

A STUDY ON THE INFLUENCE OF BIAS AND NATURAL LANGUAGE PROCESSING IN DISCRETE DOMAINS

Arun Padmanabhan, K. Devasenapathy*

Abstract

Bias influences the mindset of a person in handling situations. Bias dominates the thinking process of a person while the decision process. Due to this importance of the subject concerned, this paper reviewed the influence of bias in various domains and its effects. Natural language processing is a branch of Artificial Intelligence that makes machine to understand what humans instruct to do. Sentiment analysis and opinion mining are two different field of study in natural language processing which analyze the opinion of the people, sentiments, behaviours, attitude from text. Both can be applied when combined with some powerful machine language algorithm which is able to achieve any complex tasks. Linear regression, logistic regression, Decision tree, Cross Vector Machine algorithm, Naïve Bayes and Random forest are some among many machine learning algorithms used by the researchers. Hence this review examines the application of sentiment analysis and opinion mining in discrete domains.

Keywords: Bias; Sentiment Analysis; Opinion Mining; Confirmation Bias.

I INTRODUCTION

Characteristics that correspond to the position of the Sun and other astrological planets at the time of the birth is Astrology. Astrology exists in a variety of ways, depends on the culture of the developing person. Astrology can greatly influence and validate a person's concept of self belief. Astrologers and those who believe in it adapt and incorporate the descriptions that suit them the best. Reviewing a person's

Department of Computer Science,
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India
*Corresponding Author

natal chart reveals that person's personal astrological signs. Natal chart outlines the position of all the astronomical planets in the solar system at the time of the person's birth. Astrologers use this natal chart having clear description to relate a person's characteristics or circumstances. The influence of bias in astronomy is enormous.

Bias influences the mindset to choose one object, thing or person to another and to favor's that person or thing. It also means a strong inclination of the mind or a preconceived opinion about something or someone. Influence of bias affects the accuracy of judgment on any field. This review paper deals with the concept of influence of bias in astrology. Also the influence of confirmation bias in discrete domains is included.

II BIAS IN ASTROLOGY

Andersson, Ida, Julia Persson, and Petri Kajonius (2022) [1] suggested that astrology will be able to attract and strengthen the personal characteristics. Using an online survey among 264 random persons, data are collected from the social media for Belief in Astrology, Big Five Personality Traits, Narcissism and Intelligence. By performing multiple linear regressions for data analysis, Narcissism showed positive relationship whereas Intelligence showed a negative relationship in astrological belief.

Bozkurt, Gözde, and Ahmet (2021) [2] found that personal characteristics and qualities are influenced by beliefs existing before the scientific inventions. The influence of astrological personality on personal traits was measured using the Five Factor Personality Model. Through personality inventory, same or different individual characteristics may be determined. Characteristics of an individual are developed during the birth and in progress of

living, the influence of society shaped the development are also determined using astrology.

Helgertz, Jonas, and Kirk Scott. (2020) [3] observed that among married couples, the zodiac sign combination and the chances of getting divorce are related to each other are the biased influenced effect of the individuals knowledge and belief in astrology in marriage based on data taken from Sweden during the period 1968 to 2001 using individual data based on longitudinal values.

III INFLUENCE OF CONFIRMATION BIAS IN DISCRETE DOMAINS

Cafferata, Alessia, Marwil J. Dávila-Fernández, and Serena Sord (2021) [4] opined that confirmation bias is the tendency to influence, interpret and support the own self belief over real facts. The tendency to favor one's own self belief deteriorates the easiness in handling the environmental challenge. In human reasoning, confirmation bias is the most severely influenced bias. Kappes, A., Harvey, A. H., Lohrenz, T., Montague, P. R., & Sharot, T. (2020) [5] opined that the individual may likely to change the opinions during discord due to the influence of existing judgments. The existing judgments influence the decision pattern of the person as well as it affects the thinking ability of the brain. Millner et al. (2020) Millner, A., Ollivier, H., & Simon, L. (2020) [6] have narrated how the influence of confirmation bias affects political outcomes. The study explored that the presence of behavioural bias of the voters maintains the equilibrium welfare among the candidates in election. Decision in judgment was largely based on the biased belief of the candidate as well as the voter. Voters tend to accept the information rather than confirming whether it is genuine or not which is due to the existence of confirmation belief in thoughts. Cafferata, Alessia, and Fabio Tramontana (2019) [7] studied that some features of time series of financial market may be influenced by confirmation bias with the help of Monte Carlo method. For the study, heterogeneous agents

were used in a financial market model. Study showed that four different groups of traders coexist and some traders were affected by the influence of confirmation bias. Also, noticed that the biased belief has caused a very large disparity among the belief of the traders.

Gan, Lirong, Huamao Wang, and Zhaojun Yang (2020) [8] predicted the Asian option prices effectively using deep learning method by conducting experiment on three different sets of data set. 10,000 Asian option prices in 1 second were computed using the trained deep learning model. The efficiency and effectiveness of the data are examined using the numerical and empirical test. In order to train the deep learning model, real data of West Texas Intermediate Average Price Options, so that more accurate results are produced in the experiments conducted using numerical tests of the other. Industrial experts and financial advisors found that this prediction is helpful. While using real data, the biased prediction has a median of 0.8 % and a mean of 95%. Wang, A., Xu, J., Tu, R., Saleh, M., & Hatzopoulou, M. (2020) [9] used Land use regression and various machine learning algorithms such as Artificial Neural Networks and gradient boost, studied the dimensional distribution of air pollution by taking black carbon and fine residue from Toronto and Canada with the help of a dynamic camera to record the real time traffic. The study observed that the approaches handled by machine learning algorithms are superior to the Land use regression.

IV OVERVIEW OF SENTIMENT ANALYSIS

Pavitha, N., Pungliya, V., Raut, A., Bhonsle, R., Purohit, A., Patel, A., & Shashidhar, R. (2022) [10] described a method for movie recommendation using Sentiment Analysis based on the reviews of movie selected with the help of Naïve Bayes Classifier and Support Vector Machine Classifier. Accuracy, F1 Score, Precision and recall were taken as the parameters to be used for comparison between

the two supervised machine learning algorithms. Accuracy level obtained using Support Vector Machine Classifier outweighs the same that using Naïve Bayes Classifier. Jardim, Sandra, and Carlos Mora (2022) [11] proposed a method to inspect the mindset of the users based on the reviews and comments taken from digital tourism platform using sentiment analysis. Using the clustering algorithm, the developed method showed a high accuracy in segmenting the users based on their interests and needs.

Li, H., Chen, Q., Zhong, Z., Gong, R., & Han, G. (2022) [12] for the study have taken 4300 comments from online dating website having both positive as well as negative emotions. Using the combination of lexicon based approach and machine learning; the study could test and compare the user behavior. To make multiple classifications of user behavior, combination of machine learning with dictionary based sentiment analysis approach was used effectively.

Biradar, Shanta H., Gorabal, and Gaurav Gupta (2021) [13] developed an algorithm on the basis of sentiment analysis to make business decisions using open source framework tools such as Apache Hadoop. Sentiment analysis is combined with supervised machine learning techniques and large amount of data are extracted from twitter handled in an efficient way. The study observed that the developed algorithm is 1.5 times faster than the traditional method. 80% accuracy level is also ensured using the model. Tang, F., Fu, L., Yao, B., & Xu, W. (2019) [14] developed a model which improved the accuracy and performance while extracting aspects and opinions taken from online reviews. Using aspect based sentiment analysis and supervised machine learning algorithm, the extraction of opinions, aspects, granularities and polarities are made possible.

V. OPINION MINING

Wu, T., Hao, F., & Kim, M. (2021) [15] presented a framework model to excavate user review and opinions of

the animated film “Monkey King: Hero is Back” using crawler technology term frequency and inverse document frequency algorithm to analyze the production related problems of the movie. Using K-means algorithm, similar connotations but different expressions are combined together to perform the analysis. Sharma, Abhilasha, and Himanshu Shekhar [16] developed a framework which extracts the opinion of the public about various government policies taken from twitter and classified into three different polarities i.e positive, negative and neutral. In order to extract the sentiments of the public and to perform the analysis of polarity, Deep learning method and Natural Language Processing techniques were used. Padmavathy, P., and S. Pakkir Mohideen (2019) [17] proposed an approach to predict the satisfaction level of drug based on the review and opinions are collected from existing patients who already used the medicine. Support Vector Machine and Artificial Neural Network are used to perform this task.

Application of two – pass classifier is done to predict whether the opinion of drug satisfaction level is positive or negative. In this model, the features are extracted from the data set and performed the preprocessing task. When using the two pass classifier, the classification of the review whether it is positive or negative is performed. Based on precision, recall and f-measures, two different set of data sets are performed and are analyzed.

Karthik, Valmeekam, Dheeraj Nair, and J. Anuradha [18] proposed a method using Machine Learning Classifiers, Convolution Neural Networks and Artificial Neural Networks to analyze opinions from emojis taken from twitter. Separate mining of emojis and text are taken from the tweet which is performed to conduct an analysis to determine the polarity of the tweet. The method is applied on dataset having Unicode value of the emojis and its respective polarities. The relationship between the text polarity and that of the emoji is later obtained to inspect the presence of

sarcasm in the tweets. Rathan, M., Hulipalled, V. R., Venugopal, K. R., & Patnaik, L. M. (2018) [19] proposed a model for data mining from twitter looking for features like detecting emojis, spelling correction and detection of emoticons. By using lexicon based methodology, the model has an automated training data labelling. Observation showed that good accuracy is shown for classifier based on automated training data. In order to improve the classification accuracy, it is important to detect the emoji and the attribute specific lexicons were also demonstrated.

Tsirakis, N., Pouloupoulos, V., Tsantilas, P., & Varlamis, I. (2017) [20] proposed a solution to meet the challenges faced in handling large volumes of data while conducting the analysis of data which are taken from social media as examples. In order to obtain insights in motion on the data, the proposed model processes data in real time. For this, the real time opinion mining is developed. PaloPro – a new platform is also presented to handle the need of the time being for data handling. Yan, Zheng, Xuyang Jing, and Witold Pedrycz (2017) [21] conducted an experiment to extract information which are reputed and track the perspectives of the public based on opinion mining and fusing by filtering the opinions to trash the outliers and then using the fused method to group the related ones. Based on the grouped opinions, the conducted study observed that various kinds of recommendations could be generated. Opinions were taken from popular commercial websites in both Chinese and English.

VI CONCLUSION

This paper analyzed the influence of bias and various natural language processing methods such as sentiment analysis, opinion mining, fusing, etc on various domains. It is learnt that the influence of bias in various domains especially astrology affects the mindset and results of the concerned query. By combining the features of machine learning algorithms such as Naïve Bayes, Linear Regression, Support

Vector Machine, etc., the analysis of the effect of bias in various domains can be determined easily and effectively. Certain studies have found to be done on comparing the results by performing two different experiments using machine learning algorithm and traditional methods. Methods conducted using machine learning algorithms are superior than the result conducted using traditional method in obtaining the accuracy.

REFERENCES

- [1] Andersson, Ida, Julia Persson and Petri Kajonius, Even the stars think that I am suerios: Personality, Intelligence and belief in astrology”. *Personality and Individual Differences* 187 (2022): 111389.
- [2] Bokurt, Gozde and Ahmet. “Do Astrological Beliefs Reflects Systematic Bias in Personality measurement?.” *Alphamnumeric Journal* 9. (2021): 217-228
- [3] Helgetz Jonas, and Krik Scott. “The validity of astrological predictions on marriage and divorce: a longitudinal analysis of Swedish register data.” *Genus* 76.1 (2020): 1-18.
- [4] Cafferata, Alessia, Marwil J. Dávila- Fernández, and Serena Sordi. “Seeing what can (not) be seen: Confirmation bias, employment dynamics and climate change. “*Journal of Economic Behaciour & Organization* 189 (2021)” 567-586.
- [5] Kappes, Andreas et a;. “Confirmation bias in the utilization of others’ opinion strength. “*Nature neuroscience* 23.1 (2020): 130-137
- [6] Millner, Antony, Helene Olliver and Leo Simon. Confirmation bias and signaling in Downsian elections.” *Journal of Public Economics* 185 (2020): 104175.

- [7] Cafferata, Alessia, and Fabio Tramontana. "A financial market model with confirmation bias." *Structural Change and Economic Dynamics* 51 (2019): 252-259.
- [8] Gan, Lirong, Huamao Wang, and Zhaojun Yang. "Machine learning solutions to challenges in finance: An application to the pricing of financial products." *Technologies Forecasting and Social Change* 153 (2020): 119928.
- [9] Wang, An et al. "Potential of machine learning for prediction of traffic related air pollution." *Transportation Research Part D: Transport and Environment* 8 (2020): 102599.
- [10] Pavitha, N., et al. "Movie Recommendation and Sentiment Analysis using Machine Learning." *Global Transitions Proceedings* (2022).
- [11] Jardim, Sandra and Carlos Mora. "Customer reviews sentiment – based analysis and clustering for market oriented tourism services and products development or positioning." *Procedia Computer Science* 196(202): 199-206.
- [12] Li, Hui, et al. "E-word of mouth sentiment analysis for user behaviour studies." *Information Processing & Management* 59.1 (2022): 102784.
- [13] Biradar, Shanta H., J. V. Gorabal, and Gaurav Gupta. "Machine learning tool for exploring sentiment analysis on twitter data." *Materials Today: Proceedings* (2021).
- [14] Tang, Feilong, et al. "Aspect based fine – grained sentiment analysis for online reviews." *Information Sciences* 488 (2019): 190 – 204.
- [15] Wu, Ting, Fei Hao, and Mijin Kim. "Typical opinions mining based on Douban film comments in animated movies." *Entertainment Computing* 36 (2021): 100391.
- [16] Sharma, Abhilasha, and Himanshu Shekhar. "Intelligent Learning based Opinion Mining Model for Government Decision Making." *Procedia Computer Science* 173 (2020): 216224.
- [17] Padmavathy, P., and S. Pakkir Mohideen. "An efficient two – pass classifier system for patient opinion mining to analyze drugs satisfaction." *Biomedical Signal Processing and Control* 57 (2020): 101755.
- [18] Karthik, Valmeekam, Dheeraj Nair, and J. Anuradha. "Opinion mining on emojis using deep learning techniques." *Procedia Computer Science* 132 (2018): 197-173.
- [19] Rathan, M., et al. "Consumer insight mining: aspect based Twitter opinion mining of mobile phone reviews." *Applied Soft Computing* 68 (2018): 765-773.
- [20] Tsirakis, Nikos, et al. "Large scale opinion mining for social, news and log data." *Journal of Systems and Software* 127 (2017): 237-248..
- [21] Yan, Zheng, Xuyang Jing, and Witold Pedrycz. "Fusing and mining opinions for reputation generation." *Information Fusion* 36 (2017): 172-184.