

струйный смеситель вовлекает больше донных отложений.

Литература:

1. ГОСТ 31385-2016 Резервуары вертикальные цилиндрические стальные для нефти и нефтепродуктов. Общие технические условия (с Поправками)

2. Устройство для перемешивания жидкостей в резервуарах/ В. Ф. Салихова, Ю. Р. Галиакбаров, М. Ф. Галиакбаров, И. М. Галиакбаров: пат. 2189852 Рос. Федерация. № 2001110507/12; заявл. 17.04.2001; опубл. 27.09.2002, Бюл. 2002. 27. С. 405.

3. Галиакбарова Э.В., Бахтизин Р.Н., Галиакбаров В.Ф. Использование струйных гидравлических смесителей для интенсификации процессов подготовки нефти к переработке // Нефтегазовое дело. 2016. Т. 14. № 1. С. 145 – 149

4. Совершенствование систем предотвращения накопления донных отложений в резервуарах большой вместимости/ В. Н. Александров, В. А. Галканов, Ю. К. Кириллов, С. Н. Мальцев, Б. Н. Мастобаев, Р. Н. Бахтизин, А. А. Локшин // Нефтяное хозяйство. 2001. № 2. С.70.

5. Сидоров Г.М., Яхин Б.А., Ахметов Р.Ф. Моделирование работы статического смесителя (нефть - вода) для обессоливания нефти и опытно-промышленное испытание // Успехи современного естествознания. 2017. № 2. С. 152 - 156

6. Галиакбаров В.Ф., Галиакбарова Э.В., Шварева Е.Н., Белозеров А.Е., Жолобова Г.Н. Повышение эффективности процессов обезвоживания, обессоливания и защелачивания нефти в смесительных аппаратах ЭЛОУ // Информационные технологии. Проблемы и решения: матер. Междунар. науч.-практ. конф. / редкол.: Ф.У. Еникеев и др. Уфа: Изд-во «Восточная печать», 2015. Т. 1. С. 188 - 190

7. Галиакбарова Э.В., Валявин Г.Г., Галиакбаров В.Ф. Внедрение струйного гидравлического смесителя для пожарной безопасности и эффективной работы резервуарных парков: ФГБОУ ВПО «Уфимский государственный нефтяной технический университет», г. Уфа, Российская Федерация ООО «НТ-Центр», г. Уфа, Российская Федерация

8. Чепур П. В., Тарасенко А. А. Особенности совместной работы резервуара и устройств размыва донных отложений винтового типа //Фундаментальные исследования. Техн. науки. 2015. №2. С.1671- 1675.

9. Галиакбаров В. Ф., Кононов О. В., Мастобаев Б. Н. Анализ и классификация существующих способов борьбы с отложениями в нефтяных емкостях. Уфа: Изд-во «Реактив», 2010. 40 с.

10. Ахметов С.А. Технология глубокой переработки нефти и газа: учеб. Пособие для вузов / С.А. Ахметов. – СПб.: Недра, 2013. – 544 с.

НЕСБАЛАНСИРОВАННОСТЬ КЛАССОВ ПРИ РЕШЕНИИ ЗАДАЧИ КЛАССИФИКАЦИИ 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*

IMBALANCE OF CLASSES IN SOLVING THE PROBLEM OF 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.2413-5291.2021.2.68.449

АННОТАЦИЯ

Проблему несбалансированности данных часто недооценивают при решении задач классификации. Модель классификации, так на первый взгляд хорошо обученная на ваших данных и дающая хороший процент распознавания, может оказаться недостоверной. Рассмотрение данной проблемы в контексте

конкретной задачи классификации пользователей социальных сетей по профессиональной ориентации позволит понять, как, зачем и, самое важное, когда нужно избавляться от несбалансированности данных.

ABSTRACT

The problem of data imbalance is often underestimated when solving classification problems. A classification model that looks well trained on your data and gives a good recognition rate may not be reliable. Consideration of this problem in the specific task of classifying users of social networks will make it possible to understand how, why and, most importantly, when it is necessary to get rid from data imbalances.

Ключевые слова: классификация; несбалансированность классов; парадокс точности; точность распознавания.

Keywords: classification; imbalance of classes; the paradox of accuracy; recognition accuracy.

Несбалансированность данных представляет из себя проблему, когда классы представлены не однородно [4]. Так, при решении задачи классификации пользователей социальных сетей возникла данная проблема. Данные о

пользователях для обучения классификатора выгружались из групп IT направления, но распределение по классам получилось не равномерное. Для демонстрации часть распределения представлено на рисунке 1.

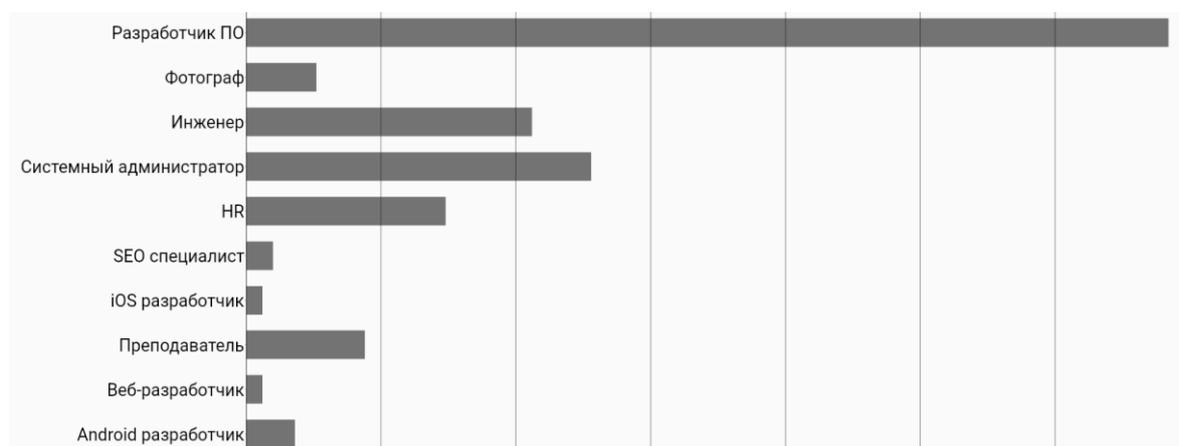


Рисунок 1. Распределение по классам

Классификаторы при обучении на несбалансированных данных смотрят на данные и решают, что лучший выбор – это всегда прогнозировать «Разработчик ПО» и добиваться высокой точности. При изучении правил в окончательной модели, можно увидеть, что оно очень вероятно предсказывает один класс независимо от данных [1]. Все это вызывает парадокс точности, когда показатель точности распознавания достаточно высок, но точность отражает только распределение базовых классов.

Рассмотрим, как же можно обнаружить данную проблему:

- проверять результаты – если процент распознавания всегда высокий, но результат неудовлетворительный это повод задуматься, этот показатель и являлся моим поводом изучения данного вопроса;
- использовать более сложные функции распознавания точности – например, одной из самых популярных оценок точности является ассигасу, при использовании которой часто могут возникать некорректные заключения о работе классификатора [5];
- смотреть на косвенные признаки – так, например, высокая скорость классификации может указывать на то, что классифицируется только класс большинства, в нашем случае это класс «Разработчик ПО»;

Теперь рассмотрим, что можно сделать, когда данная проблема обнаружилась [2].

1) Убедитесь, что проблема у вас есть. Помните, что не все модели машинного обучения плохо работают с несбалансированными классами. Большинство вероятностных моделей слабо зависят от баланса классов.

2) Попробуйте применить другой алгоритм. Как уже оговаривалось выше, не привязывайтесь к одному алгоритму. Возможно вам будет достаточно поменять алгоритм. Так, например, при решении задачи можно воспользоваться таким инструментом как Weka в качестве платформы для машинного обучения. В нем присутствует множество алгоритмов машинного обучения и еще очень интересное приложение Experimenter, которое служит для сравнения предсказательной способности алгоритмов машинного обучения [3].

3) Выполните донастройки алгоритма, например, регулировку весов или же измените алгоритм с целью повышения чувствительности к редким классам.

4) Соберите больше данных – самый простой и верный способ по решению проблемы. Так для обучения моего классификатора, была произведена выгрузка из других сообществ где преобладают представители профессий, выборки по которым были не достаточны.

5) Уберите лишние данные – это может показаться вам странным, но данный подход довольно часто применим. Так, для обучения классификатора для определения профессиональной направленности на данном этапе мне была важна классификация IT профессий. Было правильным решением удалить классы, представленные недостаточной выборкой, и не относящиеся к IT.

6) Создайте синтетические данные – этот процесс относится к расширению выборки, но происходит искусственным путем. Например, если у вас нет возможности получить дополнительных данных, вы можете применить систематические алгоритмы для генерации синтетических образцов. Так, при помощи Weka, вы можете использовать SMOTE контролируемый фильтр. SMOTE – это метод передискретизации, который создает синтетические образцы [3].

Это не единственные способы по решению данной проблемы, но они являются самыми, на мой взгляд, понятными и наиболее применяемыми. Помните, что нет единственно верного подхода для решения проблемы несбалансированности. Все подходы можно применять в комплексе или выбрать один, и избежать проблем. Данная работа

демонстрирует подходы к обнаружению и исправлению несбалансированности классов, а также показывает на решении конкретной задачи как их стоит применять.

Список литературы

1. 8 тактик для борьбы с несбалансированными классами в вашем наборе данных машинного обучения – режим доступа: <https://www.machinelearningmastery.ru/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/>
2. Samir, A. Transfer learning for class imbalance problems with inadequate data / A. Samir, K. R. Chandan // Knowl Inf Syst 48. — 2015. — P. 201–228.
3. Weka Wiki – режим доступа: <https://waikato.github.io/weka-wiki/>
4. Махсотова, Ц. В. Исследование методов классификации при несбалансированности классов / Ц. В. Махсотова // Научный журнал. – 2017. – № 5(18). – С. 35-36.
5. Старовойтов В. В. Об оценке результатов классификации несбалансированных данных по матрице ошибок / В. В. Старовойтов, Ю. И. Голуб // Информатика. – 2021. – Т. 18, № 1. – С. 61–71.

ПРИМЕНЕНИЕ LSTM-СЕТИ В РЕШЕНИИ ЗАДАЧИ ПРОГНОЗИРОВАНИЯ МНОГОМЕРНЫХ ВРЕМЕННЫХ РЯДОВ

Обрубов Максим Олегович

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

Кириллова Светлана Юрьевна

*кандидат технических наук, профессор
“Владимирский государственный университет имени Александра Григорьевича и Николая
Григорьевича Столетовых” (ВлГУ),
город Владимир*

USING LSTM NETWORK FOR SOLVING THE MULTIDIMENSIONAL TIME SERIES FORECASTING PROBLEM

Obrubov Maxim Olegovich

*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*

Kirillova Svetlana Yurievna

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

DOI: 10.31618/nas.2413-5291.2021.2.68.450

АННОТАЦИЯ

В статье рассматривается применение технологии рекуррентных нейронных сетей к задаче прогнозирования многомерных временных рядов. Выполнено экспериментальное определение архитектуры нейронной сети и основных гиперпараметров для достижения минимальной погрешности. Выявленная структура сети будет использоваться далее для определения аномалий в многомерных временных рядах.