КЛАССИФИКАЦИЯ ТЕКСТОВЫХ ЗАПИСЕЙ ПО СТЕПЕНИ ВЫРАЖЕННОСТИ ПСИХОЛОГИЧЕСКИХ ОСОБЕННОСТЕЙ[1]

Багрецов Г. И., студент Санкт-Петербургского государственного университета, gbagretsov@gmail.com

Тулупьева Т. В., к.пс.н., старший научный сотрудник лаборатории ТиМПИ СПИИРАН, доцент СПбГУ, доцент СЗИУ РАНХиГС, tvt100a@mail.ru

Аннотация

Доклад посвящён описанию модели классификации текстовых записей на страницах пользователей социальной сети «ВКонтакте». Множество классов состоит из трёх элементов — низкое, среднее и высокое значение той или иной характеристики. В работе рассматривается модель, основанная на методе опорных векторов. Такой классификатор может быть применён в задаче построения профиля психологических особенностей пользователя, на основе которого строится профиль уязвимостей пользователя.

Введение

Сегодня для нарушения информационной безопасности компаний злоумышленники всё чаще эксплуатируют не программно-технические уязвимости, а уязвимости пользователей. Несмотря на то, что для защиты от программных атак разрабатывается большое количество средств [2,4], во многих компаниях уровень информационной безопасности остаётся низким, так как одним из наиболее уязвимых мест информационной системы является пользователь [5,9]. Атаки на информационную систему, использующие не программно-технические уязвимости, а уязвимости пользователей, называются социоинженерными. В основе таких атак лежат манипулятивные воздействия на пользователя. Общая цель исследований в этой сфере заключается в построении оценки защищённости персонала информационных систем от социоинженерных атак.

В предыдущих исследованиях был разработан программный комплекс «критичные документы — информационная система — персонал — злоумышленник» [9], предназначенный для моделирования социо-

[1] Работы выполнялись в рамках проекта по государственному заданию СПИИРАН № 0073-2014-0002.

инженерных атак и для оценки уровня защищённости информационной системы от такого рода атак. На основе этого комплекса разработаны алгоритмы, имитирующие атаку злоумышленника через деревья атак и производящие оценку защищённости пользователей. Кроме того, был формализован ряд определений, таких как уязвимость пользователя и профиль уязвимостей пользователя. Последний представляет собой набор пар «уязвимость» — «выраженность уязвимости». Также был предложен подход к построению профиля уязвимостей через психологический профиль пользователя. Предполагается, что оценивать психологический профиль будет эксперт, а оценка будет производиться вручную на основе определённых методов из области психологии. Однако данный подход может быть сложно применить в случае, когда в компании работает большое количество сотрудников – эксперту необходимо обработать значительный объём информации. В то же время данные, полученные из анкет, не всегда являются достоверными.

Данная работа посвящена подходу к автоматизации построения профиля уязвимостей пользователя. В работе представлено описание моделей, с помощью которых возможно определить степень выраженности ряда психологических особенностей пользователя на основе текстовой информации. В свою очередь, текстовая информация получена со страниц пользователей в социальной сети «ВКонтакте». Эта социальная сеть является одной из самых популярных и распространённых в России [3]. Однако предложенный подход можно распространить и на другие источники текстовой информации, в т. ч. на другие социальные сети — модель не зависит от источника обучающей выборки. В работе также представлены оценки эффективности построенных моделей. Результаты показывают, что модели могут быть применены при разработке программного модуля для анализа профиля пользователя в социальной сети «ВКонтакте».

Описание задачи

Для построения психологического профиля пользователя необходимо разработать систему, способную определять степень выраженности той или иной характеристики у владельца определённой текстовой записи. Все возможные значения разбиты на три класса: низкий уровень выраженности, средний уровень выраженности, высокий уровень выраженности [10].

Поставленная задача сводится к задаче многоклассовой классификации. Требуется построить набор моделей, каждая из которых будет способна классифицировать текстовую запись по степени

выраженности той или иной характеристики.

Используемые методы

Для построения моделей, предсказывающих степень выраженности характеристик, был применён метод опорных векторов (SVM) [1]. Этот метод решает задачу бинарной классификации. Для применения его в поставленной задаче построения классификатора текстовых записей использовалась схема One-vs-One [8]. В этой схеме задача мультиклассовой классификации сводится к построению ½N(N−1) бинарных классификаторов (N – количество классов), каждый из которых разделяет объекты пар различных классов.

Также производилась оценка показателя tf-idf (term frequency – inverse document frequency) [7], которая использовалась для того, чтобы выбрать в качестве признаков в задаче классификации наиболее значимые слова. Большой вес tf-idf получат термины с высокой частотой в пределах конкретной записи и с низкой частотой употреблений в других записях.

Получение обучающей выборки

Для сбора данных, которые использовались в процессе обучения модели, были использованы результаты предыдущего исследования [9]. В частности, в ходе анкетирования, проведённого в рамках этого исследования, была получена информация о психологических профилях 90 испытуемых разного пола и возраста. Информация о каждом респонденте включает в себя адрес персональной страницы в социальной сети «ВКонтакте», пол, а также значения выраженности психологических характеристик, представленных далее. С каждой из страниц, указанных в результатах опроса, были собраны текстовые записи (посты) владельца аккаунта, находившиеся в открытом доступе. После собранные данные были сохранены для дальнейшего использования при обучении. Каждая строка полученного набора данных — это запись со страницы пользователя и значения степени выраженности каждой характеристики у этого пользователя, приведённые к соответствующим классам. Объём собранной коллекции составил 3500 единиц.

Для того, чтобы модель могла анализировать текстовые данные, необходимо предварительно их обработать. Препроцессинг текста включает в себя удаление пунктуации, цифровых и специальных символов, приведение текста в нижний регистр, удаление так называемых «стоп-слов», стемминг слов, извлечение N-грамм (последовательностей из N слов) и выбор признаков для обучения на основе меры tf-idf.

Обучение моделей и результаты

Исходя из полученной формализации задачи, были выделены параметры, которые подверглись регулировке для достижения наилучшей эффективности: параметр N, который определяет длину последовательности при извлечении N-грамм, а также количество признаков для обучения (параметр *features*).

Исходная выборка случайным образом делилась на обучающую и тестовую в соотношении 80% и 20% соответственно. Для уменьшения влияния случайного фактора при разделении исходных данных на два подмножества система проходила процедуру обучения 10 раз на каждом наборе изменяемых параметров.

Для оценки эффективности построенной модели использовались следующие показатели [6]:

- acc_{ovr} – общая достоверность (overall accuracy);
- acc_{avg} – средняя достоверность (average accuracy);
- pre_M – макро-усреднённая точность (macro-averaged precision);
- rec_M – макро-усреднённая полнота (macro-averaged recall);

Значения параметров N и features, на которых достигаются наилучшие результаты, приведены в таблице 1. Значения показателя acc_{ovr} моделей с указанными в таблице 1 параметрами показаны на рис. 1. Значения показателя acc_{avg} моделей с указанными в таблице 1 параметрами показаны на рис. 2. Значения показателя pre_M моделей с указанными в таблице 1 параметрами показаны на рис. 3. Значения показателя rec_M моделей с указанными в таблице 1 параметрами показаны на рис. 4.

Таблица 1

Характеристика	**N**	**features**
Отрицание	1	7500
Вытеснение	1	9000
Регрессия	1	9000
Компенсация	1	9000
Проекция	2	9000
Замещение	1	9000
Рационализация	1	9000
Гиперкомпенсация	1	9000

Эксперимент показал, что часть построенных моделей работают эффективнее остальных. У некоторых моделей показатели acc_{ovr}, pre_M, rec_M не превышают 0,5. Тем не менее, учитывая сравнительно небольшой объём доступной обучающей выборки и сложность задачи многоклассовой классификации, эксперимент можно назвать удачным.

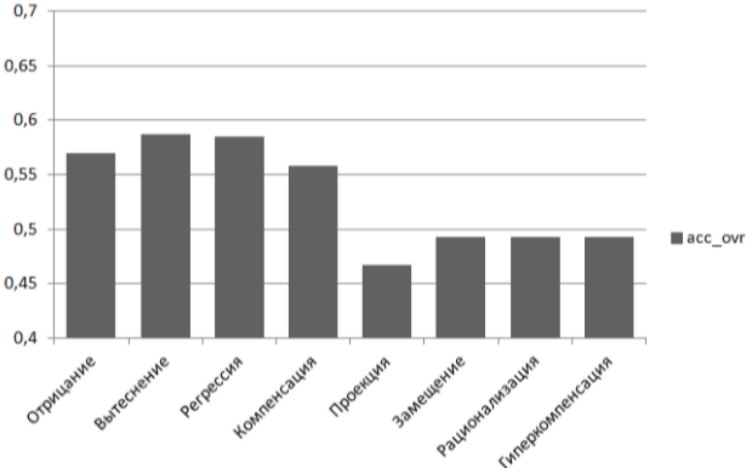

Рисунок 1: Значения показателя acc_{ovr}

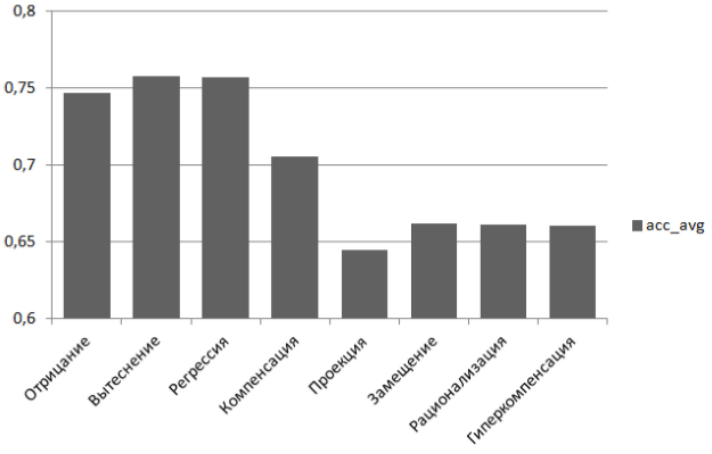

Рисунок 2: Значения показателя acc_{avg}

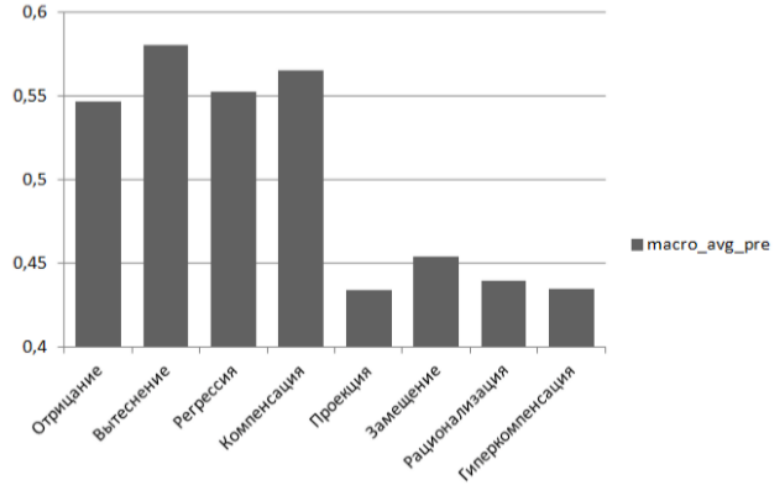

Рисунок 3: Значения показателя pre_M

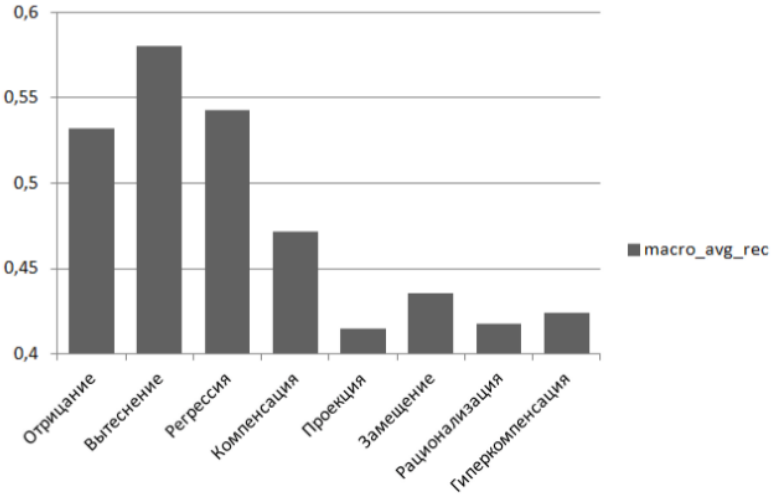

Рисунок 4: Значения показателя rec_M

Заключение

В ходе исследования был разработан метод сбора и анализа текстовой информации, который основан на применении метода опорных векторов. Была собрана обучающая выборка для модели анализа текстовых данных. Также были построены модели классификации текстовых записей по степени выраженности той или иной характеристики. Приведены данные, касающиеся эффективности моделей. Полученные модели могут быть применены при разработке программного модуля для анализа профиля пользователя в социальной сети «ВКонтакте».

Литература

1. Andrew A.M. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods by Nello Christianini and John Shawe Taylor, Cambridge University Press, Cambridge. – 2000
2. Beckers K. et al. A pattern-based method for establishing a cloudspecific information security management system //Requirements Engineering. – 2013. – Т. 18. – №. 4. – С. 343-395.
3. Brand Analytics. Социальные сети в России, зима 2015–2016. Цифры, тренды, прогнозы. URL: https://blog.br-analytics.ru/ socialnye-seti-v-rossii-zima-2015-2016-cifry-trendy-prognozy/ [дата просмотра: 01.04.2017]
4. Distefano S., Puliafito A. Information dependability in distributed systems: The dependable distributed storage system //Integrated Computer-Aided Engineering. – 2014. – Т. 21. – №. 1. – С. 3-18
5. Mitnick K.D., Simon W.L. The art of deception: Controlling the human element of security. – John Wiley & Sons, 2011.
6. Sokolova M., Lapalme G. A systematic analysis of performance measures for classification tasks //Information Processing & Management. – 2009. – Т. 45. – №. 4. – С. 427-437.
7. Sparck J.K. A statistical interpretation of term specificity and its application in retrieval //Journal of documentation. – 1972. – Т. 28. – №. 1. – С. 11-21.
8. Tax D.M.J., Duin R.P.W. Using two-class classifiers for multiclass classification //Pattern Recognition, 2002. Proceedings. 16th International Conference on. – IEEE, 2002. – Т. 2. – С. 124-127.
9. Азаров А.А., Тулупьева Т.В., Суворова А.В., Тулупьев А.Л., Абрамов М.В., Юсупов Р.М. Социоинженерные атаки: проблемы анализа. — СПб.: Наука, 2016. — 349 с.
10. Тулупьева Т.В. Психологическая защита и особенности личности в период ранней юности //СПб.: СПбГУ. – 2000.