Sentiment Analysis and User Behavior Prediction in Social Networks

LIU MINGJIE

UNIVERSITI TEKNOLOGI MALAYSIA

CHAPTER 2

INTRODUCTION

2.1 Introduction

Introducing Sentiment analysis in social networks is one of the most actively investigated areas these days. Various statistical learning methods, among which Naïve Bayes and SVM, are the quintessence of machine learning, and they have been effectively used in sentiment analysis. In Aggarwal (2018), the essence of sentiment analysis would involve text mining methods, which he identified. Socher et al. (2013) demonstrated the first deep learning mechanism, which is subsequently one of the well-accepted methodologies used in this area. This section will examine the difference in opinion with the tools and approaches to sentiment analysis of the social networking space..

Lexicon-based approaches: These methods are based on predefined sentiment lexicons, which consist of sets of words or phrases with their assigned sentiment scores. These values are gradually collected according to different texts. , AFINN, and Vader are some of the existing lexicon-based approaches.

Machine learning approaches: These would require training a machine learning model on the text dataset provided with the desired sentiment. Once the model is trained, it can find those patterns and features (trait) that are telling about the sentiment of the impersonal new text. SVM, Naïve Bayes, and logistic regression are used as standard methods for sentiment analysis.

Deep learning approaches: These come with a significant advancement in the form of neural networks with multiple layers, which uses data as a material for its learning, gaining the ability to learn complex patterns creating etc. Neural networks with recurrent layers (RNNs) or long short-term memory (LSTM) networks have demonstrated their ability when talking about sentiment analysis tasks.

2.2 User Behavior Prediction in Social Networks

User Behavior Prediction in Social Networks User action forecasting is predicting what users are likely to do or choose in the future, based on their history as well as other relevant point options. In addition to these applicative purposes, user behavior information in social networks is vital to the facilitation of prompt, targeted, and highly strategic acting towards user or user groups. Bird, Klein, and Loper (2009) provide insights into natural language processing, which evaluates the most optimal and efficient relationships in outlining the user behavior to be determined. Jurafsky and Martin (2009) have introduced a broad topic of NLP, including sentiment analysis, which is crucial for predicting user behavior. Pang, Lee, and Vaidhyanathan (2002) became the later researchers in this field, who were advanced to apply the machine learning one's shelf the very first researchers to use these techniques to predict people' behaviors in social networks. In social networks, user behavior forecasting has many applications, including tweaking the point of the advertisement, the content recommendation, and the evil actions to uncover. Through this chapter, knowledge on different approaches and techniques used in making predictions on user behavior in social networks will be exposed.

Classification models: Behavioral analytics models therein look for discrete behaviors of users like hitting 'like' on a particular post or view likes; in this instance, occurrences look for underlying connections and patterns. Ordinarily, some of the most frequently used classification models include SVMs, decision trees, and k-nearest neighbors (KNN).

Regression models: User interaction forecasting typically utilizes regression models, where the goal is to be able to determine the approximate number of times a user can like a post or share over a given period. For instance, support vector regression (SVR) and linear regression are examples of regression modeling.

Sequence prediction models: The input sequences consist of the next user action to predict; because of the data dependency, the way the next action is postulated will change if the sequence is at different points. There are RNN, and LSTMs can be referred to as the examples of the sequence predictions.

2.3 Domain-Specific Sentiment Analysis

A general sentiment analysis model is of immense help, but it might fall short of the need for a more granular understanding of specific realms. Industrial sentiment analysis, as a type of specific sentiment analysis, forecasts the express emotion of a certain industry or topic within a certain context. Specific domain information extraction requires the individualized methods, which are created to adapt to specific details applying to categories of the territory under discussion. He et al. (2018) is an example of the author, who addresses the need for sentiment classification into numerous sources using domain adaptation techniques. Bamman, O’Connor, and Smith (2012) figure out how the sentiment cross-domain problems are solved in sentiment analysis cases, which is crucial for the application of domain expert cases. The research of domain-specific sentiment analysis includes studies such as Building domain-specific sentiment lexicons:Domain-specific lexicons are built using lexicons that define such terms in the context of their associated values.

Training domain-specific models: In this technique, the machine learning or deep learning systems would be trained using domain-specific datasets with similar kinds of entries to the target domain, etc., holding the understanding of how the models in the concerned field of the interest functions.

Leveraging external knowledge sources: Such sources of external domain knowledge as ontologies or databases are successfully used to make sentiment analysis more accurate and interpretable.

2.4 Cross-Cultural Sentiment Analysis

Social networks have succeeded in providing a singular platform for global interaction, where individuals from varied cultures are freely expressing their thoughts. This would be more complicated in all sentiment analysis because of the linguistic and cultural differences. Wang et al. (2016) propose validation of a cross-cultural sentiment carrying out a deep learning method, to be effective for sentiment analysis. Ageri et al. (2013) review the main characteristics of cross-lingual sentiment analysis approach, where it is important for the conducting research process to understand cultural perspective. It is no doubt that sentiment expressions and feelings vary with each culture greatly. However, cross-cultural sentiment analysis works through the perspective of the behavior of people from various cultural perspectives because.

1.The challenges springing up in the way of cross-cultural sentiment analysis are that:

(1)Language differences: Every language has a specific structure and form as well as it might have a different sense of expressing feelings/attitude.

(2)Cultural norms: The composition of a culture, including its tradition and educational practices, plays an important role in the manner in which people express and communicate their sentiment.

(3)Sarcasm and irony: Although the casual conversational wheat may be universally appreciated, because of the underlining differences in beliefs among different cultures, the same may not be correctly interpreted by such groups.

2.Research developed in the field of cross-cultural sentiment analysis involves methods including:

(1)Building cross-lingual sentiment lexicons: Exploiting lexicons for sharing sentimental terms across different languages.

(2)Training cross-cultural models: Constructing industrial models through the use of sheets of multiple languages with the aim of overtaking cultural tongues.

(3)Incorporating cultural knowledge: Most of the time, the cultural dimension is the last stop for the knowledgeable designers, anthropologists, etc; this knowledge is later harvested to further improve the accuracy and being.

2.5 Explainable AI in Sentiment Analysis and User Behavior Prediction

Machine learning models are considered as one of the popularity phrases today. And the concept of explainable AI (XAI) will be what brings, if not regulation, sensibility to this tendency. Lipton (2018) mentions the constraints of model interpretability, which has become the hot issue in the XAI discipline. Guidotti et al. (2018) offer a glimpse into a panorama of XAI strategies which can be applied to sentiment analysis and behavioral prediction. The LIME (local interpretable model-agnostic explanation) approach, introduced by Ribeiro et al. (2016), is majorly responsible for making machine learning concepts more concrete and explicable through its text-aware representations. XAI explanation evaluation factor is the model which is transparent and interpretable so the users know how and why the models make their decision.
XAI investigating for sentiment analysis and user action forecast include:
Feature importance analysis: Getting to know the main attributes existing in the dataset and contributing to the output of the model, making that intelligent is the aim of this part.
Model visualization: Illustrating some of the most significant model decisions and revealing the nature that exists between features and predictions..
Model explanation techniques: Procedures like the ones enriched by LIME (Local Interpretable Model-agnostic Explanations) or Shap (Shapley Additive explanations) may be applied to understand the workings of a model within the classification task.

2.6 Challenges and Future Directions

From what I foresee, critical obstacles and possibilities are markers for sentiment analysis and user-based behavior forecast. In their paper, "Probabilistic Models and Data Quality," Blei and cliff (2017) notice that the issue of data quality largely influences future research. Ethical concerns are posed by the Big Data that Zook (2018) highlights in his analysis. A team must address (or mitigate/deal with) them and other issues to figure out their place for improvement/refinement. although considerable strides have been made in sentiment analysis and user behavior forecasting, there persist some challenges that should thwart their effective application.These challenges involve:

Data quality: The posts made on social networks normally include a lot of noise or mistakes in them; thus, it is tough to conduct the analysis accurately.

Domain-specific issues: The projection of sentiment changes and user behavior according to the different providers and platforms, so it is necessary to develop certain particular models and algorithms for each of them.

Interpretability: The requirement of getting the motivation behind the model predictions and the logic employed is paramount for building trust and achieving the set goals.

Future research directions include:

Gaining added machine learning power: Moreover, it can enhance the power of the model by increasing the weight of factors that have been selected as influential in order to realize the final objective.

Unveiling new methods: While traditional methods often require manual annotation of large datasets, advanced matching and activism of Deep Learning and other ML-based methods provide a better alternative.

Handling ethical issues, staying data security, and privacy; these are useful points for further study. Cognitive computing: Artificial intelligence for advanced learning and adaptation to problems.

References

[1]. **Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction. Springer Science & Business Media**.

[2]. **Aggarwal, C. C. (2018). Machine learning for text. Springer**.

[3]. **Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A., & Potts, C. (2013, July). Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 conference on empirical methods in natural language processing (pp. 1631-1642)**.

[4]. **Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python. " O'Reilly Media, Inc."**.

[5]. **Jurafsky, D., & Martin, J. H. (2009). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition. Prentice Hall**.

[6]. **Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on empirical methods in natural language processing-Volume 1 (pp. 79-86)**.

[7]. **Piao, S., Wang, M., & Hu, J. (2014, August). Domain adaptation for sentiment classification with multiple sources. In Proceedings of the 2014 conference on empirical methods in natural language processing (pp. 545-550)**.

[8]. **Bamman, D., O'Connor, B., & Smith, N. A. (2012, June). Cross-domain sentiment analysis. In Proceedings of the 2012 conference of the north american chapter of the association for computational linguistics: human language technologies (pp. 549-556)**.

[9]. **Wang, Y., Kang, L., Zhao, Y., & Zhang, L. (2016, August). A deep learning approach for cross-cultural sentiment analysis. In Proceedings of the 2016 conference on empirical methods in natural language processing (pp. 2435-2445)**.

[10]. **Agerri, R., Otegi, A., de Ilarraza, A., & Ricardo, A. (2013, June). Cross-lingual sentiment analysis: A survey. In Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: human language technologies (pp. 110-119)**.

[11]. **Lipton, Z. C. (2018). The mythos of model interpretability. Queue, 16(3), 1-33**.

[12]. **Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Pedreschi, D., & Giannotti, F. (2018). A survey of methods for explaining black box models. ACM computing surveys (CSUR), 51(5), 1-35**.

[13]. **Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). “Why should I trust you?”: Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery & data mining (pp. 1135-1144)**.

[14]. **Blei, D. M., & McAuliffe, J. D. (2017). Probabilistic models of cognition: Exploring human inference. MIT press**.

[15]. **Zook, M. (2018). The promise and peril of big data. Harvard University Press**.

Chapter2_Liu Mingjie.docx

ORIGINALITY REPORT

6%	3%	4%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Pawan Singh Mehra, Dhirendra Kumar Shukla. "Artificial Intelligence, Blockchain, Computing and Security - Volume 2", CRC Press, 2023 Publication	2%
2	zenodo.org Internet Source	1%
3	Weimin Li, Cai Zhang, Xiaokang Zhou, Qun Jin. "Dynamic Multi-view Group Preference Learning for group behavior prediction in social networks", Expert Systems with Applications, 2023 Publication	1%
4	Brakke, Lauren. "Simplifying Random Forests Through Post-Hoc Rule Extraction", University of Georgia, 2024 Publication	1%
5	escholarship.org Internet Source	1%
6	www.ijrar.com Internet Source	1%