УДК 004.624

Опыт применения методов искусственного интеллекта для выявления пользователей социальной сети, готовых к совершению противоправных действий

В. О. Старкова, Е. Ю. Никитина
Пермский государственный национальный исследовательский университет
Россия, 614990, г. Пермь, ул. Букирева, 15
kochergina.vo@mail.ru; +7-922-309-51-24

Рассмотрены различные способы анализа данных, представленных текстами на естественном языке. Осуществлена попытка построить программный классификатор для таких текстов методами искусственного интеллекта. Проведено сравнение качества классификации на большой выборке разнообразных текстов и выборке более узкой направленности. Сделаны выводы о модели представления данных для решения задач классификации.

Ключевые слова: *обработка естественного языка; искусственный интеллект; анализ данных; модель данных; онтология; выборка.*

DOI: 10.17072/1993-0550-2020-1-80-86

Введение

Целью проводимой исследовательской работы является попытка построения и последующей реализации в составе программного комплекса математической модели определения готовности человека к совершению противоправных действий на основе опубликованных им постов, репостов и комментариев в социальной сети.

Понятие готовности человека может рассматриваться с разных точек зрения:

- психологии,
- криминалистики и оперативно-розыскной службы,
- юриспруденции,
- других источников.

Чтобы с высокой точностью определить готовность человека к совершению того или иного действия, нужно иметь четкое представление о том, по каким параметрам будет оцениваться такая готовность.

Задача усложняется тем, что мы анализируем текст, который пишет пользователь, а не его личные данные: пол, возраст, состав семьи и т.д. Мы должны выявить его намерения только по тому, что он пишет или репостит.

Разработка подобной описательной модели на основе перечисленных выше областей знания является темой отдельной статьи, как и выявление четких параметров психологического состояния личности, отраженного в его постах, репостах и комментариях, предполагает глубокую совместную работу таких структур как органы внутренних дел, юристы, криминалисты, психологи, эксперты по различным видам преступлений и др.

В данной статье мы ограничились представлением данных только по одному виду противоправных действий – так называемому "school shooting" ("готовность к совершению массовых расстрелов и подрывов"). В качестве способа определения вышеуказанной угрозы на данном этапе мы рассмотрели только проявленный в постах, репостах и комментариях интерес к использованию различных видов оружия и взрывчатых веществ.

Целью выполненной промежуточной работы является анализ подобного рода данных и выбор средств для его реализации.

1. Отбор данных для анализа

Часть задач по отбору данных была осуществлена компанией "Сеуслаб". Была разработана поисковая система, включающая в себя инструменты поиска постов, репостов и комментариев социальной сети "ВКонтакте" по заданному слову или набору слов.

Кроме того, эксперты компании составили "словарь" по заданной тематике – набор слов, имеющих прямое отношение к названию или составу оружия и взрывчатки. По данному словарю поисковой системой были отобраны тексты из сети "ВКонтакте".

В результате проведенного отбора данных было обнаружено, что в выборку попадают данные, которые можно рассматривать как потенциально опасные, но также (и их большинство) – тексты, имеющие отношение совсем к иным сферам. Так тексты, в которых встречается слово "селитра", могут иметь отношение и к изготовлению взрывчатых веществ, и к области производства удобрений.

Таким образом, в обязательном порядке возникает задача проведения классификации отобранных данных.

2. Применение классификаторов

Классификация отобранных данных может быть осуществлена различными способами и различными программными средствами. Подбор таких способов и средств и является результатом проделанной нами работы.

Одним из способов классификации может быть последовательная программная проверка каждого текста на наличие других слов, способных выявить контекст сообщения.

В вышеописанном примере с "селитрой" это могут быть слова "огород", "удобрение", "подкормка" и т.д. – для случая нерелевантных текстов или слова "уголь", "сера", "взрывчатка" и прочее – для случая текстов о взрывчатых веществах. Однако такой способ имеет ряд очевидных недостатков: невозможно учесть все слова, имеющие отношение к нужному контексту, тем более, если для отбора используется не одно ключевое слово (как в примере – "селитра"), а целый словарь слов.

3. Применение онтологий

Второй способ, найденный при анализе работ в области NLP (**Natural Language Processing**, обработка **естественного языка**), – это популярное сейчас направление онтологий.

Применение данного способа для представления текстов на естественном языке подразумевает кодирование слов таким образом, чтобы учитывался контекст – семантическая близость между словами.

На сегодняшний день имеется множество способов такого кодирования. Эти способы реализованы различными программными средствами, встроенными в конкретные языки программирования, либо представляющие собой самостоятельные словари и библиотеки.

3.1. Реализация онтологического кодирования (векторизации) на выборке по полному словарю

Представление рассматриваемой выборки по оружию и взрывчатке различными способами кодирования представлено на рис.1, 2 и 3. Желтым цветом на рисунке отмечены посты, которые не являются релевантными для исследуемого примера, а синим – те, которые несут потенциальную угрозу с точки зрения экспертов.

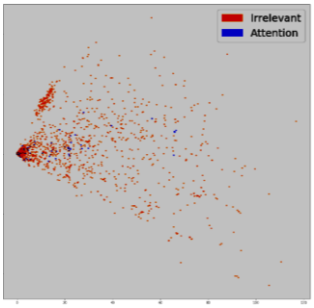

Рис. 1. *Векторизация методом "Bag of words"*

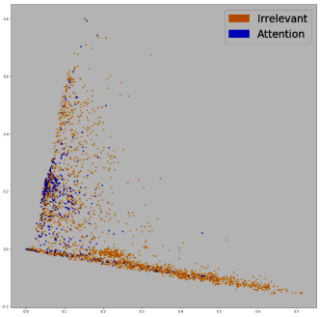

Рис. 2. *Векторизация методом TF-IDF*

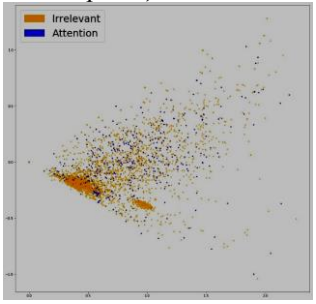

Рис. 3. *Векторизация методом Word2vec*

Как видно на графиках, разделение такой многообразной выборки на два класса посредством определения их контекста с помощью онтологических средств не выглядит перспективным.

3.2. Реализация онтологического кодирования (векторизации) на выборке по одному слову

В качестве эксперимента было принято решение сузить выборку и посмотреть, как на подобных онтологиях будет выглядеть более сбалансированное множество. Были отобраны тексты не по всему словарю "оружие и взрывчатка", а только по слову "селитра". Данная выборка точно так же была размечена экспертами на релевантное и нерелевантное множества классов.

Кроме того, при повторной разметке были учтены некоторые недочеты, выявленные на предыдущем этапе, что позволило сделать новую разметку выборки более сбалансированной. Результаты проведенной классификации по одному ключевому слову приведены на рис. 4.

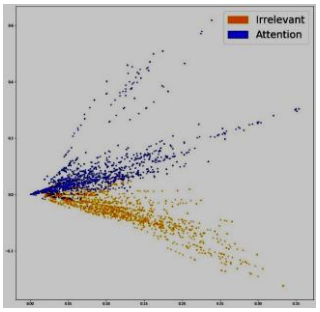

Рис. 4. *Векторизация выборки по слову "селитра". Метод TF-IDF*

На данном рисунке видно, насколько хорошо может работать онтологическое кодирование при задаче классификации.

3.3. Промежуточные выводы

По результатам проведенных экспериментов были сделаны следующие выводы:

1. Результат и качество классификации текстов напрямую зависит от качества подбора и сужения выборки данных.
2. Онтологический способ может быть использован как средство классификации. Возможность его применения будет зависеть от критериев отбора данных для конечной выборки.

Для разрабатываемой модели описания портрета потенциального преступника впоследствии будет рассмотрена возможность применения данного способа.

Необходимо отметить, что сейчас с большой скоростью развиваются технологии представления текстовых данных. В частности, такой программный комплекс как BERT компании Google обладает уже рядом дополнительных возможностей. И в случае необходимости на конечной модели данных планируется прибегнуть к этой технологии.

4. Применение нейронных сетей

Еще одним способом классификации является машинное обучение. Популярное сейчас направление искусственного интеллекта широко применяется для работы с текстовыми данными.

4.1. Выбор параметров нейронной сети

При использовании текстов на естественном языке в качестве входных данных нейронной сети необходимо представить тексты в числовом виде. Для такой кодировки были использованы все рассмотренные выше способы векторизации. В качестве выходного слоя был взят один "нейрон", который принимает значения 0 либо 1. Таким образом, получилось два массива данных:

1. Массив входных параметров в виде списка из 4322 примеров, где текст каждого примера был закодирован набором числовых значений.
2. Массив выходных параметров такой же размерности, представляющий собой набор 0 и 1, где 0 – означает, что данный текст нерелевантен, то есть не представляет угрозы, а 1 – текст потенциально опасен и может нести угрозу.

Далее оба массива были разбиты на:

- обучающее множество, состоящее из 3042 примеров, использованное для обучения сети,
- тестируемое множество, состоящее из 1128 примеров, используемое для проверки прогностических свойств,
- валидирующее множество из 152 примеров для предохранения сети от проблемы приспособления на этапе тестирования.

4.2. Проектирование нейронной сети

На следующем этапе было выполнено проектирование нейронной сети, ее обучение и тестирование, а также эксперименты над нейросетевой моделью. Оптимальная струк-

тура нейронной сети была подобрана методом перебора различных архитектур. Наиболее удачной проявила себя архитектура "Yoon Kim", разработанная японским экспертом по работе с NLP и названная в его честь (рис. 5).

Данная архитектура состоит из слоя плотных векторов (Embedding), трех параллельных сверточных слоев, двух слоев Dropout (для борьбы с ситуацией переобучения), одного скрытого слоя на 64 нейрона с функцией активации "relu", слоя flatten, который переводит многомерные вектора сверточного слоя и слоя embedding в плоский вектор. И выходного слоя с одним нейроном. На выходном слое использовалась сигмоидная функция активации.

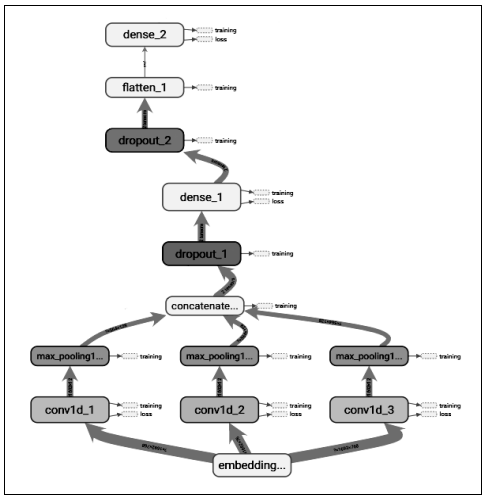

Рис. 5. *Нейросетевая архитектура Yoon Kim*

После оптимизации и обучения нейронной сети ее прогностические свойства проверялись на примерах тестирующего множества, которое в процессе обучения нейросети не участвовало. Подобранная архитектура показала хорошую сходимость: точность на обучающей выборке 99 %, на тестовой 91 % (рис. 6).

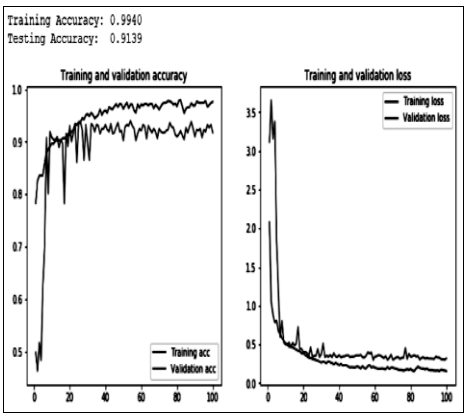

Рис. 6. *Результаты работы нейросети на полной выборке. Архетиктура Yoon Kim*

4.3. Результаты вычислительных экспериментов

Подбор оптимальной архитектуры показал хорошие результаты даже на использованной разнообразной выборке. Однако анализ слов, которые нейросеть определяла как значимые для разделения по классам, выявил, что деление на классы она осуществляет не в полной мере корректно (рис. 7, 8).

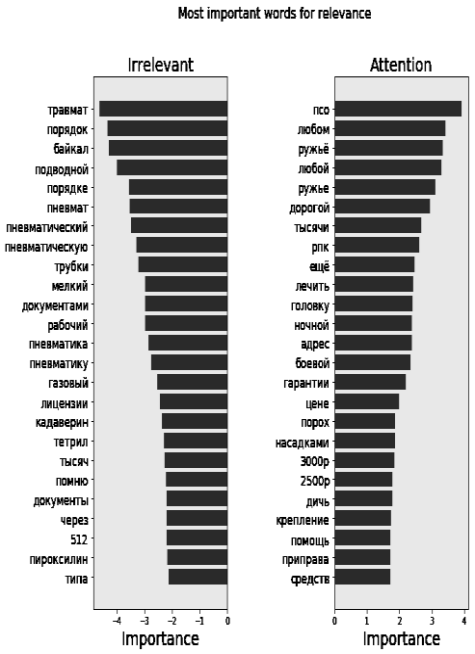

Рис. 7. *Слова, выбранные нейросетью в качестве значимых. Выборка по полному словарю. Метод векторизации Bag of words*

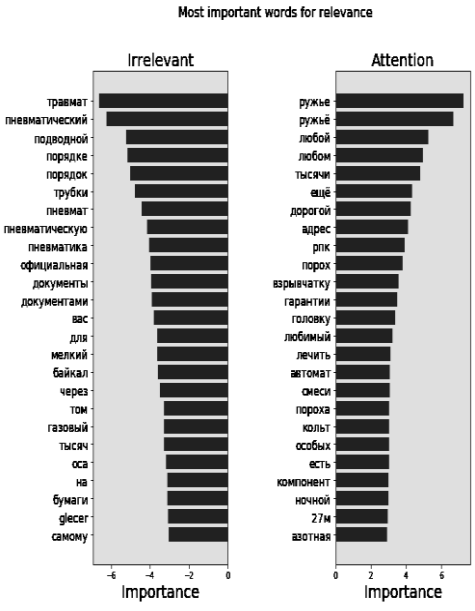

Рис. 8. *Слова, выбранные нейросетью в качестве значимых. Выборка по полному словарю. Метод векторизации TF-IDF*

Часть слов в разделенных классах действительно значимо отражает разделение. Тем не менее, многие выбранные слова являются откровенно "мусорными". Кроме того, именно на данном этапе тестирования был выявлен ряд неточностей и рассогласованностей выборки, связанный со способом разметки экспертами и особенностями выгрузки самой поисковой системы.

4.4. Работа нейронной сети на новой выборке

Созданная нейронная сеть была применена на выборке по слову "селитра". Такая выборка состояла из гораздо меньшего числа слов, поэтому был применен к ней метод аугментации. Для текстовых данных аугментация представляет собой копирование уже имеющихся текстов с четырьмя типичными "заменами":

1. Добавление слов.
2. Удаление слов.
3. Замена слов на синонимы.
4. Перестановка слов местами.

Таким образом, сгенерированная частная выборка из 600 примеров была увеличена почти в 10 раз.

В результате было соблюдено условие соответствия между количеством входных параметров и количеством примеров.

$$\frac{N_y Q}{1+\log_2(Q)} \le N_w \le N_y\left(\frac{Q}{N_y}+1\right)(N_x+N_y+1)+N_y,$$

где Q – количество примеров, Nx – число входных нейронов, Ny – число выходных нейронов, Nw – число весовых коэффициентов.

Кроме того, новая выборка была более "чистой". В дальнейшем планируется реализовать "чистку" "мусорных" слов программными средствами.

На новой выборке (рис. 9) сеть дала результат точности: 99,8 % на обучающемся и 98 % на тестовом множестве.

Анализ значимых слов, которые выявила нейросеть, показал большую содержательную точность (рис. 10). При соотнесении данных слов с текстами можно убедиться, что в большинстве случаев нейросеть выделила в качестве значимых именно те слова, которые выделил бы и человек.

В случае второй выборки нужно учитывать также проведенную нами аугментацию. Здесь представлено по 10 примеров каждого текста, отличающихся всего несколькими словами. Это зачастую увеличивает возможности сети для правильного отнесения текста к тому или иному классу.

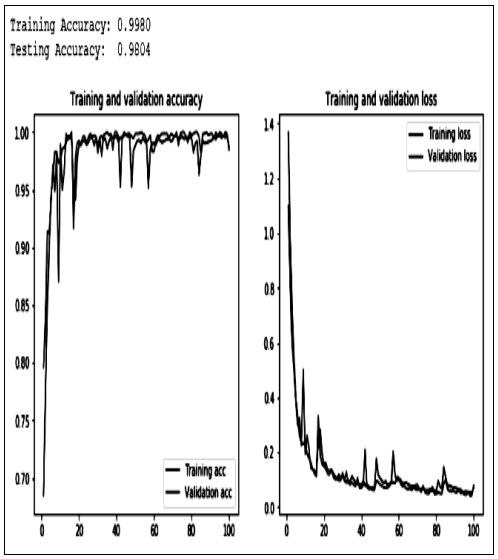

Рис. 9. *Результаты работы нейросети на новой выборке*

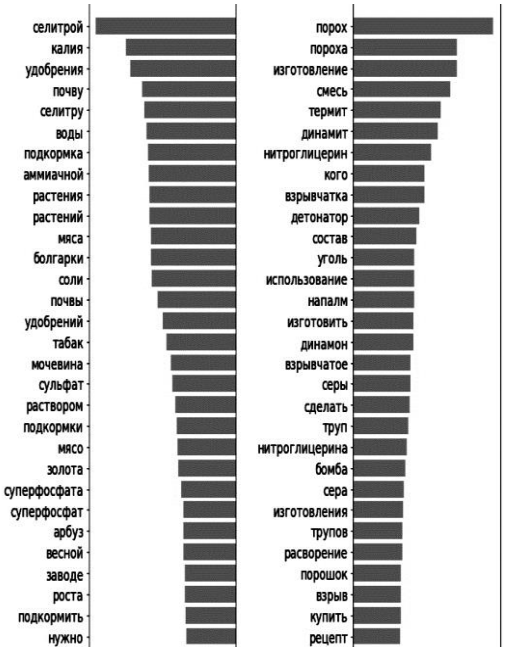

Рис. 10. *Слова, выбранные нейросетью в качестве значимых. Выборка по слову "селитра". Метод векторизации TF-IDF*

Выводы

Главный вывод, который можно сделать после проведенных экспериментов – каким бы способом ни выполнялась классификация данных, точность классификации будет зависеть от "чистоты" выборки и ее сбалансированности.

В гораздо меньшей степени точность зависит от используемого способа векторизации. И еще в меньшей степени точность зависит от выбора способа классификации.

Проведенная работа дает возможность сделать и ряд других значимых выводов:

- разметка данных экспертом не должна учитывать дополнительный анализ, не представленный в выборке – это ведет к противоречивости в данных;
- большое разнообразие ключевых слов отбора значительно усложняет возможность разделить множество только по двум классам;
- текстовые данные представляют собой большой содержательный объем, который можно интерпретировать очень разными способами. Поэтому постановка задачи, ее анализ и разработка естественной и математической моделей должны занимать самое важное место в такого рода исследованиях.

Таким образом, построение качественной модели "портрета" потенциального преступника и выделение параметров, по которым будут отбираться соответствующие тексты, является главной задачей на пути достижения поставленной цели.

Подбор выборок исключительно программными средствами может оказаться затруднителен. В связи с этим необходимо продолжить эксперименты на больших выборках, чтобы решить следующие вопросы:

- влияние аугментации данных на большой выборке на увеличение качества классификации нейросети;
- целесообразность введения мультиклассификаторов для большого множества и перечень классов для выделения;
- возможность использования более сложной системы векторизации (BERT, ELMO) для повышения "читабельности" выборки для нейросети.

Список литературы

1. *Ясницкий Л.Н.* Интеллектуальные системы: учебник. М.: Лаборатория знаний, 2016. 221 с.
2. *Тарик Р.* Создаем нейронную сеть / пер. с англ. СПб.: ООО "Альфа-книга", 2017. 272 с.
3. *Dokuka S., Koltcov S., Koltsova O., Koltsov M.* Echo Chambers vs Opinion Crossroads in News Consumption on Social Media // Analysis of Images, Social Networks and Texts: 7th International Conference, AIST 2018. Moscow, Russia, July 5–7, 2018.
4. *Kjell, O. N. E., Kjell, K., Garcia, D., & Sikström, S.* Semantic measures: Using natural language processing to measure, differentiate, and describe psychological constructs – 2019. Psychological Methods, 24(1), 92–115.
5. *Makarov I., Gerasimova O., Sulimov P., Korovina K., Zhukov L.E.* Joint Node-Edge Network Embedding for Link Prediction Media // Analysis of Images, Social Networks and Texts: 7th International Conference, AIST 2018. Moscow, Russia, July 5–7, 2018.
6. *Levent Tolga Eren, Senem Kumova Metin.* Vector Space Models in Detection of Semantically Non-compositional Word Combinations in Turkish // Analysis of Images, Social Networks and Texts: 7th International Conference, AIST 2018. Moscow, Russia, July 5–7, 2018.
7. *Лабутин И.А., Белоусов К.И., Чуприна С.И.* Кластеризация текстовых документов на основе семантических полей // Математика и междисциплинарные исследования – 2019 [Электронный ресурс]: материалы Всерос. науч.-практ. конф. молодых ученых с междунар. участием. Пермь, 2019. 10,8 Мб; 428 с.
8. *Isaeva E., Bakhtin V., Tararkov A.* Collecting the Database for the Neural Network Deep Learning Implementation // T. Antipova and A. Rocha (Eds.): DSIC 2018, AISC 850/ P. 12–18, 2019.
9. *Ji Young Lee, Franck Dernoncourt*. Sequential Short-Text Classification with Recurrent and Convolutional Neural Networks // conference paper at NAACL 2016.
10. *Alexis Conneau, Holger Schwenk, Yann Le Cun, Loic Barrault.* Very Deep Convolutional Networks for Text Classification // Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, pages 1107–1116, Valencia, Spain, April 3–7, 2017. 2017. Association for Computational Linguistics.
11. *Abdullah Alsaeedi, Mohammad Zubair Khan.* A Study on Sentiment Analysis Techniques of Twitter Data // International Journal of

Advanced Computer Science and Applications 10(2): 361–374, February 2019.
12. *Kamran Kowsari, Kiana Jafari Meimandi, Mojtaba Heidarysafa, Sanjana Mendu, Laura Barnes, Donald Brown.* Text Classification Algorithms: A Survey // Information 2019. Vol. 10(4). 150 c. URL: https://doi.org/10.3390/info10040150 (дата обращения: 20.12.2019).
13. *Brian T. Pace, Michael Tanana, Bo Xiao, Aaron Dembe, Christina S. Soma, Mark Steyvers, Shrikanth Narayanan, David C. Atkins, Zac E. Imel.* What about the Words? Natural Language Processing in Psychotherapy // Society for the Advancement of sychotherapy. URL: https://societyforpsychotherapy.org/words-natural-language-processing-psychotherapy/ (дата обращения: 20.12.2019).

Experience of applying artificial intelligence methods to identify social network users willing to commit illegal actions

V. O. Starkova, E. Yu. Nikitina
Perm State University: 15, Bukireva st., Perm, 614990, Russia
kochergina.vo@mail.ru; +7-922-309-51-24

The paper explores various methods of natural language processing (a case study of the Russian language). An attempt is made to build a text classification application using Machine Learning (ML) algorithms. The quality of classification is compared for a large sample of various texts of a general kind and a smaller sample of niche texts. Conclusions are made about the best model of data presentation for solving classification problems.

Keywords: *natural language processing; artificial intelligence; data analysis; data model; ontology; sample.*