(Сначала рассмотрим рисунки в конце статьи)

<https://education.yandex.ru/handbook/ml/article/kak-ocenivat-veroyatnosti>

Каширин И.Ю.

КАЛИБРОВКА ML-МОДЕЛЕЙ ДЛЯ ИДЕНТИФИКАЦИИ

ФЕЙК-НОВОСТЕЙ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Рязанский государственный радиотехнический университет имени В.Ф. Уткина», Рязань

Модели машинного обучения с эффективностью используются не только в задачах прогнозирования, но и в задачах идентификации объектов предметной области или сложных ситуативных структур [1,2].

В частности, такие модели используют для исследования естественно-языковых текстов, например, с целью определения эмоциональной окраски текста или отнесения электронных новостей к классам различной степени достоверности. Существует множество зарубежных разработок в области автоматического отнесения новости к фейку или истине [3]. Такие идентифицирующие системы приобретают все большую популярность, особенно в молодежной среде.

В то же время зарубежные автоматизированные системы построены на аксиоматике множества неоспоримых исторических фактов, и сделанные ими заключения об отнесении новости к фейку основаны на этих первоначальных фактах. Такими фактами, например, могут быть «Победителями во второй мировой войне были Англия и Америка» или «Вероятнее всего, газопровод «Северный поток» был уничтожен Россией». Если исходить из подобных «фактов» как из априори истинных аксиом, любые электронные новостные материалы будут оцениваться соответствующим образом.

Изложенные соображения свидетельствуют о необходимости создания отечественных программных продуктов, трактующих исторические факты, исходя из документально подтвержденных исторических документов и информационных материалов.

Предлагаемая в настоящей статье технология идентификации достоверности новостей основана на изучении особенностей существующих в информационной среде реалий и использует механизм калибровки моделей машинного обучения для вычисления вероятности отнесения новостного материала к недостоверным сведениям.

Калибровка предполагает уточнение бинарного классификационного признака («ложь», «истина») более точным числом, представляющим собой вероятность признания факта лживым или достоверным. Существующая библиотека scikit-learn 1.2.2 дает возможность использовать множество готовых методов калибровки, например OneVsOneClassifier:

**from** **sklearn** **import** datasets

**from** **sklearn.multiclass** **import** OneVsOneClassifier

**from** **sklearn.svm** **import** LinearSVC

X, y = datasets.load\_iris(return\_X\_y=**True**)

OneVsOneClassifier(LinearSVC(random\_state=0)).

fit(X, y).predict(X)

В предлагаемом здесь подходе используются лингвистические правила, основанные на наборах именных групп как фрагментов синтограмм, применяемых на этапе парсинга для получения дополнительных коэффициентов достоверности (таблица 1).

Таблица 1 – Фрагмент набора именных групп индикаторов лжи

|  |  |  |  |
| --- | --- | --- | --- |
| № | Признаки достоверной/недостоверной новости | Формула синтограммы | Базовый  коэффициент  достоверности |
| 1 | Автор избегает конкретных цифр | «было убито много» X,  «большое количество жертв» X | -0,2 |
| 2 | Лицо, заинтересованное в фейке, но опровергающее его, может говорить правду. | «мы были неправы»,  «приносим извинения за X» | +0,1 |
| 3 | Лицо, узнавшее о важной информации, делает вид, что не знает этой информации. | «нам неизвестен факт» X, X «не могло происходить» | -0,7 |
| 4 | Косвенная ложь | «представитель X сообщил, что», «X говорит, что» | -0,5 |

Более полный фрагмент правил приведен в [4]. Для применения правил используется база знаний, содержащая продукции следующего вида:

<Синтограмма> <Новостной факт X1><Новостной факт X2><…>

<Новостной факт XN> <Электронное издание><Автор> =>

KBChange(<X1,K>,<X2,K>,<…>,<XN,K>,

<Электронное издание,K>,<Автор,K>),

где <Синтограмма> − применение схемы синтагрммы при парсинге текста новости, <Новостной факт XI> − текст изложенного в новости факта, <Электронное издание> и <Автор> − наименование электронного новостного ресурса и имя автора соответственно. KBChange – метод актуализации базы знаний исправлением рейтинга новости, издания и автора коэффициентом К из таблицы 1.

Для калибровки модели обучения может быть применена дополнительная конструкция:

KBChange(LinearSVC(random\_state=0)).fit(X, y).predict(X).

В настоящее время программный инструментарий, проектируемый по предложенной технологии, находится в разработке.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Каширин И.Ю**.** Применение теории иерархических чисел в проектировании ICF-таксономии для оптимизации нейронных сетей. Вестник Рязанского государственного радиотехнического университета. Рязань, 2022. № 80 – с.36-43.

## Сатыбалдина Д.Ж., Овечкин Г.В., Калымова К.А. Система распознавания статических жестов рук с использованием камеры глубины. Вестник рязанского государственного радиотехнического университета. Рязань, 2020. № 72 с. 93-105.

1. Maria D. Molina, S. Shyam Sundar, Thai Le, and Dong-won Lee. “Fake news” is not simply false information: A concept explication and taxonomy of online content. American Behavioral Scientist, 2021, 65(2):180–212.
2. Каширин И.Ю. идентификация достоверности новостей с помощью моделей машинного обучения Вестник Рязанского государственного радиотехнического университета. Рязань, 2023. № 83. – С.36-47.









