

ЛІП'ЯНИНА-ГОНЧАРЕНКО Х. В.Західноукраїнський національний університет
ORCID ID: 0000-0002-2441-6292
e-mail: xrustya.com@gmail.com**КОМАР М. П.**Західноукраїнський національний університет
ORCID ID: 0000-0001-6541-0359
e-mail: mko@wunu.edu.ua**ЛЕНДЮК Т. В.**Західноукраїнський національний університет
ORCID ID: 0000-0001-9484-8333
e-mail: tl@wunu.edu.ua**ГРАМЯК Р. М.**Західноукраїнський національний університет
ORCID ID: 0000-0001-8698-0377
e-mail: fear3171@gmail.com

МЕТОД ВИБОРУ КОНКУРЕНТНОГО ТОВАРУ НА ОСНОВІ ЕМОЦІЙНОГО ЗАБАРВЛЕННЯ ВІДГУКІВ

У статті розроблено інтелектуальний метод вибору конкурентного товару на основі емоційного забарвлення відгуків, що ділиться на три блоки: парсер відгуків, визначення емоційного забарвлення та класифікація відгуків.

Ключові слова: товар; відгуки; парсер; емоційне забарвлення тексту; класифікація; машинне навчання.

KHRYSTYNA LIPIANINA-HONCHARENKO, MYROSLAV KOMAR,
TARAS LENDYUK, ROMAN GRAMYAK
Western Ukrainian National University

METHOD OF CHOOSING A COMPETITIVE PRODUCT BASED ON THE EMOTIONAL COLOR OF THE CALLS

Other people's opinions have always been an important piece of information for most of us in the decision-making process. The interest shown by users to online reviews and comments, as well as the potential impact of these comments on issues in discourse and decision-making, make them pay attention to this aspect of online activity. Finding the best products for sale is one of the most important stages in the process of creating a profitable company. That is why the choice of goods for an online store should be carried out deliberately, taking into account both the capabilities and analysis of prospects in the niche, as well as a number of other important parameters. One of the methods of choosing a competitive product may be the analysis of goods in marketplaces based on the emotional color of the calls. Product feedback research is an extremely popular topic, which is confirmed by the analysis of studies. Calls can be constantly reread, but when there are many goods in one segment, because there are more manufacturers, it is laborious. Therefore, the development of technology that will be able to automate this process is necessary for business sales. The article developed an intelligent method of choosing a competitive product based on the emotional color of the calls, which is divided into three blocks: a feedback parser, the definition of emotional coloring and the classification of calls. The findings will help retailers manage their websites wisely and help customers make product purchase decisions. In the next scientific researches, the implementation of the method will be carried out on the data of the Ukrainian site Rozetka. The classification of the most classical methods of classification based on machine learning will be carried out, namely Support Vector Classifier, Stochastic Gradient Descent Classifier, Random Forest Classifier, Decision Tree Classifier, Gaussian Naive Bayes, K-Neighbors Classifier, Ada Boost Classifier, Logistic Regression.

Keywords: product; reviews; parser; emotional coloring of the text; classification; machine learning.

Постановка проблеми

Думки інших людей завжди були важливою частиною інформації для більшості із нас в процесі прийняття рішень. Інтерес, який демонструють користувачі до онлайн відгуків та коментарів, а також потенціальний вплив цих коментарів на питання у дискурсі і прийняття рішень змушують звернути увагу на цей аспект. У зв'язку з цим можна вважати, що розробка методу вибору конкурентного товару з врахуванням емоційного настрою є одним із найбільш перспективних напрямків у інтернет-торгівлі.

Аналіз останніх джерел

Найбільш точними у визначенні емоційного настрою відгуку є моделі [1–3] засновані на машинному навчанні, глибокому навчанні [4, 5] та загортковій нейронній мережі [6, 7]. Представлено [8] модель для визначення корисності коментарів за допомогою текстових функцій, вилучених з оглядів, на основі різних типів функцій, включаючи емоційні, мовні та текстові особливості, значення валентності, збудження та домінування, тривалість огляду та полярність коментарів. Представлено [9] загальну структуру, яка використовує методи обробки природних мов (NLP), включаючи аналіз настроїв, аналіз текстових даних та методи кластеризації, щоб отримати нові оцінки на основі споживчих настроїв для різних характеристик товару. Проаналізовано [10] емоційний зміст великої кількості онлайн-оглядів для продуктів із використанням методів обробки натуральної мови (NLP). Запропоновано [11] експериментальне дослідження SentiME, підхід до вилучення полярності настрою повідомлення, на корпусі на основі огляду Amazon.

У зв'язку із цим, метою статті є розробка методу вибору конкурентного товару на основі емоційного забарвлення відгуків.

На відміну від аналогів [8–11] розроблений метод вибору конкурентного товару на основі емоційного забарвлення відгуків дозволить проводити парсер відповідних даних з цільового сайту. Також класифікація буде протестована вісьмома класичними методами класифікації, а саме Support Vector Classifier, Stochastic Gradient Decent Classifier, Random Forest Classifier, Decision Tree Classifier, Gaussian Naive Bayes, K-Neighbors Classifier, Ada Boost Classifier, Logistic Regression. На основі отриманого методу продавець може приймати управлінські рішення стосовно вигідного вкладання коштів в новий товар.

Виклад основного матеріалу

Для зменшення часових витрат на вибір популярного нового товару авторами розроблено метод вибору конкурентного товару на основі емоційного забарвлення відгуків. Запропонований метод представлений наступними кроками:

Крок 1. Підключаємо необхідні бібліотеки. Вони надають загальні математичні і числові операції у вигляді пре-скомпільованих, швидких функцій. Вони об'єднуються в високорівневі пакети.

Крок 2. Вибір потрібного сайту для подальшої роботи.

Крок 3. Введення посилання на категорії товару, які нам необхідні.

Крок 4. Виконується парсинг відгуків користувачів на товари. Ділимо їх по id товару і додаємо в текстовий файл. На кожен рядок окремий відгук.

Крок 5. Створення БД зібраних разом відгуків.

Крок 6. Підключення відгуків з БД для подальшої роботи.

Крок 7. Необроблений текст досить безладний для цих оглядів, тому потрібно зробити очистку тексту, щоб алгоритму було легше розпізнавати настрій.

Крок 8. Проводиться векторизація. Векторизація – вид розпаралелювання програми, при якому однопоточні додатки, які виконують одну операцію в кожен момент часу, модифікуються для виконання декількох однотипних операцій одночасно. Для виконання процесу векторизації потрібно провести токенизацію, ігнорування поодиноких символів, перетворення тексту в числові вектори та n-грами після чого формуємо масив значень.

Крок 8.1. Токенизація потрібна для заміни конфіденційного елемента даних на неконфіденційний еквівалент, званий токеном, який не має самостійного сенсу/значення для зовнішнього або внутрішнього використання.

Крок 8.2. Процес ігнорування поодиноких символів, які мішають роботі програми проводиться на основі готового набору «стоп-слів». Стоп-слова – це дуже поширені слова, такі як «якщо», «але», «ми», «він», «вона» і «вони».

Крок 8.3. На цьому кроці проводимо перетворення тексту в числові вектори та додаємо n-грами. N-грами – це просто послідовність з n елементів (звуків, складів, слів або символів), що йдуть в якомусь тексті поспіль.

Крок 8.4. Формування масиву векторизованих значень тексту для подальшого зберігання в БД.

Крок 9. Формування БД перетвореного тексту для вибору моделі класифікатора.

Крок 10. Введення очищених даних.

Крок 11. Наступним кроком є вибір найоптимальнішого класифікатора алгоритмами машинного навчання. Для цього потрібно використати, перехресну перевірку вісьмома класичними методами класифікації: Support Vector Classifier, Stochastic Gradient Decent Classifier, Random Forest Classifier, Decision Tree Classifier, Gaussian Naive Bayes, K-Neighbors Classifier, Ada Boost Classifier, Logistic Regression [12].

Крок 12. Далі проводиться класифікація даних. На позитивні та негативні коментарі. А саме позитивні відгуки діляться на категорії: відмінно, ідеально, чудово, дивовижно та чудово. Негативні діляться на категорії: найгірший, відходи, жахливо, погано та нудно.

Крок 13. Останнім кроком є виведення отриманого результату на основі, якого можна приймати рішення стосовно вигідного вкладання коштів в новий товар.

Висновки

Розроблено метод вибору конкурентного товару на основі емоційного забарвлення відгуків, на основі якого продавець може приймати управлінські рішення стосовно вигідного вкладання коштів в новий товар, і, тим самим, зменшить ризики від неприбуткових операцій купівлі-продажу. Також запропонований метод дозволяє зменшити затрати часу на пошук популярних та якісних товарів на основі відгуків користувачів.

Література

1. Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., & Roy, P. K. (2017). Predicting the “helpfulness” of online consumer reviews. *Journal of Business Research*, 70, 346-355.
2. Samal, B., Behera, A. K., & Panda, M. (2017). Performance analysis of supervised machine learning techniques for sentiment analysis. 2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS). DOI: 10.1109/ssps.2017.8071579
3. Sun, Xinyu; Han, Maoxin; Feng, Juan (2019). Helpfulness of online reviews: Examining review informativeness and classification thresholds by search products and experience products. *Decision Support Systems*, 113099. DOI: 10.1016/j.dss.2019.113099

4. Khedkar S., Shinde S. (2020) Deep Learning-Based Approach to Classify Praises or Complaints from Customer Reviews. In: Bhalla S., Kwan P., Bedekar M., Phalnikar R., Sirsikar S. (eds) *Proceeding of International Conference on Computational Science and Applications. Algorithms for Intelligent Systems*. Springer, Singapore, 391-402. https://doi.org/10.1007/978-981-15-0790-8_38
5. Wu, P., Li, X., Shen, S., & He, D. (2020). Social media opinion summarization using emotion cognition and convolutional neural networks. *International Journal of Information Management*, 51, 101978
6. Smetanin, S., & Komarov, M. (2019). Sentiment Analysis of Product Reviews in Russian using Convolutional Neural Networks. 2019 IEEE 21st Conference on Business Informatics (CBI), 01, 482-486.
7. Saumya, S., Singh, J.P. & Dwivedi, Y.K. (2020) Predicting the helpfulness score of online reviews using convolutional neural network. *Soft Comput* 24, 10989–11005. <https://doi.org/10.1007/s00500-019-03851-5>
8. F. Fouladfar, M. N. Dehkordi and M. E. Basiri, (2020). Predicting the Helpfulness Score of Product Reviews Using an Evidential Score Fusion Method, in *IEEE Access*, vol. 8, pp. 82662-82687. DOI: 10.1109/ACCESS.2020.2988872.
9. Kauffmann, Peral, Gil, Ferrández, Sellers, Mora, (2019). Managing Marketing Decision-Making with Sentiment Analysis: An Evaluation of the Main Product Features Using Text Data Mining. *Sustainability*, 11(15), 4235. DOI: 10.3390/su11154235
10. Ullah, R., Amblee, N., Kim, W., & Lee, H. (2016). From valence to emotions: Exploring the distribution of emotions in online product reviews. *Decision Support Systems*, 81, 41-53.
11. Sygkounas E., Rizzo G., Troncy R. (2016) Sentiment Polarity Detection from Amazon Reviews: An Experimental Study. In: Sack H., Dietze S., Tordai A., Lange C. (eds) *Semantic Web Challenges. SemWebEval 2016. Communications in Computer and Information Science*, vol 641. Springer, Cham. 108-120. https://doi.org/10.1007/978-3-319-46565-4_8
12. Lipyana H., Maksymovych V., Sachenko A., Lendyuk T., Fomenko A., Kit I. (2020) Assessing the Investment Risk of Virtual IT Company Based on Machine Learning. In: Babichev S., Peleshko D., Vynokurova O. (eds) *Data Stream Mining & Processing. DSMP 2020. Communications in Computer and Information Science*, vol 1158, 167-187 Springer, Cham. https://doi.org/10.1007/978-3-030-61656-4_11
13. Ullah, R., Amblee, N., Kim, W., & Lee, H. (2016). From valence to emotions: Exploring the distribution of emotions in online product reviews. *Decision Support Systems*, 81, 41-53.
14. Sygkounas E., Rizzo G., Troncy R. (2016) Sentiment Polarity Detection from Amazon Reviews: An Experimental Study. In: Sack H., Dietze S., Tordai A., Lange C. (eds) *Semantic Web Challenges. SemWebEval 2016. Communications in Computer and Information Science*, vol 641. Springer, Cham. 108-120. https://doi.org/10.1007/978-3-319-46565-4_8
15. Lipyana H., Maksymovych V., Sachenko A., Lendyuk T., Fomenko A., Kit I. (2020) Assessing the Investment Risk of Virtual IT Company Based on Machine Learning. In: Babichev S., Peleshko D., Vynokurova O. (eds) *Data Stream Mining & Processing. DSMP 2020. Communications in Computer and Information Science*, vol 1158, 167-187 Springer, Cham. https://doi.org/10.1007/978-3-030-61656-4_11

Рецензія/Peer review : 23.12.2021

Надрукована/Printed :30.12.2021