ISSN 2278-3091
Volume 9 No.2, March - April 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse164922020.pdf
https://doi.org/10.30534/ijatcse/2020/164922020

Improving Travel Recommendation Accuracy using Fusion of Machine Learning Techniques

ZainabKahid Jouhar[1], Ashwini V.Zadgaonkar[2]
[1]Shri Ramdeobaba College of Engineering and Management,KatolRd,Lonand,Gittikhadan,Nagpur, India,
jouharzk@rknec.edu,zadgaonkarav1@rknec.edu

ABSTRACT

Data recommendation is a multi-processing task, which requires a set of operations ranging from input pre-processing, clustering, similarity analysis to prediction of data-trend. The trend prediction unit performs different kinds of calculations like classification, pattern analysis, error correction, etc. In order to develop a highly accurate recommendation system, the software designers have to accurately measure the performance of each of these algorithms and then identify the best fitting method that optimizes the accuracy of prediction. In our research, we many different combinations of these algorithms, and came up with a list of algorithms which may be best suited for obtaining higher level of prediction accuracy. These algorithms were later implemented, interfaced and integrated with each other in order to evaluate their performance. The performance of the proposed hybrid recommender is found to be 20% more effective than the existing state-of-the art recommenders. This paper also suggests improvement in the proposed system in order to further enhance the performance in terms of speed of operation and accuracy of recommendation.

Key words: Accuracy, classification, pre-processing, recommendation.

1. INTRODUCTION

Data analysis is the process of systematically searching and arranging the interview transcripts, field notes, and other materials that researcher accumulate to increase understanding and to enable the researcher to present what the researcher has discovered to others. It means that, the researcher should analyse and present his or her data in order to make the reader know which the steps taken in the processing of arranging data.

The technique of data analysis is the way data to be analysed by the researcher. The technique of data analysis in this research is qualitative approach.

a. Qualitative data: is process to organize, to choose, to look and to find important aspects from the data.

b. The qualitative data could be written in a form of words or symbol.

It means that, qualitative data is technique to analyse data from the research field and can be form words or symbols. This step is taken by the researcher in order to know how far the influence of the teacher's motivation strategies toward the students' learning English attitude. Then the researcher will describe more about the research problem. Data or information is collected data to be accurate, relevant and appropriate with the problem. Data are all of fact and member that can be used by the researcher as information, whereas information is result of data process that used necessary. In other word, data are everything that the researcher finds and writes from the field of study. Data in this research are getting by the researcher from in-depth interview with the English teacher and observation (about the English teacher and the students activities and attitude in the classroom during teaching and learning process), and supported by some documentation as follows the picture of the English teacher explanation, the English teacher's strategy to motivate the students, the students activity (doing task) and the students attitude on learning English. The next section describes various data analysis systems & their nuances, followed by our approach, and finally the result evaluation of the proposed approach on various input conditions. We conclude the paper with some interesting observations about the proposed analyser, and some further analytics which can be done in order to extend this work.

2. LITERATURE REVIEW

Work in [1] describe data analysis as the process of bringing order, structure and meaning to the mass of collected data. It is described as messy, ambiguous and time-consuming, but also as a creative and fascinating process. Broadly speaking - while it does not proceed in linear fashion -it is the activity of making sense of, interpreting and theorizing data that signifies a search for general statements among categories of data [2]. Therefore, one could infer that data analysis requires some sort or form of logic applied to research. In this regard, work in [3] clearly posit that the analysis and interpretation of data represent the application of deductive and inductive logic to the research. Work in [4] on the other hand, state that the interpretive approach [5], which involves deduction from the data obtained, relies more on what it feels like to be a participant in the action under study, which is part of the qualitative research. Very often the researchers rely on their experience of particular settings to be able to read the information provided by the subjects involved in the study. While this thesis employed a mixed method of data collection, namely a combination of qualitative [6] and quantitative methods [7], it focused on the adoption of a pragmatic position and also used a phenomenological approach in conducting this research. Work in [8] succinctly states that the word data points to information that is collected in a systematic way and organized and recorded to enable the reader to interpret the information correctly. As such, data are not collected haphazardly, but in response to some questions that the researcher wishes to answer. Work in [9] capture the essences of capturing data well when they further add, that data are not given as a fixed, but are open to reconfiguration and thus alternative ways of seeing, finding answers to questions one wishes to answer. Implicated in the preceding views of [10] are the two methods used to analyse data, namely qualitative and quantitative. Also, they state that a qualitative study involves an inseparable relationship between data collection and data analysis in order to build a coherent interpretation of data. An assumption of the qualitative researcher is that the human instrument is capable of ongoing fine tuning in order to generate the most fertile array of data. Work in [11] provides important views when they reiterate that the analysis of qualitative methods must be systematic, sequential, verifiable and continuous. It requires time, is jeopardized by delay, is a process of comparison, is improved by feedback, seeks to enlighten and should entertain alternative explanations. As with qualitative methods for data analysis, the purpose of conducting a quantitative study, is to produce findings, but whereas qualitative methods use words (concepts, terms, symbols, etc.) to construct a framework for communicating the essence of what the data reveal, procedures and techniques are used to analyse data numerically, called quantitative methods. On the whole, regardless of the method the purpose of conducting a study, is to produce findings, and in order to do so, data should be analysed to transform data into findings. In this study, data will be analysed using both the qualitative and quantitative method. At this point in time, one has to take a closer look at both methods of analysis. Regarding qualitative and quantitative analysis of data, offer a useful outline of the differences and similarities between qualitative and quantitative methods of data analysis. According to these authors, qualitative and quantitative analyses are similar in four ways. Both forms of data analysis involve:

- Inference - the use of reasoning to reach a conclusion based on evidence;
- A public method or process - revealing their study design in some way;
- Comparison as a central process – identification of patterns or aspects that are similar or different; and
- Striving to avoid errors, false conclusions and misleading inferences.

Apart from this work in [12] also offers an equally important view on analysis and interpretation of data, when he posits that the process and products of analysis provide the bases for interpretation and analysis. It is therefore not an empty ritual, carried out for form's sake, between doing the study, and interpreting it, nor is it a bolt-on feature, which can be safely ignored until the data are collected. The proposed analysis system is described in the next section.

3. PROPOSED DATA ANALYSIS SYSTEM

In order to provide an entire system for data-analytics, we used the following subsequent categories of algorithms,

- Clustering, wont to divide the info into groups of comparable values
- Classification, for categorizing the info into one of N categories
- Prediction, to gauge subsequent value of the series from a group of given input values
- Recommendation, used for recommending trends supported the input file

The developed engine uses k-Means for clustering, and k-NN for classification. The advice is additionally through with the assistance of nearest neighbour recommender, while the prediction engine uses random forest regression so as to predict subsequent value during a set of your time series data. Usually the closet neighbour recommender provides good recommendation accuracy, but thank to lack of exhaustive pattern analysis capabilities, there's always a bottleneck within the accuracy values of the algorithm. Thanks to this drawback, we've proposed the utilization of a hybrid fusion-based recommender system which utilizes the concepts of clustering, prediction, classification and eventually provides these results to the prediction engine. The diagram of the proposed recommender system are often observed from the subsequent figure,

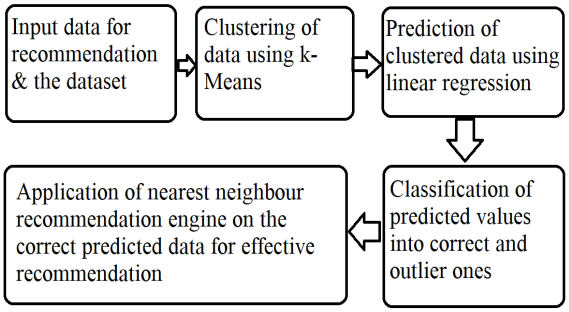
Input data for recommendation & the dataset
Clustering of data using k-Means
Prediction of clustered data using linear regression
Classification of predicted values into correct and outlier ones
Application of nearest neighbour recommendation engine on the correct predicted data for effective recommendation

Figure1. Proposed recommendation engine

The input data for recommendation alongside the dataset are given to the system for processing. The algorithm first performs k-Means clustering on the input data. This clustering allows the info to be segregated into different parts. Each part are going to be having internal similarity while different parts are going to be dissimilar at least one another. The parts which have highest intra-cluster similarity are considered for further processing. The K-means algorithm assigns each point to the cluster whose center (also called centroid) is nearest. The middle is that the average of all the points within the cluster — that's, its coordinates are the first moment for every dimension separately over all the points within the cluster.

Using the foremost similar clusters for further processing, we are removing all the info which is un-related to each other. Thereby the algorithm's delay of the advice is reduced. Once this is done, then the similar data is given to the Linear Regression-based prediction engine. Simple rectilinear regression may be a statistical procedure for obtaining a formula to predict values of 1 variable from another where there's a causal relationship between the 2 variables. Central to simple rectilinear regression is that the formula for a line that's most ordinary represented as $y = mx + c$or $y = a + bx$. Statisticians however generally like better to use the subsequent form involving betas:

$$y = \beta_{0+\beta_{1x}}$$

The variables y and x are those whose relationship we are studying. β_0 and β_1 are constants and are parameters (or coefficients) that require to be estimated from data. Their roles in the straight-line formula are as follows:

- β_0: intercept;
- β_1: gradient.

The deterministic component is a linear function of the unknown regression coefficients which require to be estimated in order that the model 'best' describes the info. This is often achieved mathematically by minimizing the sum of the squared residual terms (*least squares*). The fitting also produces an estimate of the error variance which is important for things like significance test regarding the regression coefficients and for producing confidence/prediction intervals. Using these intervals, we are ready to predict the patterns which could occur in the clustered data.

These patterns are given to a kNN classifier unit. This unit performs classification in order to validate the results from the linear regressor. The classification allows the system to see whether the responses from the regressor are accurate, or do they need some modifications. These responses are then given to a nearest neighbour recommendation engine. The recommender can be represented by the sebsequent diagram,

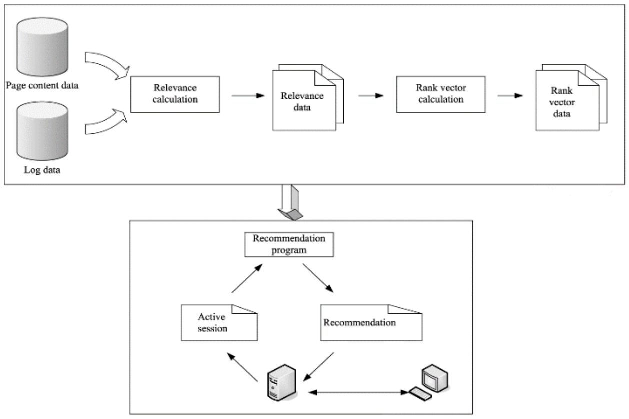

Figure 2. Nearest neighbour recommender

From the diagram we will observe that the input documents are given to a relevance calculation engine, followed by a rank calculator, which is given to a recommendation program. The program performs matching, and eventually produces recommendation outputs based on the given input file. We evaluated the results on different datasets, and therefore the results and analysis of the integrated proposed algorithm is given within the next section.

4. RESULT AND ANALYSIS

Our evaluation is done based on 2 main parameters, which are delay of execution and accuracy of recommendation. The delay of execution is taken as the time needed by the system to perform recommendation, while the accuracy of recommendation is the ratio of total number of correctly recommended results to the total number of recommendations suggested by the system. The following table showcases the delay needed by the system for recommendation,

Table 1: Delay results for recommendation

No. of Recommendations	Delay (ms) NN Recommender	Delay (ms) Hybrid Recommender
10	15.30	16.70
20	17.40	17.90
30	20.10	20.50
40	31.30	31.70
50	48.90	49.10
60	59.80	60.30
70	62.13	62.83
80	71.11	71.82
90	81.97	82.72
100	92.55	93.36

From the results we can observe that the delay of recommendation is slightly increased in case of the hybrid recommender. This is majorly due to the addition of 3 other algorithms before the actual recommendation engine. Due to this the accuracy of the recommendation is increased drastically. This can be seen from the following table,

Table 2: Results of accuracy

No. of Recommendations	**Accuracy (%) NN Recommender**	**Accuracy (%) Hybrid Recommender**
10	60.00	80.00
20	75.00	85.00
30	78.00	88.50
40	79.00	90.20
50	81.00	91.60
60	83.50	93.10
70	84.60	93.90
80	85.20	94.70
90	86.30	95.30
100	87.10	95.70

From the results we can observe that the overall accuracy has been increased drastically. Due to this the system has wide variety of applications in travel, medical and other recommendation areas.

5. CONCLUSION AND FUTURE WORK

From the results we can observe that the overall accuracy of the system is improved by more than 10%, due to which the system can be applied to any real-time application like medical recommenders, e-commerce recommenders, etc. Moreover, the delay is increased infinitesimally by 2%, this increase is not drastic due to removal of unwanted entries from the system. Due to which only the required entries which are needed for recommendation are used, while other repetitive and non-similar entries are removed. Moreover, due to addition of prediction engine the accuracy of recommendation is further improved.

The delay can be handled more efficiently by adding machine learning and artificial intelligence techniques which are delay and accuracy oriented. Algorithms like Q-learning and re-enforcement learning can be applied in order to further improve the overall efficiency of the system.

REFERENCES

1. Venkatesh, R., Balasubramanian, C. &Kaliappan, M. **Development of Big Data Predictive Analytics Model for Disease Prediction using Machine learning Technique**. *J Med Syst* 43, 272 (2019). https://doi.org/10.1007/s10916-019-1398-y
2. H. Purwins, B. Sturm, B. Li, J. Nam and A. Alwan, **"Introduction to the Issue on Data Science: Machine Learning for Audio Signal Processing,"** in IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 2, pp. 203-205, May 2019. https://doi.org/10.1109/JSTSP.2019.2914321
3. K. I. Moharm, E. F. Zidane, M. M. El-Mahdy and S. El-Tantawy, **"Big Data in ITS: Concept, Case Studies, Opportunities, and Challenges,"** in IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 8, pp. 3189-3194, Aug. 2019. https://doi.org/10.1109/TITS.2018.2868852
4. F. Amalina et al., **"Blending Big Data Analytics: Review on Challenges and a Recent Study,"** in IEEE Access, vol. 8, pp. 3629-3645, 2020.
5. A. Moubayed, M. Injadat, A. B. Nassif, H. Lutfiyya and A. Shami, **"E-Learning: Challenges and Research Opportunities Using Machine Learning & Data Analytics,"** in IEEE Access, vol. 6, pp. 39117-39138, 2018. https://doi.org/10.1109/ACCESS.2018.2851790
6. N. Sharma, D. Sawai and G. Surve**, "Big data analytics: Impacting business in big way,"** *2017 International Conference on Data Management, Analytics and Innovation (ICDMAI),* Pune, 2017, pp. 111-116.
7. Sowmya R and Suneetha K R, **"Data Mining with Big Data,"** *2017 11th International Conference on Intelligent Systems and Control (ISCO),* Coimbatore, 2017, pp. 246-250.
8. K. Jayamalini and M. Ponnavaikko**, "Research on web data mining concepts, techniques and applications,"** *2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET),* Chennai, 2017, pp. 1-5. https://doi.org/10.1109/ICAMMAET.2017.8186676
9. H. A. Madni, Z. Anwar and M. A. Shah, **"Data mining techniques and applications — A decade review,"** *2017 23rd International Conference on Automation and Computing (ICAC),* Huddersfield, 2017, pp. 1-7.
10. P. Akulwar, S. Pardeshi and A. Kamble, **"Survey on Different Data Mining Techniques for Prediction,"** 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2018 2nd International Conference on, Palladam, India, 2018, pp. 513-519. https://doi.org/10.1109/I-SMAC.2018.8653734
11. T. R. Kumar, T. Vamsidhar, B. Harika, T. M. Kumar and R. Nissy, **"Students Performance Prediction Using Data Mining Techniques,"** *2019 International*

Conference on Intelligent Sustainable Systems (ICISS), Palladam, Tamilnadu, India, 2019, pp. 407-411. https://doi.org/10.1109/ISS1.2019.8907945

12. M. Dawodi, J. A. Baktash and T. Wada**, "Data-Mining Opportunities in E-Government: Agriculture Sector of Afghanistan,"** *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, Vancouver, BC, Canada, 2019, pp. 0477-0481. https://doi.org/10.1109/IEMCON.2019.8936193
13. Adeagbo M. A., Akhigbe B. I. 2, Afolabi B. S, **"Towards A Job Recommender Model: An Architectural-Based Approach"** Volume 8, No.6, November 2019 ISSN 2278-3091, International Journal of Advanced Trends in Computer Science and Engineering. https://doi.org/10.30534/ijatcse/2019/94862019
14. MarveeCheska B. Natividad, Bobby D. Gerardo, Ruji P. Medina, **"A Career Track Recommender System for Senior High School Students using Fuzzy Logic"** Volume 8, No.5, September - October 2019 ISSN 2278-3091, International Journal of Advanced Trends in Computer Science and Engineering. https://doi.org/10.30534/ijatcse/2019/97852019