

Sentiment Analysis Based on 100 Doctoral Acknowledgments

Zhou Shihao

College of Humanities and Social Development, Northwest A&F University, Yangling, Shaanxi, China
18615089789@163.com

Keywords: LDA, Sentiment Analysis, Acknowledgments

Abstract: This paper conducts sentiment analysis and LDA topic model on 100 doctoral thesis acknowledgement texts to explore the changes in emotional inclination and attitudes towards different subjects, events, and objects in the process of acknowledgement writing. The aim is to build a doctoral thesis acknowledgement corpus by collecting 28,152 segmented words and establish a time span of 2017-2024. The results are as follows: (1) There are five types of high-frequency words in doctoral acknowledgements: important interpersonal relationships (children, parents, etc.), key actions, focus objects, critical processes, and psychological feelings; (2) On a statistical level, the positive emotions in doctoral acknowledgements are significantly greater than the negative emotions. However, considerable negative emotions need to be paid attention to, such as anxiety and guilt; (3) The positive topics of doctoral acknowledgements include support and companionship, guidance and assistance, life, academic and school, and personal growth. The negative themes include thesis writing, research experience, mentor-disciple relationship, and institutional assistance.

1. Introduction

Every year, some college graduates' thesis acknowledgments attract public discussion, no matter what form they take or whether they are written in narrative or poetry, they all express the author's spirit of perseverance, striving for progress, and gratitude. Unlike the rigorous language expression in the thesis, the author can express their emotions freely in the thesis acknowledgments, expressing gratitude to their teachers, classmates, and family members who have helped them during their studies.

The research on thesis acknowledgments by domestic scholars mainly focuses on the lexical expressions and pragmatic identity analysis in linguistics. For example, Li Juan constructed the "professional identity, social identity, and personal identity" model based on the sentence structures of doctoral thesis acknowledgments from different disciplines and different periods in China, and differentiated the motivations of different identities. ^[1] Some scholars in the field of library and information studies have also conducted descriptive statistical analysis of the content of thesis acknowledgments. Jin Wugang et al. studied whether the text of doctoral thesis acknowledgments mentioned the help of the school library in completing academic papers, pointing out that China's academic elites use library resources at ease but lack gratitude. ^[2] Yang Yihong et al. analyzed the

text of 90 excellent doctoral thesis acknowledgments in China from multiple angles, statistically analyzed the categories of acknowledgment objects, the gender differences of doctoral candidates and supervisors, and the support role of funding for excellent doctoral theses.^[3] In summary, these studies are all descriptive statistical studies, lacking research on the combination of thesis acknowledgments and doctoral students' emotions. Based on this, this paper intends to use emotion analysis and LDA topic modeling for research.

2. Materials and Methods

We collected 100 doctoral theses of sociology majors on Wanfang Data knowledge service platform. Firstly, we performed lexical analysis on the data, made word cloud map for an overall overview, then used snownlp library for sentiment analysis, and finally used Python for LDA theme modeling to obtain positive and negative themes in the field of paper acknowledgements. The specific process is shown in the figure below.

2.1. Data acquisition

In November 2024, through the Wanfang Data knowledge service platform, 860 doctoral theses were retrieved with the theme of professional "sociology" and degree "doctor", and 100 doctoral thesis PDF files were captured in chronological order. These PDFS were extracted with the pypdf2 library in Python, and the extracted acknowledgement paragraphs were saved as txt text. The jieba library is then used for word segmentation operation, and the operation of removing noise, special characters and blank lines is carried out simultaneously in this process, and the stop word table is directly adopted by the stop word table issued by the laboratory of HIT University, and the Wan Fang Sociology doctoral thesis acknowledgements corpus is obtained, with a total of 860pieces of corpus. In particular, it should be noted that the most authoritative database of doctoral thesis acknowledgments should be the excellent master's thesis database of China National Knowledge Network, but China National Knowledge Network no longer provides the thesis acknowledgments page, fortunately, Wan Fang can also download the thesis containing the acknowledgments page.

2.2. Sentiment analysis

The snownlp library of Python can be used for quick Chinese sentiment analysis, especially for the sentiment trend analysis of each sentence in the acknowledgments text, so that readers can recognize the emotional changes of the author when writing acknowledgments. Its sentiment analysis function works by calculating the text's sentiment inclination score, which varies between 0 and 1.

In addition, Baidu Intelligent Cloud api is also used for emotion analysis under artificial intelligence. After registering an account and obtaining the two keys of API Key and Secret Key, Baidu emotion analysis API supports the judgment of emotional tendency of Chinese text, including three emotional tendencies of positive, negative and neutral, in addition to the confidence level indicating emotional polarity. The range is 0 to 1, the higher the value, the higher the confidence of the model in the judgment, and finally the probability of representing the positive and negative emotions of the text is output.

2.3. Latent Dirichlet allocation

Latent Dirichlet Allocation (LDA) is a topic model widely used in text mining to discover hidden topic distributions from a large number of documents. It is based on the assumption that documents are generated from a mix of implicit topics, and that each topic is made up of a series of words. In the

LDA model, each document is treated as a bag-of-words, ignoring the order of words but keeping information about their frequency. With follow-up measures, you can find the topic words for each topic and the topic distribution for each document

3. Results & Discussion

3.1. Doctoral thesis acknowledgments mainly express positive emotions.

The result of word segmentation of acknowledgments is 28,152 phrases in total, among which, it can be found that there are roughly 5 categories: (1) important interpersonal relationships ^[4] : Teachers, parents, senior brothers, senior sisters, family members, colleges, students; (2) Key actions: research, study, research, support, writing, revision, research, defense, companionship, advice; (3) Focus on: thesis, academic, life, society, sociology, doctoral thesis, study; (4) Key processes: period, process, study, time, growth, experience, master; (5) Psychological feelings: gratitude, encouragement, concern, gratitude.

From the above five categories of words, it can be seen that the overall emotion of the sociology doctor was mainly positive during his/her doctoral study. Besides his/her family, the most important interpersonal support during the doctoral study was his/her supervisor, fellow teachers and the college. During the doctoral study, he/she seemed to be more focused on academics, with more words such as research, writing and research, while fewer words reflecting his/her own life. At the psychological level, there are more positive emotions such as gratitude, of course, there are also words such as confusion, pressure, anxiety, etc., but the word frequency is about 20, accounting for a relatively small number

3.2. Acknowledgements have emotional ups and downs, but they can be adjusted over time.

Baidu Cloud api interface was used to analyze the overall emotion of each document. Among the 78 documents, 67 could be identified, and 11 could not be identified, and only one of the 67 documents had a negative overall emotional tendency (propensity tendency value was 0.193, but a closer look showed that the acknowledgements were written in classical Chinese. Its writing technique may not be fully understood by ai), and the rest are positive and above 0.95.

Through the line chart, we can intuitively see the author's emotional changes when writing acknowledgements. Although positive emotions are the main ones, we can still see the author's emotional fluctuations. When the authors express their gratitude, their emotional tendency is obviously positive, and the emotional curve is stable and close to 1; When the author recalls the doctoral experience, especially how to get married with the teacher and the psychological course of research and writing, the emotional curve will fluctuate significantly. The doctoral life is not easy, and at this moment of writing acknowledgements, the author looks back on the study time naturally with a lot of emotions, but more for the pride of the end of the doctoral life and the vision of the future academic; Carefully examine those statements whose emotion tends to 0, which represents extremely negative, accounting for a small proportion in the whole text, mainly can be divided into: (1) worry about their own illness or the illness of relatives; (2) The author's guilt for not being able to help his parents or lovers during his studies; (3) Feelings about poverty and other social facts found in the course of research; (4) The signing at the end of the paper. Almost every thank you in this part has an extremely negative emotional trend, which may contain the mentality of the author when the writing is completed.

The acknowledge of the doctoral thesis is the doctor's summary of his past study, and is one of the few arbitrary chapters in the thesis, carrying the author's strong emotions, by using snownlp for each sentence of the emotional trend analysis, the author tries to glimpse some of the regularity of

things. In general, the doctor expressed positive feelings in the thesis acknowledgements, especially affirming his original choice and adherence to the academic path, although he suffered some material or psychological obstacles in the process, he could eventually overcome and become an unforgettable achievement in his doctoral career. However, some authors feel guilty about their lives and social aspects, thinking that they could not help their families share the pressure during their studies, or they were unable to do anything about some phenomena they saw during the research period, and could not "reach and benefit the world".

3.3. Latent Dirichlet allocation

Combining the confusion degree and consistency curve, the optimal theme number can be obtained in the range of 4 to 15, and the theme dispersion degree is visualized by pyLDAvis toolkit to determine the large range and non-overlapping when the theme number is 5.

The table 1 shows that the theme 1 is mainly about the support and companionship during the whole process of studying for a PhD. The author expresses his gratitude to "classmates", "parents" and "children", etc. For example, "I am grateful to my family for their sincere care and selfless efforts for me in recent years, and for their tolerance, understanding, support and help", "In most cases, The reason why the postscript of the doctoral thesis is a kind of grateful style is mainly because the acquisition of the cultural capital presented by the physical form, material form and institutional form as Bourdieu said, embodies the blood and sweat of too many people." This is a holistic review of the acknowledgments.

Topic 2 is mainly about the specific guidance and help of teachers during doctoral study ^[5]. The help of teachers has always been the most important support during the doctoral period, and most doctors will focus on "teacher", "thesis", "guidance", "support" and "revision" in their acknowledgments. For example, "Teacher Jiang is a simple, kind, in the shoes of the teacher, for the sake of the students" "Although the teacher usually does not speak much, but talking about academics is endless, and, usually busy, as long as the student has a request, the teacher is willing to respond. And there will be a response soon."

Topic 3 is mainly about the support and help in life, "life", "responsibility", "care", "second elder" and "husband". During the doctoral study, the family undertook a large number of family responsibilities, and these support and help inspired the author's strong gratitude. For example, "They still send greetings, send beautiful flowers and delicious food to me when I am stuck in the rental house and bored, under their care, more than half a year, the weight continues to rise, the face becomes bigger and bigger, which is probably the so-called" happiness fat "" I wish through my efforts, take on more family responsibilities, so that they can live less labor," More easy life."

Topic 4 mainly focuses on the gratitude to the academic and school institutions. Although their own efforts are indispensable during the doctoral study, the guidance of the academic circle and the school is also the key to their growth. For example, "To this day, after the doctoral thesis has gone through the relevant procedures such as review, blind review, defense, and so on, when it is submitted to the final draft, I am still nervous, deeply afraid that the prestige of the teacher in the academic world will be damaged by the bad work of this unqualified student, and my heart is sad" "In the more than 1,700 days and nights in Wuhan University, It washed away my youthful energy when I first came here, and laid the foundation for my future academic road."

Topic 5 expresses gratitude to the experience of personal growth, although there is "anxiety" during the doctoral degree, but because there are "parents", "brothers and sisters" and "society" to "support and accompany", doctoral students can constantly improve themselves in the process.

Table 1: The positive theme

Topic	Keywords
Support and accompany	Acknowledgements, PhD, Study, Thesis, Support, Accompany, Give, During, Work, Children, Research, Writing, Parents, Ten years, Life, Encourage, Classmates, Finally, Study
Guidance and help	Teacher, Acknowledgements, Thesis, Research, Sociology, Help, Give, Society, Process, Survey, Writing, Guidance, PhD thesis, Academic, Study, Provide, Support, During, Student, Revise
livelihood	Husband, Mentor, Society, Always, Take care, Mother, Bear, Now, Give up, Difficult, Parents, Elderly parents, Hardship, Limited, Achieve, Life, Hustle, Learned, Husband
Academic and school	Sociology, Teacher, Mentor, Jilin University, Academic, Future, Society, In heart, Grateful, Philosophy, Path, Life, Study, College, Exactly, Look back, Acknowledge, Experience, Move forward, My teacher
Personal growth	Teacher, Academic, Acknowledgements, School, Without, Parents, Pursue PhD, Siblings, Society, Support, Growth, Strength, Improve, Encourage, Study, Together, A, Anxiety, Choice, Relatives

4. Conclusion

Through the combination of emotion analysis and LDA thematic model, the author makes a conclusion: (1) There are five types of high frequency words of doctoral acknowledgments: important interpersonal relations (children, parents, etc.), key actions, objects of concern, key processes, and psychological feelings; (2) On the whole, the positive emotion of doctoral dissertation acknowledgments is significantly more than the negative emotion in statistical significance, but there are also considerable negative emotions that need attention, such as anxiety and guilt; (3) The doctoral thesis acknowledges the topics of positive emotion such as support and companionship, guidance and help, life, academic and school and personal growth, and the negative topics such as thesis writing, research experience, teacher friendship and institutional help. In short, the doctoral students in the acknowledgments, as the final winners, may have a certain nature of exhibition, but also need to combine their achievements after becoming scholars to conduct comprehensive research, perhaps can get more effective results.

References

- [1] LI Juan. *Research on acknowledgments from the perspective of pragmatic identity*. *Foreign Language Studies*, 2016, 33(2): 33-38. (in Chinese)
- [2] Jin Wugang, RAO Shengping, Zhan Huaqing. *An empirical study on Library Evaluation from the perspective of readers: An analysis of the content of Library evaluation in the text of university doctoral dissertation "Acknowledgments"* [J]. *Journal of University Library*, 2006(6): 11-17.
- [3] Yang Yihong, Wan Xiaoying, Wu Yishan. *Analysis of the contents of acknowledgments of China's Top 100 Outstanding doctoral Theses in 2012* [J]. *Journal of Information*, 2014, 33(1): 62-66, 119.
- [4] Traore Sekou Amadou. *Socioeconomic and Perceptive Factors of HIV Transmission from Mother to Child in Bamako—Mali* [D]. *Central China Normal University*, 2020.
- [5] Liang Han, Hu Zhigang, Luo Wen, et al. *The more sincere the gratitude, the more successful the future? Research on the relationship between the content of acknowledgments of doctoral theses and academic achievements based on fsQCA* [J]. *Library Forum*, 2022, 42(11): 50-59.