

Список литературы:

1. Dozat, T. Incorporating Nesterov Momentum into Adam / T. Dozat // ICLR Workshop. — 2016.
2. Dupond, S. A thorough review on the current advance of neural network structures / S. Dupond // Annual Reviews in Control. — 2019. — Vol. 14. — P. 200-230.
3. Kingma, D. Adam: A Method for Stochastic Optimization / D. Kingma, J. Ba // ICLR. — 2014.

**КЛАССИФИКАЦИЯ ПОЛЬЗОВАТЕЛЕЙ СОЦИАЛЬНЫХ СЕТЕЙ ПО
ПРОФЕССИОНАЛЬНОЙ ОРИЕНТАЦИИ**

Обрубова Василиса Денисовна

*Магистрант 2 курса кафедры информационных систем и программной инженерии
"Владимирский государственный университет имени Александра Григорьевича и Николая
Григорьевича Столетовых" (ВлГУ),
город Владимир*

Озерова Марина Игоревна

*кандидат технических наук, доцент
"Владимирский государственный университет имени Александра Григорьевича и Николая
Григорьевича Столетовых" (ВлГУ),
город Владимир*

SOCIAL NETWORKS USER CLASSIFICATION FOR PROFESSIONAL ORIENTATION

Obrubova Vasilisa Denisovna

*2nd year master's student of the Department of Information Systems and Software Engineering
"Vladimir State University named after Alexander Grigorievich and Nikolai Grigorievich Stoletovs"
(VISU),
Vladimir*

Ozerova Marina Igorevna

*candidate of technical sciences, associate professor
"Vladimir State University named after Alexander Grigorievich and Nikolai Grigorievich Stoletovs"
(VISU),
Vladimir*

DOI: 10.31618/nas.2413-5291.2021.2.68.451

АННОТАЦИЯ

В статье рассматривается комплексная постановка темы классификации пользователей социальных сетей для определения профессиональной ориентации.

ABSTRACT

The article deals with a complex formulation of the topic social networks users classification to determine professional orientation.

Ключевые слова: классификация; алгоритмы классификации; профориентология; социальные сети.

Keywords: classification; classification algorithms; vocational guidance; social networks.

Предметная область по данной теме находится на стыке трех направлений (рисунок 1). Первое направление – это профессиональное ориентирование или так называемая наука

профориентология. Второе направление – это исследование социальных сетей. И третье направление – это сам подход по решению, а именно машинное обучение.



Рисунок 1. Предметная область

Профориентология – это дисциплина, важной частью которой является профессиональное ориентирование [3]. Данное направление активно развивается и начинает затрагивать сферу социальных сетей, так как именно социальные сети содержат в себе неисчерпаемое количество информации о человеке которое можно активно

применять в психологических, социологических и прочих методах профессионального ориентирования.

Одним из подходов к обработке информации может служить машинное обучение. Математическую модель можно представить в следующем виде:

$$CLS(u_i f_j) = cls_{ij} \quad (1)$$

где CLS – функция-классификатор, ставящая в соответствие профилю и признаку классификации конкретное значение профессии. Каждому профилю u_i должно соответствовать множество значений признаков $u_i \rightarrow \{cls_{ij}\}$ [1].

Система признаков для классификации пользователей по социальным сетям f_j имеет следующий вид:

- должность;
- образование;
- год рождения;
- пол;
- языки;
- подписки – страницы на которые подписан пользователь;

- группы – группы на которые подписан пользователь.

Используя данные признаки выполняется обучение классификатора. Для проверки классификаторов предлагается использовать инструмент Weka, который отлично подходит для начала изучения различных алгоритмов и выбора наиболее подходящего, для решения поставленной задачи [2].

В помощь, при выборе классификатора, в Weka есть инструмент Experimenter. Он позволяет выбрать различные алгоритмы на одном и том же наборе данных для оценки модели. Результаты проведения проверки представлены на рисунке 2.

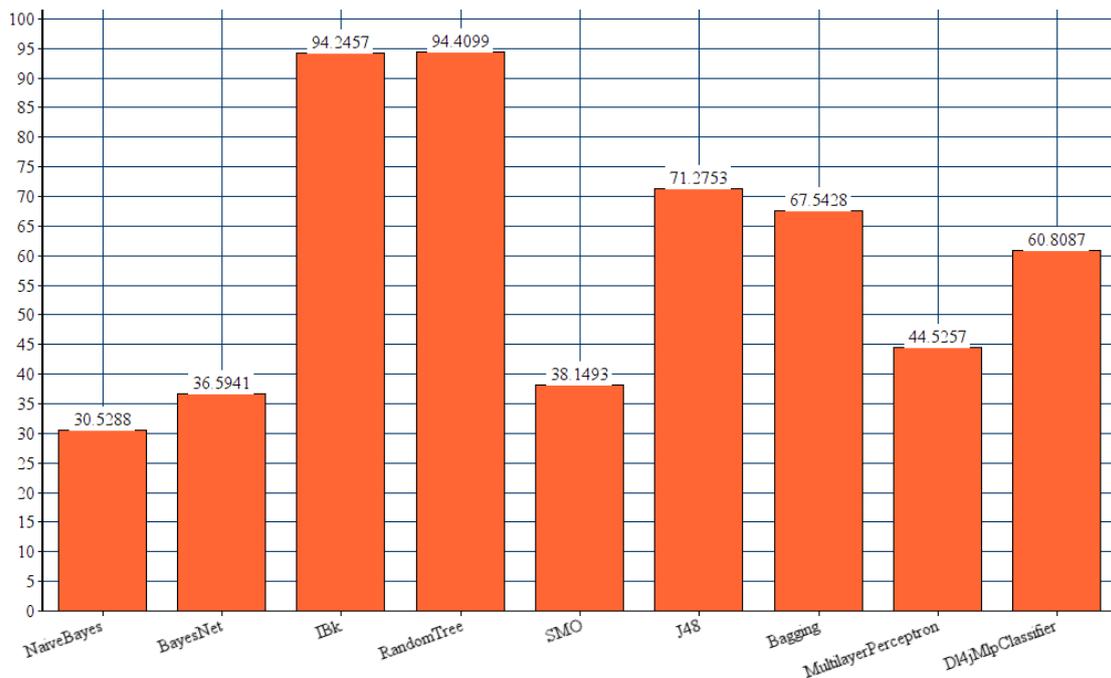


Рисунок 2. Точность оценки алгоритмов классификации

Оцениваемые классификаторы:

- NaiveBayes – метод Байеса;
- BayesNet – Байесовская сеть;
- IBk – k-Ближайшие соседи;
- SMO – машины опорных векторов;
- J48 – дерево решений;
- RandomTree - комитет решающих

деревьев;

- Bagging – метаалгоритм композиционного обучения машин, предназначенный для улучшения стабильности и точности алгоритмов машинного обучения;

- MultilayerPerceptron и D14jMlpClassifier – искусственные нейронные сети.

В результате проверки заявленных моделей выявлено, что наибольшей точностью, под текущий набор данных, обладает RandomTree. Данная модель подходит по всем критериям. Стоит заметить, что данные для обучения могут изменяться, соответственно стоит предусмотреть гибкую смену алгоритмов.

В данной работе кратко рассмотрены все три направления на стыке которых представлена

заявленная тема, данный подход рассмотрения каждой составляющей в отдельности позволяет изучить предметную область подробно за ограниченное время. Математическая модель представляет нам схему решения задачи, на базе которой выполнено применение распространенных алгоритмов классификации и выбран наиболее точный.

Список литературы

1. Tang B. et al. A Bayesian classification approach using class-specific features for text categorization // IEEE Transactions on Knowledge and Data Engineering. – 2016. – Т. 28. – № 6. – С. 1602–1606.
2. Weka Wiki – режим доступа: <https://waikato.github.io/weka-wiki/>
3. Зеер, Э. Ф. Профориентология : теория и практика : учеб. пособие для студентов вузов / Э. Ф. Зеер, А. М. Павлова, Н. О. Садовникова ; Э. Ф. Зеер, А. М. Павлова, Н. О. Садовникова. – Москва : Акад. проект ;, 2006. – (Учебное пособие для высшей школы). – ISBN 5829106841.