 60 годами XX века, работами исследователей Austin and J. Searle [1,10], которые разработали теорию речевых актов. J. Austin разделяет речевые акты на три составляющих - локутивный, иллокутивный, перлокутивный. Наше исследование связано только с иллокутивными актами в направлении политического дискурса. Как и многие другие исследователи [14], мы используем конечный набор интенционных типов. Олешков М.Ю. предложил типизировать не речевые акты, а коммуникативную стратегию как основную цель речевого акта [8]. Мы используем это конечное множество типов интенций, предложенных Радиной Н.К [2].

Задача, которая заключается в автоматическом определении типа коммуникативной стратегии из заданного конечного набора состояний, может быть сформулирована как классификационная проблема. Проблема классификации коротких текстов - это хорошо известная проблема в области исследований естественных языков. Например, предложения могут быть классифицированы согласно их удобочитаемости с использованием заранее выбранных признаков и SVM, деревьев решений и т.д. [4]. Еще короткие сообщения из социальных сетей часто классифицируются по типу эмоций [5].

Прорывные результаты последних лет, связанные с применением нейронных сетей, показывают, что передача знания (transfer learning) позволяет улучшать классические методы машинного обучения [12,3]. Это особенно критично при условии, что тренировочная выборка достаточно небольшая. Именно поэтому в работе было решено использовать нейронные сети с пред обученным слоем в качестве классификатора.

Одним из ключевых направлений нашего исследования была подготовка подходящего набора данных. Было решено использовать самую популярную русскую социальную сеть - ВКонтакте¹ - в качестве источника диалогов и текстов. Однако, возникла необходимость в тщательной фильтрации полученных диалогов. В частности, были отброшены все фразы короче 40 символов, а также бесполезные с точки зрения анализа кода комментарии - картинки, пустые сообщения и т.д.

Крайне важно разметить каждый текст в соответствии с классом интенции. Более того, крайне важно также определить направленность интенции произвольного текста. Разметка данных производилась экспертами, которые использовали буквы для обозначения типа интенции и цифры для обозначения ее направленности. Такая разметка позволяет сконцентрировать внимание не на традиционных методологиях интен-анализа, но и на смешанных, которые позволяют классифицировать текст достаточно точно с интенционной точки зрения.

Интеграция полученной выборки в построение автоматического интенционного классификатора будет подробно описана в следующем Разделе.

Для автоматического предсказания интенций из заданного множества интенций нам необходимо создать классификатор пользовательских сообщений. Олешко М.Ю. предложил использовать классификацию, включающую 25 типов интенций [8]. Радина Н.К. предложила группировать эти типы в группы по 5 штук [2] согласно идеям Хабер-

маса, данная классификация представлена в Таблице 2. Кроме того, мы добавляем 4 типа направленности интенции, показанные в Таблице 1. Таким образом, у нас есть следующие группы классов:

- 25x4 = 100 типов интенций + направленности;
- 25 типов интенции;
- 5 типов агрегированных интенций;
- 4 направления интенций;

В качестве классификатора были использованы две ключевые нейросетевые архитектуры. Для оценки работы классификаторы были использованы следующие метрики как: точность, полнота и т.д.

Таблица 1.

Направленность интенций		
№	Трактовка	Пример
1	"Я"/"Мы комментатор о себе и о "нас"	я не согласен, я считаю мы вместе можем и т.п.
2	"Они» автор/ы статьи	они правы, они не правы, они заблуждаются и т.п
3	"Они другие комментаторы	они не понимают, они говорят глупость и т.п.
4	кто-то вообще, не "Я"/"Мы не "Они"	Россия, Америка, капиталисты и т.п.

Идея представления каждого слова в тексте или целого текста в виде вектора является ключевой в современных методологиях анализа текста. Такие векторы представляют слова и могут быть натренированы с помощью алгоритмов word2vec, Glove и т.д. [7,9]. Эти вектора могут быть натренированы на любой выборке, хотя данная выборка должна быть достаточно большой, например, объем всей российской Википедии достаточен для получения осмысленных результатов. Выборка, которую мы используем, содержит 21192 текст и 100 классов, что означает с точки зрения классификатора недостаток данных для тренировки векторов слов с нуля. Поэтому было решено уже готовые векторы² 4, натренированные на Национальном корпусе русского языка (НКРЯ) [6] и русской Википедии.

Однако, изучения данных векторов выявило необходимость в предварительной обработке каждого слова из исходной выборки в вид, в котором это слово бы имело соответствие в выбранном списке векторов. В частности, каждое слово должно быть приведено к форме "нормальная форма"+ "часть речи, например, "бежал" должно превратиться в "бегать_Verb". Такая обработка требует наличия автоматического инструмента для проведения морфологического анализа для каждого слова в заданном тексте. В рамках данного исследования был выбран инструмент MyStem³, разрабатываемый компанией Яндекс. Данный инструмент позволяет создавать необходимые словоформы и переиспользовать для них уже натренированные векторы.

Слои свертки. Свёрточные нейронные сети - наиболее продвинутые модели для решения задачи классификации текста [13]. Используются трехслойные объединенные композитные слои с maxpooling.

Рекуррентные слои. Рекуррентные слои оказываются удобными для хранения длинных последовательностей [13]. Используются ячейки longshort-termmemory (LSTM).

Регуляризация. В качестве регуляризатора используется dropout для того, чтобы модель не переобучилось [11]. Наш слой регуляризатора случайным образом выбирает половину скрытых нейронов и выставляет их выходы как нули, что предотвращает объединение свойств.

Обучение. Мы используем слой векторов из списка этих векторов из НКРЯ [6] и русской Википедии. Другие слои нашей сети были инициализированы случайным образом. Затем происходила тренировка сети с использованием Adam подхода для стохастиче-

ской оптимизации функции цели.

Далее будет произведена демонстрация результатов (см. в Табл. 3), а также будет произведен анализ полученных результатов. Полученные результаты демонстрируют способностей сверточных 1b и рекуррентных 1a моделей решать классификационную задачу на заданной выборке в зависимости от количества эпох тренировки. Во-первых, обе модели были использованы для предсказания классов интенции заданного текста, и полученные результаты достаточно низки для обеих моделей (точность меньше чем 0.05). Это, в свою очередь, доказывает гипотезу, что использованная выборка слишком мала (21192 текстов) для такого большого количества классов (100 классов). Принимая это во внимание, было решено проводить дальнейшие эксперименты на меньшем числе классов, и авторами были предложены три стратегии для решения проблемы низкой точности.

Первая стратегия заключалась с попытки предсказать только тип интенции. Вместо попыток предсказывать 100 классов, мы можем предсказывать только 25, основываясь на общей классификации типов интенций - Таблица 2. Этот эксперимент позволил получить немного лучшие результаты, хотя размер точности по-прежнему оставался меньше 0.1. В качестве второй стратегии было решено использовать предсказание только направления интенции. В таком случае, модели приходилось предсказывать один из четырех классов направленности интенций - Таблица 1. Валидация второй стратегии показала, что основной причиной низкой точности модели на предыдущих шагах выступало различие в соответственном размере количества классов и размера выборки. В качестве финальной стратегии производилась попытка предсказывать тип интенции на основе ее супер-интенции. В самом деле, Nabernas (см. Табл. 2) позволяет объединять типы интенций в группы по 5. Результаты этого эксперимента показали, что при попытке предсказать 5 классов, обе модели могут корректно распознавать интенцию каждого третьего текста.

Анализируя итоговый результат, можно понять, что рекуррентные сети обгоняют сверточные сети почти во всех сценариях, хотя и требует большего времени для тренировки модели на большом количестве эпох. Это позволяет сделать вывод о подходемости рекуррентных сетей для решения классификационных задач текстов, а также дороговизны с вычислительной точки зрения.

Наконец, сложно переоценить важность первой в мире выборки в русском языке, которая была размечена согласно типу и направленности интенции. Выборка содержит 21992 реплик, представляющих 100 классов. Кроме того, натренированные модели и исходный код доступны публично для пользования и улучшения⁴ под "мягкой" для бизнеса лицензией.

Таблица 3.

Экспериментальные результаты на сверточных (CNN) и рекуррентных сетях (LSTM) в зависимости от количества эпох

	Супертип интенции		Направление интенции		Тип интенции		Тип интенции + направленности	
количество классов	5		4		25		100	
	LSTM	CNN	LSTM	CNN	LSTM	CNN	LSTM	CNN
количество эпох	5	20	5	20	5	5	100	5
precision	0.35	0.06	0.63	0.36	0.07	0.01	0.06	0.01
recall	0.35	0.24	0.65	0.60	0.15	0.10	0.05	0.07
f1-score	0.34	0.09	0.63	0.45	0.08	0.02	0.05	0.01

Цель данного исследования заключалась в построении модели, подходящей для предсказания интенций текста пользователя, выраженного в виде диалогов из социальных сетей.

Была подготовлена выборка с помощью сбора диалогов из социальных сетей и

разметки диалога. Класс содержит в себе два элемента: тип интенции и ее направление. Все диалоги были связаны с политическими темами. Использование данной выборки и заранее натренированных векторов слов были построены две модели на основе нейронных сетей, которые оказались способны предсказывать интенцию произвольного текста любого автора. Классификация по направленности интенции показала наивысшую точность. Мы объясняем это не только низким количеством классов, но и тем, что направления часто задаются внешними словами.

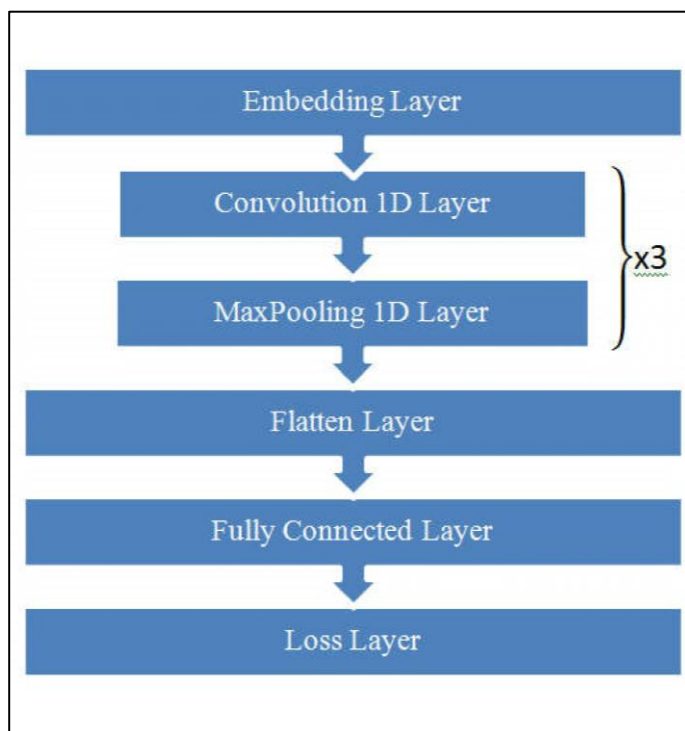


Рис. 1.а: Структура сети CNN, использованной в экспериментах.

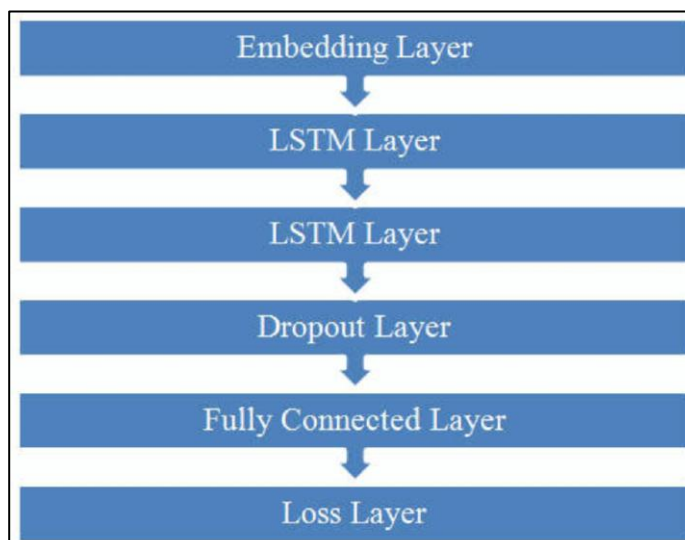


Рис. 1.б: Структура сети LSTM, использованной в экспериментах.

Информативно-воспроизводящий тип, Воспроизвести в речи наблюдаемое		Эмотивно-консолидирующий тип, Предложение собственной картины мира для кооперативного взаимодействия		Манипулятивный тип, Социальное доминирование, установление иерархии		Волюнтивно-директивный тип, Побудить адресата к действию, внести изменения в фрагмент действительности		Контрольно-реактивный тип, Выразить оценочную реакцию на ситуацию	
А	Удивление Вопрос	Е	Самопрезентация	Л	Оскорбления	Р	Побуждение к позитивному действию/рекомендации	Х	Одобрение Похвала
Б	Выражение несогласия Сомнение	Ж	Привлечение внимания (рассуждения, риторические вопросы)	М	Устрашение, угрозы	С	Подстрекательство к негативному	Ц	Сарказм Зло-радство
В	Выражение согласия Поддержка	З	Успокоение аудитории Утешение	Н	Дискредитация (под-рыв авторитета)	Т	Обвинение	Ч	Критика
Г	Непринятие, отказ от коммуникации	И	Прогнозы, претензии на истину	О	Демонстрация силы (без явной угрозы)	У	Предупреждение о последствиях	Ш	Ирония
Д	Сочувствие, симпатия	К	Оправдание (как самооправдание, т.е. без обвинений)	П	Морализаторство, поучения	Ф	Отвод обвинений (если обвиняют)	Щ	Разоблачение

Таблица 2: Классификация интенций по Хабермасу.

Благодарности

Статья подготовлена в результате проведения исследования (№ 17-05-0007) в рамках Программы «Научный фонд Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ)» в 2017 г. и в рамках государственной поддержки ведущих университетов Российской Федерации "5-100".

Список использованной литературы:

1. John Langshaw Austin. How to do things with words. Oxford university press, 1975.
2. Radina Nadezhda K. Intent alysis of online discussions (using examples from the internet portal inosmi.ru). Mediascope, (4), 2016.
3. Nikolay Karpov. Nru-hse at semeval-2017 task 4: Tweet quantification using deep learning architecture. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), pages 681–686, Vancouver, Canada, August 2017. Association for Computational Linguistics.
4. Nikolay Karpov, Julia Baranova, and Fedor Vitugin. Single-sentence readability prediction in russian. In International Conference on Analysis of Images, Social Networks and Texts_x000D_, pages 91–100. Springer, 2014.
5. Nikolay Karpov, Alexander Porshnev, and Kirill Rudakov. NRU-HSE at SemEval- 2016 Task 4: Comparative Analysis of Two Iterative Methods Using Quantification Library. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), pages 171–177, San Diego, California, June 2016. Association for Computational Linguistics.
6. Andrey Kutuzov and Elizaveta Kuzmenko. WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models, pages 155–161. Springer International Publishing, Cham, 2017.
7. Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems, pages 3111–3119, 2013.
8. M. Yu. Oleshkov. Simulation of the communication process: monograph. Nizhny Tagil gos.sots.-ped.akademiya (et al.), 2006.
9. Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In EMNLP, volume 14, pages 1532–1543, 2014.
10. John R Searle. Speech acts: An essay in the philosophy of language, volume 626. Cambridge university press, 1969.
11. Nitish Srivastava. Improving neural networks with dropout. PhD thesis, University of Toronto, 2013.
12. Dario Stojanovski, Gjorgji Strezoski, Gjorgji Madjarov, and Ivica Dimitrovski. Finki at semeval-2016 task 4: Deep learning architecture for twitter sentiment analysis. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), pages 149–154, San Diego, California, June 2016. Association for Computational Linguistics.
13. Duyu Tang, Bing Qin, and Ting Liu. Document modeling with gated recurrent neural network for sentiment classification. In EMNLP, pages 1422–1432, 2015.
14. Latynov V.V., Сепцов V.A., and Alexeev K.I. Words in action: Intent-analysis of political discourse. 2000.