

Научная статья

DOI 10.15593/2499-9873/2022.4.09

УДК 659.4.012 : 004.8

А.А. Минин, С.А. Федосеев

Пермский национальный исследовательский
политехнический университет, Пермь, Россия

ПРИМЕНЕНИЕ МЕТОДА СЛУЧАЙНОЙ ДИХОТОМИЧЕСКОЙ КЛАССИФИКАЦИИ ДЛЯ ОЦЕНКИ ЗАСТРОЙЩИКА ПО ОТЗЫВАМ ПОТРЕБИТЕЛЕЙ

Бренд влияет на успешность и устойчивость застройщика. В настоящей работе авторы продолжают решать задачу управления брендингом застройщика на основе информации в сети Интернет. Предложен метод дихотомической классификации, который позволяет классифицировать отзывы о застройщике на основе случайной маркировки данных. Метод дихотомической классификации позволяет работать с небольшими базами данных, в отличие от методов на основе построения искусственных нейронных сетей, для которых необходимы большие выборки данных. Метод был реализован в системе «Эйдос». В основе системы «Эйдос» заложен автоматизированный системно-когнитивный анализ (АСК-анализ). Решение об изменении маркера отзыва принималось на основе изменения достоверности модели по F-мере Ван Ризбергена.

Апробация применения метода дихотомической классификации для классификации отзывов о застройщиках произведена на примере Акционерного общества «Пермский завод силикатных панелей» (ПЗСП). При обработке текстовой информации методом дихотомической классификации удалось в собранной выборке выделить 20 негативных и 63 позитивных отзыва.

Показана работоспособность и адекватность использованного метода. В частности, показано, что можно разделить собранную текстовую информацию о застройщике на два кластера без предварительного обучения.

Метод дихотомической классификации позволяет сократить время реакции на упоминание о застройщике. В настоящей работе предложен алгоритм кластеризации отзывов на основе метода дихотомической классификации. Данный алгоритм может быть положен в основу программного обеспечения для сбора и анализа данных о застройщике, а также управления брендингом застройщика.

Ключевые слова: дихотомическая классификация, кластеризация, анализ отзывов, обучение без учителя, репутационный риск, управление брендингом, бренд, брендинг застройщика, системно-когнитивный анализ, АСК-анализ, F-мера Ван Ризбергена.

A.A. Minin, S.A. Fedoseev

Perm national research polytechnic university, Perm, Russian Federation

APPLICATION OF THE RANDOM DICHOTOMOUS CLASSIFICATION METHOD TO DEVELOPER'S BRAND ASSESSMENT BASED ON CONSUMER'S REVIEWS

The brand influences the success and sustainability of the builder. In this paper, the authors continue to address the problem of managing developer branding based on information on the Internet. The paper proposes a dichotomous classification that allows you to classify developer reviews based on

randomly labeled data. The dichotomous classification method allows you to work with small databases, in contrast to methods based on the construction of artificial neural networks, which require large samples of data. The method was implemented in the "Eidos" system. The Eidos system is based on automated system cognitive analysis (ASC-analysis). The decision to change the recall marker was made on the basis of changing the credibility of the model by Van Riesbergen's F-measure.

The application of the dichotomous classification method to classify developers' reviews was tested on the example of OJSC "Perm Silicate Panels Plant" (PZSP). The dichotomous classification method allowed us to distinguish 20 negative and 63 positive reviews in the collected sample.

The performance and adequacy of the method used is shown. In particular, it is shown that it is possible to divide the collected textual information about the developer into two clusters without prior training.

The dichotomous classification method reduces the developer's response time to a reference. This paper proposes an algorithm for reference clustering based on the dichotomous classification method. This algorithm can be used as a basis for software to collect and analyze developer data, as well as to manage developer branding.

Keywords: dichotomous classification, clustering, feedback analysis, teacherless learning, reputational risk, branding management, branding, developer branding, system-cognitive analysis, ASK-analysis, Van Riesbergen's F-measure.

Введение

В современном мире успешность и устойчивость предприятия в значительной степени определяются его брендом. В работе [1] была показана значимость репутационных рисков для бренда предприятия-застройщика. Особого внимания при оценке репутационных рисков застройщика заслуживают факторы, определяющие взаимодействие застройщика с клиентами. Именно клиенты в наибольшей мере оказывают влияние на имидж и репутацию застройщика. В наши дни основным каналом передачи и сбора информации о застройщиках является Интернет. Отзывы и упоминания о застройщике в Интернете – это первое, с чем знакомятся заинтересованные лица, когда начинают изучать рынок. Таким образом, актуальной для предприятий-застройщиков является задача мониторинга и управления своей репутацией в сети Интернет.

В настоящее время в компаниях застройщиках существуют отделы маркетинга, которые работают над созданием бренда, в том числе и в сети Интернет. Маркетологи и копирайтеры вручную или через сторонние сервисы осуществляют поиск в Интернете информации о застройщике, определяют её окрас, значимость, прогнозируют, какой эффект окажут эти упоминания на застройщика, и принимают решения, каким образом на такие упоминания реагировать. Однако проблема заключается в том, что у большинства застройщиков автоматизация подобных исследований находится на низком уровне, что существенно ограничивает скорость и масштаб исследований и в результате значительно снижает их своевременность и точность.

Так, в работе [1] при комплексной оценке репутационного риска застройщиков учитывались отзывы о застройщиках в сети Интернет. Эти отзывы собирались и обрабатывались вручную, а именно каждый отзыв исследователю приходилось читать и классифицировать как положительный или отрицательный.

В данной работе предложен подход, который позволит решить часть обозначенной выше проблемы, а именно автоматизировать разделение собранных упоминаний о застройщике на положительные и отрицательные. Как следствие, предложенный подход поможет сократить время требуемой реакции на упоминания о застройщике.

Теория и методы

Для решения задач кластеризации широко используются различные методы, связанные с применением искусственных нейронных сетей. Так, в работах [2–4] рассматривается вопрос машинного обучения нейронных сетей в задачах, близких по содержанию к задаче, которая поставлена в настоящей работе. Стоит отметить, что использование нейронных сетей требует значительного объема исходной информации. Однако при оценке репутационного риска для застройщиков может быть найдено или получено относительно небольшое количество отзывов (до нескольких десятков).

В настоящей работе для решения задачи кластеризации отзывов о застройщиках использован метод случайной дихотомической классификации. Данный метод впервые был предложен в работе [5] для классификации текстовой информации на основе случайной маркировки данных. Основная идея метода заключается в том, чтобы сформировать два кластера на основе случайно промаркированных данных без предварительного обучения.

В рамках задачи оценки и управления репутационными рисками застройщика интерес представляют только положительные и отрицательные отзывы о застройщике. Благодаря полярной противоположности исходной текстовой информации, необходимо выявить совпадения в случайно сформированных двух группах отзывов. Найденные отзывы будут типичны для одной группы или для другой. Далее при помощи последовательной смены маркера случайно выбранной группы у каждого отзыва предполагается сформировать два результирующих кластера отзывов и в заключение продемонстрировать эти кластеры

эксперту, который определит, какой из кластеров положительный, а какой отрицательный.

Применение метода дихотомической классификации позволяет в автоматизированном режиме осуществить разделение собранных упоминаний о застройщике на положительные и отрицательные, что приведет к значительной экономии времени. Метод дихотомической классификации был реализован в системе «Эйдос», в которой используется автоматизированный системно-когнитивный анализ (АСК-анализ) [6]. В данной системе реализованы инструменты для формализации предметной области (конструирование классификационных и описательных шкал, градаций, а также формирование обучающей выборки) [7–9].

Пример применения метода дихотомической классификации

В работе [1] самый низкий уровень репутационного риска был определен для Акционерного общества «Пермский завод силикатных панелей» (ПЗСП), поэтому данный застройщик был выбран в качестве примера не только для дополнительной оценки его репутации, но и для проверки адекватности метода дихотомической классификации.

В качестве исходных данных были собраны 82 отзыва о застройщике ПЗСП, оставленных клиентами на площадке «Яндекс.Карты» в период с 09.08.2018 по 13.06.2022. Каждый отзыв представляет собой текст, в котором клиент описывает достоинства ПЗСП, которые он смог для себя отметить во время работы с предприятием, либо описывает негативный опыт и даже предостерегает других потенциальных клиентов от работы с данным предприятием. На рис. 1 представлен короткий положительный текстовый отзыв, а на рис. 2 представлен объёмный текстовый отзыв, который несет в себе негативную информацию. На рис. 3 представлены текстовые отзывы среднего объема, которые одновременно несут позитивную и негативную информацию.

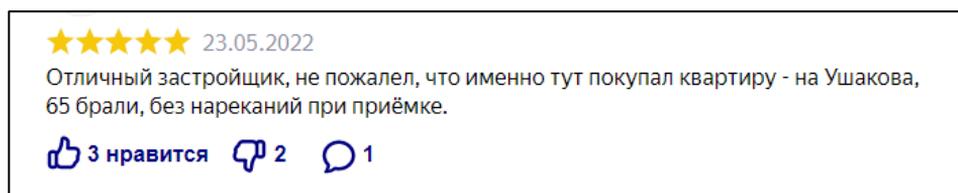


Рис. 1. Пример текста короткого положительного отзыва

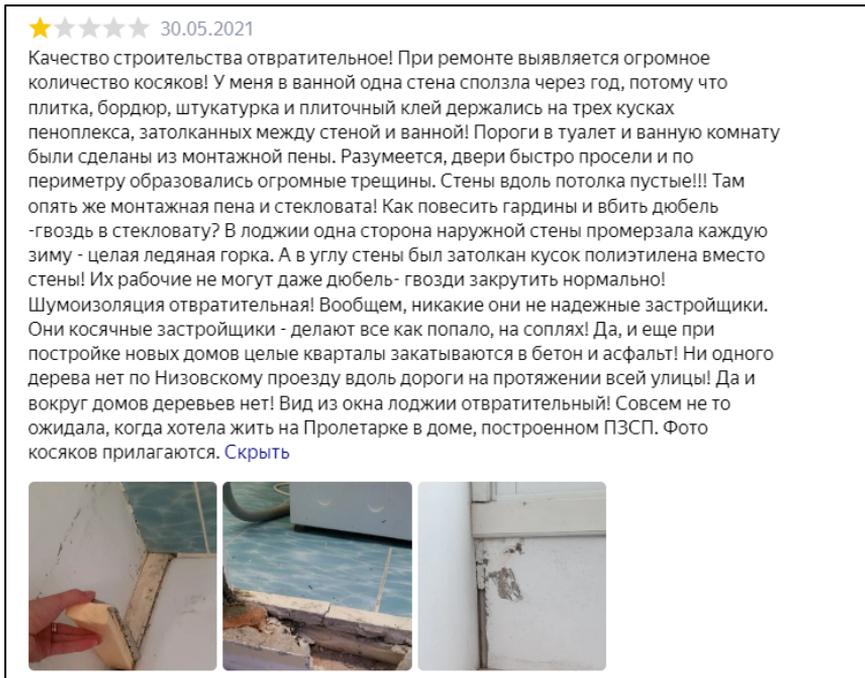


Рис. 2. Пример текста объемного отрицательного отзыва

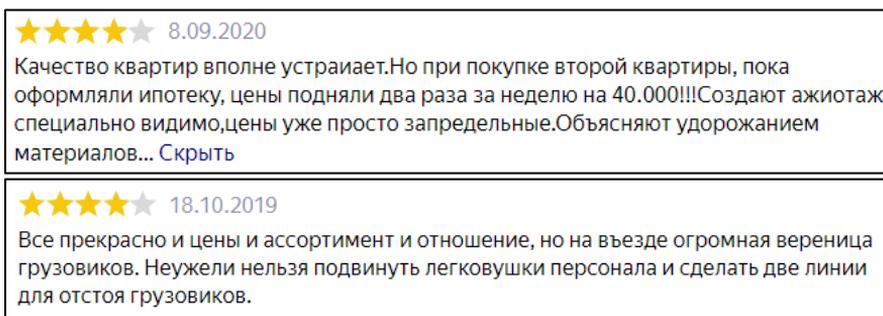


Рис. 3. Примеры одновременно положительного и отрицательного отзыва

Полученная выборка из 82 отзывов была загружена в таблицу Excel и промаркирована случайным образом с помощью следующей формулы =ЕСЛИ(СЛУЧМЕЖДУ(0;100)<50;"А";"Б"). Процедура случайной маркировки отзывов была выполнена 10 раз. Каждый из 10 вариантов сравнивался с остальными по F -мере Ван Ризбергена [10; 11]. Для варианта № 7 по модели INF3 было получено максимальное значение F -меры Ван Ризбергена, равное 0,909 (рис. 4), именно поэтому данный

вариант маркировки отзывов был выбран для дальнейших исследований.

14. Обобщенная форма по достоям моделей при разн. крит. Текущая модель: "INF3"

Наименование модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно положительных решений (TP)	Число истинно отрицательных решений (TN)	Число ложно положительных решений (FP)	Число ложно отрицательных решений (FN)	Точность модели	Полнота модели	F-мера Ван Ризбергена
1. ABS - частный критерий: количество встреч слов/знак. "Акс."	Корреляция абс. частот с обр...	82	82	27	55		0.599	1.000	0.749
1. ABS - частный критерий: количество встреч слов/знак. "Акс."	Средняя абс. частот по признакам...	82	82	4	78		0.513	1.000	0.678
2. PRIC1 - частный критерий: усл. вероятность его признака сред.	Корреляция усл.отн. частот с о...	82	82	27	55		0.599	1.000	0.749
2. PRIC1 - частный критерий: усл. вероятность его признака сред.	Средняя усл.отн. частот по признакам...	82	82	4	78		0.513	1.000	0.678
3. PRIC2 - частный критерий: условная вероятность его признака.	Корреляция усл.отн. частот с о...	82	82	27	55		0.599	1.000	0.749
3. PRIC2 - частный критерий: условная вероятность его признака.	Средняя усл.отн. частот по признакам...	82	82	4	78		0.513	1.000	0.678
4. INF1 - частный критерий: количество значений по А/Заркевину, в.	Семантический резонанс зна...	82	65	81	1	17	0.985	0.793	0.878
4. INF1 - частный критерий: количество значений по А/Заркевину, в.	Средняя значения	82	78	63	19	4	0.804	0.951	0.872
5. INF2 - частный критерий: количество значений по А/Заркевину, в.	Семантический резонанс зна...	82	65	81	1	17	0.985	0.793	0.878
5. INF2 - частный критерий: количество значений по А/Заркевину, в.	Средняя значения	82	77	66	16	5	0.828	0.939	0.880
6. INF3 - частный критерий: Условный резонанс между факт...	Семантический резонанс зна...	82	74	74	8	8	0.902	0.902	0.902
6. INF3 - частный критерий: Условный резонанс между факт...	Средняя значения	82	74	74	8	8	0.902	0.902	0.902
7. INF4 - частный критерий: ROI (Return On Investment), веро...	Семантический резонанс зна...	82	65	81	1	17	0.985	0.793	0.878
7. INF4 - частный критерий: ROI (Return On Investment), веро...	Средняя значения	82	79	62	20	3	0.798	0.963	0.873
8. INF5 - частный критерий: ROI (Return On Investment), веро...	Семантический резонанс зна...	82	65	81	1	17	0.985	0.793	0.878
8. INF5 - частный критерий: ROI (Return On Investment), веро...	Средняя значения	82	78	64	18	4	0.813	0.961	0.876
9. INF6 - частный критерий: ранг усли и безусли вероятностей, вер.	Семантический резонанс зна...	82	64	74	8	18	0.889	0.780	0.831
9. INF6 - частный критерий: ранг усли и безусли вероятностей, вер.	Средняя значения	82	74	58	24	8	0.755	0.902	0.822
10. INF7 - частный критерий: ранг усли и безусли вероятностей, ве.	Семантический резонанс зна...	82	64	74	8	18	0.889	0.780	0.831
10. INF7 - частный критерий: ранг усли и безусли вероятностей, ве.	Средняя значения	82	73	61	21	9	0.777	0.890	0.820

Рис. 4. Пример интерфейса системы «Эйдос»

F-мера Ван Ризбергена [12–14] определяется как взвешенное гармоническое среднее от точности и полноты информационного поиска

$$F = \frac{1}{\alpha \frac{1}{P} + (1 - \alpha) \frac{1}{R}}, \quad \alpha \in [0, 1],$$

где P – точность (precision) информационного поиска; R – полнота (recall) информационного поиска.

В рамках настоящего исследования $\alpha = 0,5$, при этом F-мера называется сбалансированной и может быть определена следующим образом:

$$F = \frac{2PR}{P + R}.$$

Точность и полнота необходимы для оценки эффективности информационного поиска. Точность – это доля объектов, действительно принадлежащих данному классу, относительно всех объектов, которые были отнесены к данному классу. Точность может быть определена следующим образом:

$$P = \frac{TP}{TP + FP},$$

где TP – количество истинных положительных (True Positive) результатов, т.е. число объектов, правильно отнесенных к классам, к которым они действительно относятся;

FP – количество ложных положительных (False Positive) результатов, т.е. число объектов, неправильно отнесенных к классам, к которым они на самом деле не относятся.

Полнота – это доля найденных объектов, принадлежащих классу, относительно всех объектов этого класса среди найденных. Полнота может быть определена следующим образом

$$R = \frac{TP}{TP + FN},$$

где FN – количество ложных негативных (False Negative) результатов, т.е. число объектов, неправильно не отнесенных к классам, к которым они на самом деле относятся.

В рамках настоящего исследования объекты – это отзывы о застройщике, а классы – это отзывы отрицательные и отзывы положительные.

Модель INF3 представляет собой разность между фактическими и теоретически ожидаемыми абсолютными частотами

$$I_{ij} = N_{ij} - \overline{N_{ij}} = N_{ij} - \frac{N_i N_j}{N},$$

где N_{ij} – фактическое число появлений i -го признака у объектов j -го класса;

$\overline{N_{ij}}$ – теоретическое число появлений i -го признака у объектов j -го класса;

N_i – суммарное число появлений i -го признака во всех классах;

N_j – суммарное число появлений всех признаков в j -м классе;

N – суммарное число появлений всех признаков во всех классах.

В рамках настоящего исследования признаки – это слова из собранных отзывов о застройщике.

В системе «ЭЙДОС» также реализованы и другие модели, представленные на рис. 5.

После того как была найдена начальная достоверность модели по F -мере, последовательно менялись маркеры для каждого отзыва. Если

при изменении маркера достоверность модели увеличивалась, то данное изменение фиксировалось. Если при изменении маркера достоверность модели уменьшалась, то изменение отвергалось, т.е. у отзыва маркер оставался таким, каким был до изменения.

Наименование модели знаний и частных критерий	Выражение для частного критерия	
	через относительные частоты	через абсолютные частоты
ABS , матрица абсолютных частот, N_{ij} - фактическое число встреч i -го признака у объектов j -го класса; \bar{N}_{ij} - теоретическое число встреч i -го признака у объектов j -го класса; N_i - суммарное количество признаков в i -й строке; N_j - суммарное количество признаков или объектов обучающей выборки в j -м классе; N - суммарное количество признаков по всей выборке (таблица 1)	---	$N_i = \sum_{j=1}^M N_{ij}; N_j = \sum_{i=1}^M N_{ij}; N = \sum_{i=1}^M \sum_{j=1}^M N_{ij};$ $N_{ij} - \text{фактическая частота,}$ $\bar{N}_{ij} = \frac{N_i N_j}{N} - \text{теоретическая частота.}$
PRC1 , матрица условных P_i и безусловных P_j процентных распределений, в качестве N_j используется суммарное количество признаков по классу	---	$P_j = \frac{N_{ij}}{N_j}; P_i = \frac{N_i}{N}$
PRC2 , матрица условных P_i и безусловных P_j процентных распределений, в качестве N_j используется суммарное количество объектов обучающей выборки по классу	---	$P_j = \frac{N_{ij}}{N_j}; P_i = \frac{N_i}{N}$
INF1 , частный критерий: количество знаний по А.Харкевичу, 1-й вариант расчета вероятностей: N_j - суммарное количество признаков по j -му классу. Вероятность того, что если у объекта j -го класса обнаружен признак, то это i -й признак	$I_{ij} = \Psi \times \text{Log}_2 \frac{P_{ij}}{P_i}$	$I_{ij} = \Psi \times \text{Log}_2 \frac{N_{ij}}{N_j}; I_{ij} = \Psi \times \text{Log}_2 \frac{N_i N_j}{N_j N_j}$
INF2 , частный критерий: количество знаний по А.Харкевичу, 2-й вариант расчета вероятностей: N_j - суммарное количество объектов по j -му классу. Вероятность того, что если предьявлен объект j -го класса, то у него будет обнаружен i -й признак.	---	$I_{ij} = N_{ij} - \bar{N}_{ij} = N_{ij} - \frac{N_i N_j}{N}$
INF3 , частный критерий: Хи-квадрат : разности между фактическими и теоретическими ожидаемыми абсолютными частотами	---	$I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$
INF4 , частный критерий: ROI - Return On Investment, 1-й вариант расчета вероятностей: N_j - суммарное количество признаков по j -му классу	---	$I_{ij} = \frac{N_{ij}}{N_j} - 1 = \frac{N_i N_j}{N_j N_j} - 1$
INF5 , частный критерий: ROI - Return On Investment, 2-й вариант расчета вероятностей: N_j - суммарное количество объектов по j -му классу	---	$I_{ij} = P_{ij} - P_i$
INF6 , частный критерий: разность условной и безусловной вероятностей, 1-й вариант расчета вероятностей: N_j - суммарное количество признаков по j -му классу	---	$I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$
INF7 , частный критерий: разность условной и безусловной вероятностей, 2-й вариант расчета вероятностей: N_j - суммарное количество объектов по j -му классу	---	$I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$

Рис. 5. Различные аналитические формы частных критериев знаний, применяемые в АСК-Анализе и системе «Эйдос» [15]

Имя критерия	Имя критерия	Число объектов обучающей выборки	Число объектов тестовой выборки	Число объектов обучающей выборки	Число объектов тестовой выборки	Точность	Достоверность	Формат	Средняя нагрузка	Средняя нагрузка	Средняя нагрузка	
1. ABS - частный критерий: количество признаков по классу	Корреляция abs: частот с корр...	82	81	43	39	1	0.675	0.988	0.802	31.425	3.170	4
2. PRC1 - частный критерий: усл. вероятность того признака с корр.	Корреляция усл: частот с корр...	82	81	43	39	1	0.675	0.988	0.882	31.425	3.170	4
3. PRC2 - частный критерий: усл. вероятность того признака с корр.	Средняя усл: частот с корр...	82	82	6	76	1	0.519	1.000	0.483	24.701	9	5
4. INF1 - частный критерий: количество знаний по А.Харкевичу, 1-й вариант	Средняя усл: частот с корр...	82	81	43	39	1	0.675	0.988	0.802	31.424	3.170	4
5. INF2 - частный критерий: усл. вероятность того признака с корр.	Средняя усл: частот с корр...	82	82	6	76	1	0.519	1.000	0.483	28.121	8	4
6. INF3 - частный критерий: Хи-квадрат: разности между фактическими и теоретическими	Средняя усл: частот с корр...	82	68	82	17	14	1.000	0.829	0.987	17.483	32.914	8
7. INF4 - частный критерий: ROI - Return On Investment, 1-й вариант	Средняя усл: частот с корр...	82	82	47	35	0	0.791	1.000	0.824	19.583	2.104	0
8. INF5 - частный критерий: ROI - Return On Investment, 2-й вариант	Средняя усл: частот с корр...	82	63	82	19	1	1.000	0.748	0.969	14.942	30.954	8
9. INF6 - частный критерий: разность условной и безусловной вероятностей, 1-й вариант	Средняя усл: частот с корр...	82	80	34	48	2	0.625	0.974	0.782	15.942	1.741	2
10. INF7 - частный критерий: разность условной и безусловной вероятностей, 2-й вариант	Средняя усл: частот с корр...	82	88	88	88	1	2	0.974	0.974	24.812	34.812	8
11. INF8 - частный критерий: разность условной и безусловной вероятностей, 1-й вариант	Средняя усл: частот с корр...	82	88	88	88	1	0.974	0.974	0.974	19.248	18.248	8
12. INF9 - частный критерий: ROI (Return On Investment), 1-й вариант	Средняя усл: частот с корр...	82	44	82	16	16	1.000	0.885	0.992	18.074	22.447	8
13. INF10 - частный критерий: ROI (Return On Investment), 2-й вариант	Средняя усл: частот с корр...	82	82	44	36	1	0.875	1.000	0.838	17.388	0.848	1
14. INF11 - частный критерий: разность условной и безусловной вероятностей, 1-й вариант	Средняя усл: частот с корр...	82	62	82	20	1	1.000	0.756	0.961	16.677	30.834	8
15. INF12 - частный критерий: ROI (Return On Investment), 1-й вариант	Средняя усл: частот с корр...	82	81	33	49	1	0.623	0.988	0.764	14.384	0.619	1
16. INF13 - частный критерий: разность условной и безусловной вероятностей, 1-й вариант	Средняя усл: частот с корр...	82	71	81	1	11	0.986	0.844	0.922	19.679	16.230	0
17. INF14 - частный критерий: разность условной и безусловной вероятностей, 1-й вариант	Средняя усл: частот с корр...	82	80	48	34	2	0.782	0.974	0.838	12.711	1.837	1
18. INF15 - частный критерий: разность условной и безусловной вероятностей, 1-й вариант	Средняя усл: частот с корр...	82	67	77	5	15	0.911	0.817	0.870	15.181	13.330	0
19. INF16 - частный критерий: разность условной и безусловной вероятностей, 1-й вариант	Средняя усл: частот с корр...	82	74	34	48	8	0.687	0.902	0.725	10.846	1.170	2

Рис. 6. Максимальная достоверность полученной модели

Например, изначально отзыв, представленный на рис. 1, имел маркер «Б». После того как маркер был заменен на «А», достоверность модели по F -мере возросла. Поэтому для отзыва, представленного на рис. 1, в итоге был зафиксирован маркер «А».

Изначально отзыв, представленный на рис. 2, также имел маркер «Б». После того, как маркер был заменен на «А», достоверность

модели по F -мере уменьшилась. Поэтому для отзыва, представленного на рис. 2, в итоге был оставлен маркер «Б».

Вычисления были завершены после того, как был выполнен перебор всех найденных отзывов. Результирующая максимальная достоверность модели представлена на рис. 6.

После завершения вычислений из отзывов были сформированы два кластера. Эксперт, который ознакомился с полученными кластерами, определил, что отзывы с меткой «Б» относятся к положительным, а отзывы с меткой «А» – к отрицательным. Среди собранных 82 отзывов 62 оказались положительными и 20 – отрицательными. Таким образом, во-первых, показана работоспособность метода дихотомической классификации, во-вторых, продемонстрирована адекватность данного метода, так как застройщик ПЗСП обладает объективно высокой репутацией.

Блок-схема алгоритма кластеризации отзывов на основе метода дихотомической классификации представлена на рис. 7.

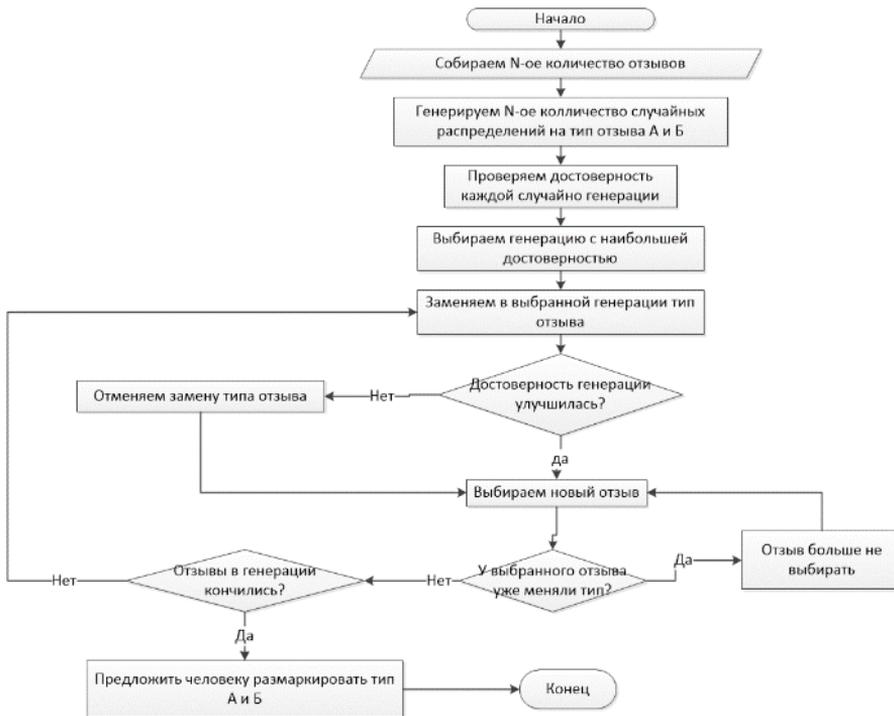


Рис. 7. Блок-схема алгоритма кластеризации отзывов на основе метода дихотомической классификации

Заключение

Использованный в настоящей работе метод дихотомической классификации позволяет автоматизировать распределение собранных отзывов о застройщике на положительные и отрицательные. В результате удается значительно увеличить скорость и масштаб анализа данных о застройщике и, как следствие, значительно повысить своевременность и точность такого анализа.

Следует отметить, что негативные отзывы выявляются лучше, чем позитивные. Данный результат объясняется тем, что, как правило, отрицательные отзывы обладают большим объемом текстовой информации, чем положительные, и к тому же содержат больший эмоциональный окрас. Следовательно, у отрицательных отзывов более плотное семантическое ядро и совпадение слов в отзывах выше, чем у положительных.

К существенным достоинствам метода дихотомической классификации относится возможность его использования при относительно небольших объемах собранной информации (несколько десятков отзывов).

Стоит отметить, что полученный в настоящей работе результат оценки деятельности застройщика ПЗСП коррелирует с результатом комплексной оценки уровня репутационного риска данного застройщика, описанным в работе [1]. Более того, использованный в настоящей работе метод дихотомической классификации позволяет уточнить оценку уровня репутационного риска застройщика.

В настоящей работе предложен алгоритм кластеризации отзывов на основе метода дихотомической классификации. Данный алгоритм может быть положен в основу программного обеспечения для сбора и анализа данных о застройщике, а также управления брендингом застройщика.

Список литературы

1. Minin A., Fedoseev S. Assessment of the Developer's Reputational Risk Level // Proceedings of 2022 4th International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency, SUMMA 2022, Virtual, Lipetsk, 9–11 November 2022. – IEEE, 2022. – P. 292–295.
2. Торкоз Д.А. Алгоритм машинного обучения широких нейронных сетей с использованием алгебры гиперразмерных двоичных векторов и

генетических алгоритмов // Южно-Сибирский научный вестник. – 2020. – № 6 (34). – С. 148–154.

3. Петрухин С.А., Глухов Г.Е., Ладыгина Н.Н. Двоичная классификация авиационных текстов с использованием нейронной сети // Научный вестник ГОСНИИ ГА. – 2021. – № 34. – С. 50–58.

4. Ораков О.Т., Сериков Е.К., Тулеген А.Т. Проблема многоклассовой классификации в машинном обучении // StudNet. – 2021. – Т. 4, № 5. – С. 1743–1748.

5. Жарова О.Ю., Чевычелов А.В. Использование методов машинного обучения для классификации вредоносного ПО // Электронный журнал: наука, техника и образование. – 2018. – № 4 (22). – С. 32–39.

6. Луценко Е.В. Синтез адаптивных интеллектуальных измерительных систем с применением АСК-анализа и системы «Эйдос» и системная идентификация в эконометрике, биометрии, экологии, педагогике, психологии и медицине // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2016. – № 116 (02). – С. 1–60.

7. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): монография (научное издание). – Краснодар: КубГАУ, 2002. – 605 с.

8. Alekseev A., Galiaskarov E., Koskova K. Application of the Matrix Rating Mechanisms and System Cognitive Analysis Methods at the Task of Residential Real Estate Conceptual Designing // Proceedings of 21st IEEE Conference on Business Informatics, CBI 2019, Moscow, 15–17 July 2019. – IEEE, 2019. – P. 111–116. – DOI: 10.1109/CBI.2019.10105.

9. Носкова А.Р., Алексеев А.О. Исследование достоверности прогнозирования банкротства при введении новой категории финансового состояния предприятий // Прикладная математика и вопросы управления. – 2020. – № 3. – С. 105–122. – DOI 10.15593/2499-9873/2020.3.06.

10. Луценко Е.В. Инвариантное относительно объёмов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергена в АСК-Анализе и системе «Эйдос» // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета – 2017. – № 126 (02). – С. 1–32. – DOI 10.21515/1990-4665-126-001.

11. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар: КубГАУ, 2014. – 600 с.

12. Оценка классификатора (точность, полнота, F-мера) [Электронный ресурс]. – URL: <http://bazhenov.me/blog/2012/07/21/classification-performance-evaluation.html> (дата обращения: 28.11.2022).

13. Как заставить работать бинарный классификатор чуточку лучше [Электронный ресурс]. – URL: <https://habr.com/ru/post/228963/> (дата обращения: 28.11.2022)

14. Гармоническое среднее [Электронный ресурс]. – URL: <http://-bazhenov.me/blog/2012/05/05/harmonic-mean.html> (дата обращения: 28.11.2022).

15. Луценко Е.В., Печурина Е.К., Сергеев А.Э. Развитый алгоритм принятия решений в интеллектуальных системах управления на основе АСК-анализа и системы «ЭЙДОС» // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2020. – № 160 (06). – С. 95–114. – DOI 10.21515/1990-4665-160-009.

References

1. Minin, A.A., Fedoseev S.A. Assessment of the Developer's Reputational. Proceedings of 2022 4nd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA 2022), 2022, pp. 292–295.

2. Torkoz D.A. Algoritm mashinnogo obucheniia shirokikh neironnykh setei s ispol'zovanie algebrы giperrazmernykh dvoichnykh vektorov i geneticheskikh algoritmov [Algorithm of machine learning wide neural networks using the algebra of hyperdimensional binary vectors and genetic algorithms]. *South-Siberian scientific bulletin*, 2020, no. 6 (34), pp. 148–154.

3. Petrukhin S.A., Glukhov G.E., Ladygina N.N. Dvoichnaia klassifikatsiia aviatsionnykh tekstov s ispol'zovaniem neironnoi seti [Binary classification of aviation texts using a neural network]. *Scientific bulletin of the state scientific research institute of civil aviation*, 2021, no. 34, pp. 50–58.

4. Ораков О., Сериков Е., Тулеген А. Problema mnogoklassovoi klassifikatsii v mashinnom obuchenii [The problem of multiclass classification in machine learning]. *StudNet*. 2021, vol. 4, no 5, pp. 1743–1748.

5. Zharova O.Iu., Chevychelov A.V., Ispol'zovanie metodov mashinnogo obucheniia dlia klassifikatsii vredonosnogo PO [Use of machine training methods for malware classification]. *Elektronnyi zhurnal: nauka, tekhnika i obrazovanie*, 2018, no. 4 (22), pp. 32–39.

6. Lutsenko, E.V. Sintez adaptivnykh intellektual'nykh izmeritel'nykh sistem s primeneniem ASK-analiza i sistemy «Eidos» i sistemaia identifikatsiia v ekonometrike, biometrii, ekologii, pedagogike, psikhologii i meditsine [Adaptive synthesis of intelligent measurement systems with the use of ASC-analysis and “Eidos” system. System identification in econometrics, biometrics, ecology, pedagogy, psychology and medicine]. *Politematicheskii setevoi elektronnyi nauchnyi zhurnal*

Kubanskogo gosudarstvennogo agrarnogo universiteta, 2016, no. 116(02), pp. 1–60.

7. Lutsenko, E.V. Avtomatizirovannyi sistemno-kognitivnyi analiz v upravlenii aktivnymi ob"ektami (sistemnaia teoriia informatsii i ee primenenie v issledovanii ekonomicheskikh, sotsial'no-psikhologicheskikh, tekhnologicheskikh i organizatsionno-tekhnicheskikh sistem) [Automated system-cognitive analysis in the control of active objects (system theory of information and its application in the study of economic, socio-psychological, technological and organizational-technical systems)]. Krasnodar, KubGAU, 2002, 605 p.

8. Alekseev A., Galiaskarov E., Koskova K. Application of the Matrix Rating Mechanisms and System Cognitive Analysis Methods at the Task of Residential Real Estate Conceptual Designing. *Proceedings of 21st IEEE Conference on Business Informatics*, 2019, pp. 111–116. DOI: 10.1109/CBI.2019.10105.

9. Noskova A.R., Alekseev A.O., Issledovanie dostovernosti prognozirovaniia bankrotstva pri vvedenii novoi kategorii finansovogo sostoianiia predpriatii [Research of the bankruptcy prediction reliability in case to introduce a new category of the companies financial position]. *Applied Mathematics and Control Sciences*. 2020, no.3, pp.105-122. DOI 10.15593/2499-9873/2020.3.06.

10. Lutsenko E.V. Invariantnoe otnositel'no ob"emov dannykh nechetkoe mul'tiklassovoe obobshchenie F-mery dostovernosti modelei Van Rizbergena v ASK-analize i sisteme «Eidos» [Invariant to volumes of data, a fuzzy multiclass generalization of F-measure of plausibility in Van Rijsbergen models in ASC-analysis and in the “Eidos” system]. *Politematicheskii setevoi elektronnyi nauchnyi zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta*, 2017, no. 126 (02), pp. 1–32. DOI 10.21515/1990-4665-126-001.

11. Orlov A.I., Lucenko E.V. Sistemnaja nechetkaja interval'naja matematika [System Fuzzy Interval Mathematics]. Krasnodar, KubGAU, 2014, 600 p.

12. Otsenka klassifikatora (tochnost', polnota, F-mera) [Classifier score (accuracy, recall, F-measure)], OpenMP, available at: <http://bazhenov.me/blog/2012/07/21/classification-performance-evaluation.html> (accessed 28.11.2022)

13. Kak zastavit' rabotat' binarnyi klassifikator chutochku luchshe [How to make the binary classifier work a little better], OpenMP, available at: <https://habr.com/ru/post/228963/> (accessed 28.11.2022)

14. Garmonicheskoe srednee [Harmonic mean]. OpenMP, available at: <http://bazhenov.me/blog/2012/05/05/harmonic-mean.html> (accessed 28.11.2022)

15. Lutsenko E.V., Pechurina E. K., Sergeev A. E. Razvityi algoritm priniatiia reshenii v intellektual'nykh sistemakh upravleniia na osnove ASK-analiza i sistemy «EIDOS» [A developed decision-making algorithm in intelligent control systems based on the ASC-analysis and the “Eidos” system]. *Politematicheskii setevoi elektronnyi nauchnyi zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta*, 2020, no. 160 (06), pp. 95–114. DOI 10.21515/1990-4665-160-009.

Сведения об авторах

Минин Антон Андреевич (Пермь, Россия) – аспирант кафедры «Вычислительная математика, механика и биомеханика» Пермского национального исследовательского политехнического университета (614990, г. Пермь, Комсомольский проспект, 29, e-mail: minin.anton96@yandex.ru)

Федосеев Сергей Анатольевич (Пермь, Россия) – доктор технических наук, доцент, профессор кафедры «Вычислительная математика, механика и биомеханика» Пермского национального исследовательского политехнического университета (614990, г. Пермь, Комсомольский проспект, 29, e-mail: fsa@gelicon.biz)

About the authors

Anton A. Minin (Perm, Russian Federation) – Ph. D. student, Department of Computational Mathematics, Mechanics and Biomechanics, Perm National Research Polytechnic University (29, Komsomolsky ave., Perm, 614990, e-mail: minin.anton96@yandex.ru)

Sergey A. Fedoseev (Perm, Russian Federation) – Dr. Habil. in Engineering, Docent, Professor, Department of Computational Mathematics, Mechanics and Biomechanics, Perm National Research Polytechnic University (29, Komsomolsky ave., Perm, 614990, e-mail: fsa@gelicon.biz)

Библиографическое описание статьи согласно ГОСТ Р 7.0.100–2018:

Минин, А.А. Применение метода случайной дихотомической классификации для оценки застройщика по отзывам потребителей / А. А. Минин, С. А. Федосеев. – текст : непосредственный. – DOI: 10.15593/2499-9873/2022.4.09 // Прикладная математика и вопросы управления / Applied Mathematics and Control Sciences. – 2022. – № 4. – С. 159–173.

Цитирование статьи в изданиях РИНЦ:

Минин, А.А. Применение метода случайной дихотомической классификации для оценки застройщика по отзывам потребителей / А. А. Минин, С. А. Федосеев // Прикладная математика и вопросы управления. – 2022. – № 4. – Р. 159–173. – DOI: 10.15593/2499-9873/2022.4.09

Цитирование статьи в references и международных изданиях

Cite this article as:

Minin A.A., Fedoseev S.A. Application of the random dichotomous classification method to assessment builder's brand based on consumer's reviews. *Applied Mathematics and*

Control Sciences, 2022, no. 4, pp. 159–173. DOI: 10.15593/2499-9873/2022.4.09 (in Russian)

Финансирование. Исследование не имело спонсорской поддержки.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Вклад авторов равноценен.

Поступила: 01.12.2022

Одобрена: 12.12.2022

Принята к публикации: 12.12.2022