

СЕКЦИЯ 4. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

УДК 519.8+004.855.5

Классификация школьников по цифровым следам на основе нейронной сети MLP

Р.Д. Айтенов, А.С. Маничева, В.В. Журавлева

АлтГУ, г. Барнаул

Статья посвящена исследованию различимости отдельных групп школьников по цифровым следам, оставляемым в профиле социальной сети «ВКонтакте». Классификация на отдельные группы основана на нейронной сети многослойного персептрона (MLP).

Ключевые слова: *классификация, нейронная сеть, школьник, цифровой след.*

Социальные сети используются людьми для развлечения, социализации, потребления новостей и других целей, в том числе, образовательных. Цифровые следы, оставляемые пользователями социальных сетей, позволяют проводить анализ и интерпретацию пользовательских данных в целях улучшения информационного воздействия, и представляют интерес для исследователей различных областей: маркетинговой и политической [1–2], отбора персонала [3], научной коммуникации [4], образовательного процесса, включающего в себя как сохранение контингента обучающихся вуза [5–6], так и проведение рекрутинговых мероприятий в рамках приемной кампании вуза [7–8].

Специфика использования цифровых профилей пользователей социальных сетей заключается в том, что данные профиля не всегда содержат информацию, необходимую для проведения исследования. В результате возникает задача восстановления данных в профиле пользователя [9–10]. Данная работа продолжает исследования по восстановлению данных в профиле пользователей социальной сети «ВКонтакте», проживающих на определенной территории и соответствующих определенному возрасту, с целью расширения аудитории потенциальных абитуриентов вуза и проведения с ними рекрутинговых мероприятий [11].

В работе рассматривается задача классификации пользователей социальной сети «ВКонтакте», соответствующих возрасту школьников 9-го класса и проживающих на территории Алтайского края, по месту жительства (город или сельская местность). В качестве исходных данных рассматривается наличие/отсутствие подписки на сообщество из списка, соответствующего заявленным классификационным группам. Построенную модель классификации планируется применять для восстановления отсутствующей информации по месту жительства для построения индивидуальной траектории работы с потенциальными абитуриентами вуза.

Исходные данные были получены при помощи платформы по сбору и анализу данных Университетского консорциума исследователей больших данных [12]. После предварительного анализа и обработки данных было оставлено 10200 записей, соответствующих пользователям социальной сети «ВКонтакте» возраста школьников 9-го класса, проживающих в Алтайском крае (8193 – «город», 2007 – «село»). Количество сообществ, наличие подписки на которые должно определять принадлежность пользователя к группе «город» или «село», составило 9600. Таким образом, размерность таблицы исходных данных, содержащей булевы значения – (10200 x 9600).

Классификация проводилась на основе одного из классов нейронной сети – многослойного персептрона (MLP, Multilayered perceptron) с использованием библиотек языка python. Количество нейронов на входном слое – 9600 (количество признаков), на выходном слое – 1 (метка группы, определяемая по модели). Количество скрытых слоев, количество нейронов в них, функция активации и функция оптимизации весовых коэффициентов подбирались для достижения наилучшего значения метрик качества модели: достоверность (accuracy) и, т.к. существует существенная разница в размерности рассматриваемых групп, среднее гармонической точности и полноты для каждой группы (F-мера). Объем обучающей и тестовой выборки был взят 75 и 25 процентов соответственно. Количество скрытых слоев рассматривалось от одного до трех, количество нейронов на скрытом слое – 100; 1000; 2000. Функции активации нейронов на скрытом слое: logistic (сигмоида), tanh (гиперболический тангенс), relu (линейный выпрямитель). Функции оптимизации: lbfgs, sgd, adam.

Наилучшее значение достоверности 0,78 достигается при двух скрытых слоях по 1000 нейронов при выборе relu и sgd. При этом значение F-меры для групп следующее: «город» – 0,88; «село» – 0,10, что говорит о низком качестве результатов классификации объектов, принадлежащих ко второй группе. Наилучшее значение F-меры для

группы «село» составило 0,29, при этом для «города» этот же показатель равен 0,74, а достоверность составила 0,61, что также говорит о низком качестве результатов классификации.

В дальнейшем планируется улучшить качество классификации школьников по цифровым следам на основе нейронной сети MLP за счет использования более сбалансированного набора исходных данных и применения других классов нейронных сетей.

Работа выполнена в рамках реализации Программы поддержки научно-педагогических работников ФГБОУ ВО «Алтайский государственный университет», проект «Разработка инструментов управления приемной кампанией и учебным процессом вуза с использованием технологий искусственного интеллекта посредством анализа цифровых следов школьников и студентов».

Библиографический список

1. Kosinski M. et al. Manifestations of user personality in website choice and behaviour on online social networks // Machine learning. – 2014. – Vol. 95, №3. – P. 357–380. DOI: 10.1007/s10994-013-5415-y.

2. Mangal N., Niyogi R., Milani A. Analysis of Users' Interest Based on Tweets // Computational Science and Its Applications – ICCSA 2016: 16th International Conference. – 2016. – Part V. – P. 12–23. DOI: 10.1007/978-3-319-42092-9_2.

3. Фурсов А.Л. Новые методы в отборе персонала: основные характеристики «цифрового следа» соискателя должности // Актуальные проблемы социально-гуманитарных наук и образования: сущность, концепции, перспективы: материалы VII Межд. науч. конф., Саратов, 15 апреля 2019 года. – Саратов, 2019. – С. 1057–1062.

4. Николаенко Г.А. Перспективы использования цифровых следов исследователей для анализа их коммуникативных стратегий (на примере социальной сети ResearchGate) // Социология науки и технологий. – 2019. – Т. 10, №2. – С. 93–109. DOI: 10.24411/2079-0910-2019-12005.

5. Тулупьева Т.В., Суворова А.В., Азаров А.А., Тулупьев А.Л., Бордовская Н.В. Возможности и опыт применения компьютерных инструментов в анализе цифровых следов студентов-пользователей социальной сети // Компьютерные инструменты в образовании. – 2017. – №5. – С. 3–13.

6. Feshchenko A. et al. Modelling of an educational profile of a student by analyzing public user data from social networks // INTED2018: 12th International Technology, Education and Development Conference. – Valencia, Spain, 2018. – P. 640–646. DOI: 10.21125/inted.2018.1115.

7. Feshchenko A., Goiko V., Stepanenko A. Recruiting university entrants via social networks // EDULEARN17 Proceedings: 9th International Conference on Education and New Learning Technologies, Barcelona, Spain, July 3rd-5th 2017. – Barcelona, 2017. – P. 6077–6082.

8. Журавлева В.В., Маничева А.С., Фещенко А.В., Журавлев Е.В., Журенков О.В., Козлов Д.Ю. Примеры использования машинного обучения для разработки инструментов управления приемной кампанией и учебным процессом вуза // EdCrunch-Томск: сборник тезисов Международной конференции по передовым технологиям обучения, Томск, 2-4 декабря 2020 г. – Томск, 2020. – С. 112-114.

9. Абрамов М.В., Тулупьев А.Л., Тулупьева Т.В. Агрегирование данных из социальных сетей для восстановления фрагмента мета-профиля пользователя // КИИ–2018: XVI национальная конференция по искусственному интеллекту с международным участием. Москва, 24–27 сентября 2018 года. – Москва, 2018. – С. 189–197.

10 Zhuravleva V.V., Manicheva A.S., Feshchenko A.V., Berestov A.V. Optimization of the algorithm for identifying digital traces of schoolchildren in the Altai Territory // Journal of Physics: Conference Series. – 2020. – V. 1615. DOI: 10.1088/1742-6596/1615/1/012013.

11. Журавлева В.В., Маничева А.С., Фещенко А.В., Берестов А.В. Исследование различимости цифровых следов у различных групп школьников на территории Алтайского края // Высокопроизводительные вычислительные системы и технологии. – 2020. – Т. 1, № 4. – С. 121–125.

12. Университетский консорциум исследователей больших данных [Электронный ресурс]. – Заглавие с экрана. Режим доступа: <http://opendata.university/> – Дата обращения: 20.05.2021.

УДК 528.8

Оценка экологического состояния водных экосистем Центрального Ямала

А.В. Ефанов, Л.А. Хворова

Алтайский государственный университет, г. Барнаул

Работа посвящена исследованию динамики параметров качества воды Обской губы Карского моря на основе архивной и оперативной спутниковой информации на тестовых участках, где ведется круглогодичная отгрузка в танкеры нефти и газа. В работе проанализирована динамика экологического состояния водных систем.