
HANDLE WITH ML: OPEN ENDED QUESTIONNAIRE FOR QUALITATIVE RESEARCH

Seema V. Purandare¹, Dr. Ranjit D. Patil²

¹Ness Wadia College of Commerce, Pune.

²D. Y. Patil Arts, Science and Commerce College, Pune.

Abstract

Open ended questions are considered as free form questions which gives freedom to respondents to answer the questions in natural language. Qualitative research which is based on open ended questionnaire plays important role in social sciences, customer satisfaction, customers' experience and loyalty analysis. It is challenging and problematic to analyze such responses which are in natural languages. This paper tries to understand these challenges and takes review of various methods used in analyzing these questionnaire. Advancement in technology has come up with a new way of making computer work in the same manner as human being using Machine Learning (ML). Various algorithms supported by ML also opened a new dimension for analyzing qualitative data. This paper also discusses how ML works and recent applications of ML in analysis of open ended questions.

Keywords: Open ended questionnaire, qualitative research, AI, NLP, ML.

► *Corresponding Author: Seema V. Purandare*

Introduction

The basic aim of research is to investigate, realize and update knowledge on various facets of a specific subject. Researchers can identify a problem in specific domain and can find solutions to it by applying an organized and systematic approach (Walia & Uppal, 2020). Cambridge dictionary defines research as "a detailed study of subject, especially in order to discover (new) understanding". Clifford woody says that Research comprises defining and redefining problems, formulating hypothesis or suggested solutions, collecting, organizing and evaluating data, making deductions and reading conclusions, and finally testing the conclusion to decide whether they fit the formulating hypothesis" (Kothari, 2004). Research plays a key role in identifying and solving various problems from different domains such as manufacturing companies, business, health care and so on.

Research process is considered as a step by step procedure for solving the problem identified. One of the important step in this procedure is to agree on the data to be used for the study and then accumulating the necessary data. Most common methods used for data collection are observation, interview and survey based on questionnaires given to targeted respondents. Questionnaire can be designed either as close ended i.e. in the form of multiple choice questions or as open ended. Close ended questions are easy to quantify and thus support quantitative research. When a questionnaire is not provided with set of possible responses, it is considered as open ended. Open ended questions are designed to give freedom to the respondent to answer the question and is mainly used for qualitative research.

A. Need for Open Ended Questions

The main goal of qualitative research is to try and understand the habits, tendencies, opinions and feelings of respondents related to specific event, product or situation. So it needs open ended

questions as they allow respondents to express their issues, experiences and depth to answers than close ended questions (Lahmar, 2022). Here the respondent can give response in his own words and can express his feelings, opinions or emotions in textual form. It can reveal hidden attributes and their correlation with others. Thus open ended questions work as a basis for qualitative research (J. sutton, 2015).

B. Advantages of Open Ended Questionnaire

Open ended questionnaire give certain advantages over close ended questionnaire such as (OptimistMinds, 2023).

- i. Respondents rarely give wrong answers
- ii. Respondents remember their answers
- iii. It helps researchers in getting detailed information about the research topic
- iv. It helps to gain answers which do not belong to specific range, thus gives valid and reliable information
- v. No need to prepare long list of responses for certain questions

C. Challenges faced by Open Ended Questionnaire

There are many disadvantages faced by open ended questions such as (Dr. Michael R. Hyman, 2016)

- i. The way a question is framed can affect the response
- ii. Respondents can take freedom of answering a question with long answers
- iii. Due to different types of respondents, response bias may be added
- iv. In some cases incomplete response may be generated
- v. There are interpretation and processing difficulties in analysis of open ended questions
- vi. It is time consuming to get the required data from respondents.

RELATED WORK

Open ended questionnaire are rarely used as compared to close ended questionnaire and if at all used, they are mostly analyzed manually or with partial support of a computer. This section tries to summarize few methods used and proposed for analysis of open ended questionnaire.

In her research article, Cornelia Züll specifies various methods of analysing open ended questionnaire, such as quantitative content analysis which can be performed with the help of computer. This method involves manual coding of responses and their categorization. Other coding and analysis approaches discussed in the article are semi-automatic coding, which is based on supervised machine learning or structural topic model and Co-occurrence analysis which focuses on words which occur together (Züll, 2016).

In their research work, A. Hiramatsu and others provided a description of a support system which can help in analyzing open ended questionnaire. Their system tries to segregate typical and atypical responses and then extraction of atypical responses using three different methods for further analysis. This separation of responses can be done first on the basis of ratio of typical words in a sentence, second method considers the order of words and difference between positions of words and third method splits the opinion into phrases. Researchers claimed that the third method is more effective and efficient for analysis (Ayako Hiramatsu, 2004).

Tim Gell in his article mentioned 4 different approaches for analyzing open ended questions which are as follows (Gell, 2019)

- i. Code responses are segregated into buckets. This involves assigning codes to each response and putting them in specific buckets which is a tedious and time consuming method.
- ii. Sharing full list of responses as it is and then apply cleaning i.e. grammar check, spell check and so on.
- iii. Input the text into word cloud which checks frequency of words appearing in responses using software.
- iv. Using text analytics which uses artificial intelligence (AI) to do coding and bucketing automatically but faces the problem of accuracy.

M. Roberts together with other researchers presented a semiautomated approach. Structural Topic Model (STM) which is based on machine learning based text analysis. This makes it easy to analyse open-ended questions and can extract hidden patterns. The researcher concludes that this approach will be useful for exploration, description, prediction and will support qualitative research in social sciences and experimentation (Margaret E. Roberts, October, 2014).

AI AND NLP

Artificial Intelligence (AI) is an evolving field in computer science which tries to inculcate human intelligence in computer. Thus, making computers capable of doing tasks which need human intelligence.

Whit Andrews, vice president and distinguished analyst with Gartner says, “AI is the use of intricate logic or advanced analytical methods to perform simple tasks at greater scale in ways that mean we can do more at large scale with the workers we have, allowing them to focus on what humans are best at, like handling complex exceptions or demonstrating sympathy”.

Zachary Jarvinen, head of technology strategy, AI and analytics at OpenText, “AI is essentially some computerized simulation of human intelligence that can be programmed to make decisions, carry out specific tasks, and learn from the results”.

AI has many subsets such as Machine Learning (ML), Deep Learning (DL) and Natural Language Processing (NLP).

Machine Learning is a field of AI which tries to add human intelligence to machine by using various statistical models.

Natural Language Processing (NLP) is a branch of AI which basically works on understanding and realization of natural language just like human beings.

Deep Learning is a subset of ML which makes use of Artificial Neural Network for the implementation of machine learning.

Figure 1 depicts the relation between the various terms AI, NLP, ML and DL.

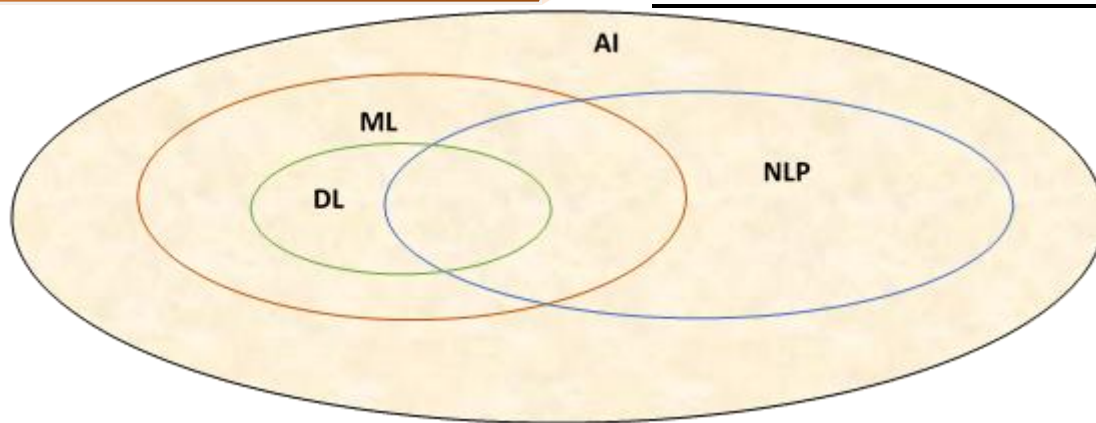


Fig. 1 Relation between AI, NLP, ML and DL

NLP tries to build machines which can understand textual data or speech and responds back either with text or speech of its own just like human beings.

NLP makes it possible for computers to communicate with humans by understanding, interpreting and manipulating human language (Overby, 2020). There are two subsets of NLP, Natural Language Understanding (NLU) and Natural Language Generation (NLG). NLU focuses on understanding the underlying text by following various steps as tokenization, stemming, lemmatization, speech tagging and name entity recognition. It basically converts the unstructured data from natural language to structured data which computers can understand. NLG performs exactly reverse process i.e. it tries to convert response generated by computer which is in the form of structured data to natural language or unstructured data which human can understand. Steps which are followed by NLG are content analysis, data understanding, sentence aggregation, grammatical structuring and language presentation.

Areas where NLP can be applied are content categorization, content extraction, opinion mining or sentiment analysis, document summarization, language translation, for handling chat driven or voice driven interfaces (Overby, 2020). As NLP is powerful enough to take care of human languages it can also be applied for analysing qualitative data generated through open ended questionnaire.

Text can be analysed using three different approaches (Dorash, 2017)

1. Rule-based (pure NLP) : It uses grammar based rules for text analytics. This approach does not need training of model.
2. Machine learning based (pure ML) : It uses algorithms which help a machine to learn a language of its own without explicit programming.
3. Hybrid (a combination of ML and NLP) : This approach uses combination of Rule- based and ML.

NLP through ML IN QUALITATIVE RESEARCH

Machine Learning (ML) mainly aims to make machines learn based on past experiences or illustrations. In ML, programs are not written for solving a specific problem but various algorithms and data are used to train a machine to solve a problem and then it's learning accuracy can be tested.

Figure 2 depicts general flow of Machine Learning model development.

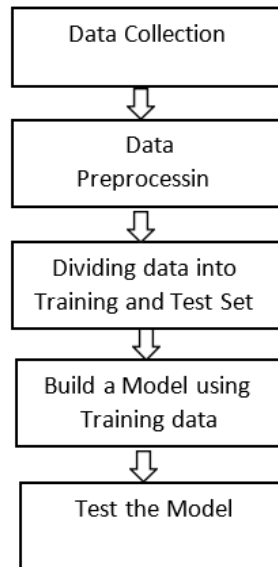


Fig. 2 General flow of ML model building

Machine Learning algorithms can be categorized into three broad categories

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Various ML algorithms used in qualitative research are:

1. Latent Dirichlet Allocation (LDA) :

It is an unsupervised ML algorithm that works on statistical probability theory. This algorithm tries to distribute each word to each defined topic. It tries to find out frequency of the word in a topic together with other words occurring with it. . LDA is one of the popular method used for NLP as it is easy to interpret. Most challenging task in LDA is determining optimal number of topics and iterations (Formoso, 2022).

Application of LDA :

Gaurav Nanda with other researchers applied LDA for analyzing open ended responses from about 1,50,000 respondents who have learnt through MOOCs. Main objective of the study was to identify significant aspects of MOOC from learners, their learning experience and their overall opinion about the course (Gaurav Nanda, 2021).

2. Support Vector Machine (SVM) :

Support Vector Machine (SVM) is one of the supervised machine learning algorithm which is used for classification problems. SVM plots the data item in n-dimensional space and then tries to classify these points into classes using optimal hyper-plane (Ray, 2017).

Application of SVM :

Ayşe Çınar with other researchers tried to develop an automated system for evaluation of descriptive answers for short answer questions. For their study they have considered four questions from physics and 246 student responded in Turkish. For analysis, this qualitative data machine learning algorithm SVM was used. In addition to SVM the researchers also used Gini, k-Nearest Neighbours (KNN) algorithm and concepts like bagging and boosting. The study concluded by

proving that it is possible to analyse open-ended questions and the model designed can be used for measuring knowledge expressed by answering open ended questions (Ayşe Çınar, 2020).

Conclusion

Open ended questionnaire helps respondents to express their feelings, emotions, experience and in depth knowledge related to specific subject, scenario or situation. The main problem which makes it very tedious to work with qualitative data is its analysis and interpretation. Advancement in technology opens the world of ML and different ML algorithms are applied in analysis of qualitative data.

References

1. Ayako Hiramatsu, H. O. (2004). Support System for Analyzing Open-ended Questionnaires Data by Culling Typical Opinions. 2004 IEEE International Conference on Systems, Man and Cybernetics, 1377-1382.
2. Ayşe Çınar, E. I. (2020). Machine learning algorithm for grading open-ended physics questions in Turkish. Springer.
3. Cambridge Dictionary. (n.d.). Retrieved from <https://dictionary.cambridge.org/dictionary/english/research>:
4. Dorash, M. (2017, December 27). Machine Learning vs. Rule Based Systems in NLP. Retrieved from Friendly Data: <https://medium.com/friendly-data/machine-learning-vs-rule-based-systems-in-nlp-5476de53c3b8#:~:text=Overall%2C%20a%20rule-based%20system%20is%20good%20at%20capturing,in%20general%20a%20better%20fit%20for%20query%20analysis>.
5. Dr. Michael R. Hyman, N. . (2016, February). Open- versus Close-Ended Survey Questions. Business Outlook.
6. Formoso, J. B. (2022, January 15). 4 Steps of Using Latent Dirichlet Allocation (LDA) for Topic Modeling in NLP. Retrieved from Omdena: <https://omdena.com/blog/latent-dirichlet-allocation/>
7. Gaurav Nanda, K. A. (2021). Analyzing Large Collections of Open-Ended Feedback From MOOC Learners Using LDA Topic Modeling and Qualitative Analysis. IEEE , 146- 160.
8. Gell, T. (2019, august 7). Retrieved from driveresearch: <https://www.driveresearch.com/market-research-company-blog/4-approaches-to-analyzing-open-ended-survey-responses/>
9. J. sutton, Z. A. (2015). Qualitative Research: Data Collection, Analysis, and Management. CJHP, 68(3): 226-231.
10. Kimola. (2020, June 12). How to Use Machine Learning in Qualitative Research? Retrieved from Kimola Cognitive: <https://kimola.com/cognitive/resources/how-to-use-machine-learning-in-qualitative-research>
11. Kothari, C. R. (2004). Research Methodology methods and techniques. New Delhi: New Age International (P) Limited Publishers.
12. Lahmar, M. (2022, November). Analyzing_Open-ended_Questions_for_Qualitative_Research. Retrieved from Researchget.net: https://www.researchgate.net/publication/365775711_Analyzing_Open-ended_Questions_for_Qualitative_Research
13. Margaret E. Roberts, B. M.-L. (October, 2014). Structural Topic Models for Open-Ended Survey Responses. American Journal of Political Science, 1064-1082.

14. OptimistMinds. (2023, January 24). Advantages of open-ended questions in qualitative research. Retrieved from [optimistminds.com](https://optimistminds.com/advantages-of-open-ended-questions-in-qualitative-research/): <https://optimistminds.com/advantages-of-open-ended-questions-in-qualitative-research/>
15. Overby, S. (2020, February 26). Artificial intelligence (AI) vs. natural language processing (NLP): What are the differences? Retrieved from The enterprisers project: <https://enterpriseproject.com/article/2020/2/artificial-intelligence-ai-vs-natural-language-processing-nlp-differences>
16. Ray, S. (2017, September 13). Learn How to Use Support Vector Machines (SVM) for Data Science. Retrieved from Analytics Vidhya: <https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/>
17. Walia, A. M., & Uppal, M. (2020). Fundamentals Of Research. Chennai: Notion Press.
18. William Leeson, A. R. (November 13, 2019). Natural Language Processing (NLP) in Qualitative Public Health Research: A Proof of Concept Study. Sage Journals <https://journals.sagepub.com/doi/full/10.1177/1609406919887021>.
19. Züll, C. (2016, December). Retrieved from [gesis.org](https://www.gesis.org): https://www.gesis.org/fileadmin/upload/SDMwiki/Zuell_Open-Ended_Questions.pdf