



Sentiment Analysis of Teaching Evaluation Based on Complex Semantic Analysis

Lina Chen¹(✉), Xianzhi Huang², Hongjie Guo¹, Fangyao Shen¹, and Hong Gao¹

¹ School of Computer Science and Technology, Zhejiang Normal University, Jinhua, China
chenlina@zjnu.cn

² College of Physics and Electronic Information Engineering, Zhejiang Normal University,
Jinhua, China

Abstract. In the era of digital education, student evaluations of teachers constitute a crucial component of digital education. They serve as a driving force for promoting teaching reforms and are the fundamental basis for enhancing the quality of education and teaching. The digitalization of education reform provides students with more opportunities and avenues for evaluating teachers, including classroom teaching evaluations, online network teaching evaluations, and periodic assessments. Effectively utilizing this evaluation data to identify student needs and discover teaching issues is one of the effective ways to implement student-centered educational reforms. Through analysis, it has been found that existing student evaluation data exhibits characteristics such as implicit expression and complex emotional semantics, posing significant challenges for data analysis. This paper addresses these challenges by constructing a sentiment lexicon in the educational evaluation domain and employing complex semantic analysis to more accurately analyze the underlying emotional states within the evaluation data. The methodology involves the expansion of a general sentiment lexicon. Using active learning algorithms, sentiment seed words are selected from the evaluation data. Based on these seed words, an educational domain sentiment vocabulary is generated using the SO-PMI algorithm. The normalized educational domain sentiment vocabulary is then merged with the expanded general sentiment lexicon, resulting in the construction of the educational evaluation domain sentiment lexicon. During the evaluation phase, complex semantic analysis is applied to the evaluation data, and the educational evaluation domain sentiment lexicon is used for data analysis. Experimental results indicate that the proposed method achieves consistency with the actual data ranking in terms of sentiment classification and evaluation scores (MAE = 1.06, RMSE = 1.28). The F1 values for positive and negative teaching comments increased by 7.3% and 34.9%, respectively, compared to a general sentiment lexicon. Furthermore, when compared with common supervised learning algorithms, the proposed algorithm demonstrates superior sentiment classification performance.

Keywords: Student Evaluation of Teaching · Sentiment Analysis · Teaching Evaluation Domain Lexicon · Active Learning

1 Introduction

In the realm of digital education, Student Evaluations of Teaching (SET) are crucial for educational reform but face challenges due to the lack of domain-specific sentiment lexicons. The complex nature of teaching comments leads to discrepancies between evaluation outcomes and actual performance, hindering the evaluation process. SET involves collecting student feedback on course instruction to enhance teaching quality [1]. Text Sentiment Analysis (SA) in education has been explored by various researchers, such as Balahadia et al. [2], who used sentiments from evaluations to develop a performance evaluation system. Lin et al. [3] applied machine learning to extract sentiment from SET, and Wang et al. [4] used sentiment lexicons for emotional analysis in educational news. However, relying on general sentiment lexicons poses challenges to uncover deeper information [5].

Some studies address this limitation using different approaches. Hatzivassiloglou et al. [6] demonstrated the reliability of polarity relationships in English text, while Huang et al. [7] used conjunctions and emotional polarity constraints. Liu et al. [8] expanded HowNet to create an emotional lexicon, and Yang et al. [9] utilized HowNet and NTUSD for emotional tendency analysis. Zhou et al. [10] adopted cross-lingual techniques to extract semantic elements from HowNet.

Knowledge-based methods, like those employed by Zhang et al. [11] and Cai et al. [12], offer versatility but may lack domain specificity. Bollegala et al. [13] annotated word polarity using PMI, while Wawer [14] used search engines for sentiment seed words. Yang et al. [15] determined sentiment polarity using Baidu search results, and Gao et al. [16] enhanced a general sentiment lexicon with specialized lexicons for sentiment analysis of user reviews.

In conclusion, sentiment analysis in SET faces challenges, especially with implicit and complex language. Various approaches, including machine learning and knowledge-based methods, are employed to address these challenges, each with its strengths and limitations.

This paper addresses the mentioned issues by introducing a Sentiment Lexicon for Teaching Evaluation (SL-TeaE [17]) and proposing a method called SL-TeaE(CSA) based on this lexicon. The key contributions are:

1. Generation of a sentiment lexicon for evaluating teaching. Sentiment seed words are chosen using an active algorithm to create a domain-specific sentiment lexicon for teaching, employing the SO-PMI algorithm. This enhances the model's generalizability and sentiment classification accuracy. Varying weights, determined by a gradient descent formula, are assigned to different intensity adverbs, forming an adverbs of degree list. Additionally, a negative word list is constructed based on negation words. The incorporation of the adverbs of degree list and negative word list into the general sentiment lexicon enables a more precise emotional analysis of teaching comments, expanding the sentiment lexicon for teaching evaluation. Integration of domain-specific sentiment words into the expanded general sentiment lexicon improves the performance of the generated teaching evaluation domain sentiment lexicon in sentiment analysis of teaching evaluations.

2. Complex Semantic Analysis. Complex semantic analysis is applied to teaching evaluation data to more accurately extract semantic features from the evaluation comments.

2 SL-TeaE(CSA) Model Diagram

The model diagram of SL-TeaE(CSA) is shown in Fig. 1.

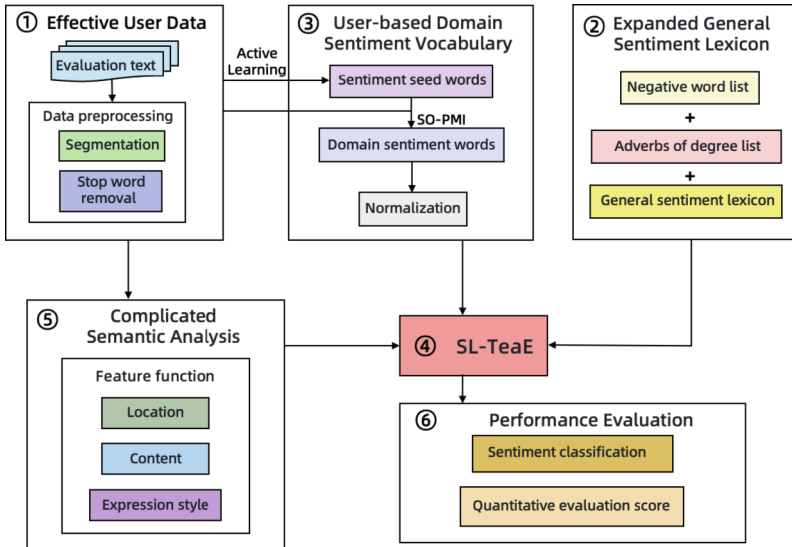


Fig. 1. The model diagram of SL-TeaE(CSA)

The teaching evaluation sentiment analysis model based on semantic analysis is divided into six sections:

1. Teaching Evaluation Data Preprocessing. The evaluation text undergoes preprocessing operations, including tokenization and stop-word removal, to enhance data quality. This step involves breaking down the text into individual words (tokenization) and eliminating common stop words, contributing to improved data quality.
2. The expansion process involves selecting a foundational sentiment lexicon, constructing a list of negation words, and creating a list of adverbs denoting intensity. These steps collectively contribute to enlarging the general sentiment lexicon, providing a more comprehensive set of words for sentiment analysis.
3. Generation of Teaching Evaluation Domain Sentiment Words.
 - a. Generation of Sentiment Seed Words. An active learning algorithm is used to select sentiment seed words from preprocessed teaching evaluation data. These seed words are used to generate domain-specific sentiment words. The active learning algorithm selects words with maximum coverage for annotation. The TextRank algorithm [18], combined with the K-Means clustering algorithm, is used to generate sentiment seed words.

- b. Domain-Specific Sentiment Words Generation: Utilizing the selected sentiment seed words, the SO-PMI algorithm identifies necessary domain-specific sentiment words from teaching evaluation data, determining their sentiment polarity and tendency values.
 - c. Normalization of Sentiment Inclination Values: Aligning the sentiment intensity of domain-specific sentiment words with the general foundational sentiment lexicon ensures a consistent scale for SA.
4. Generation of SL-TeaE. The normalized domain-specific sentiment words merge with the expanded general sentiment lexicon, forming the teaching evaluation domain sentiment lexicon.
 5. Complex Semantic Analysis. Semantic analysis is conducted on the evaluation data to extract emotional central sentences representing the overall viewpoint of the reviewer as sentences representing the overall viewpoint of the reviewer.
 6. Performance Evaluation. This section comprises sentiment classification and quantitative evaluation scores analysis. It involves evaluating the performance of general sentiment lexicon expansion, domain-specific sentiment word enrichment, and complex semantic analysis by comparing their performance on teaching evaluation data in terms of SA and sentiment computing.

3 Complex Semantic Analysis

3.1 Sentiment Center Sentences

Usually, the semantics in Student Evaluations of Teaching (SET) comments are more complex, expressions are more implicit, and emotions are more subtle. When a student writes a SET comment, they typically do not express negative emotions directly but rather use relatively implicit expressions. For example, phrases like “Compared to Professor Wang, Professor Li’s teaching could be improved” or “It would be better if this instructor had a teaching assistant” are common. Due to the complex nature of these comments, it is challenging to extract emotional features from SET comments.

Typically, the sentiment polarity in SET comments is determined by the most critical opinions of the reviewers rather than minor details. Therefore, it is essential to focus on extracting sentences that represent the overall opinions of the reviewers from SET comments. Here, we refer to the sentences that can represent the overall opinions of the reviewers as “sentiment center sentences”. We evaluate SET comments’ sentiment center sentences from three angles: the position angle, the content angle, and the expression style angle.

Firstly, from the position angle, in a SET comment, sentences at the beginning and end of the comment are more likely to become sentiment center sentences. Therefore, the position feature function should assign higher scores to sentences at the beginning and end. Experimental results show that a negative Gaussian function can be used as the position feature function. Thus, given a sentence “s”, the position feature function is determined as shown in Eq. (1):

$$f_1(s) = -\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(s-\mu)^2}{2\sigma^2}}, 1 \leq s \leq len \quad (1)$$

In Eq. (1), μ represents the mean, σ represents the standard deviation, and len represents the length (the number of sentences in one comment). In the subsequent experiments, μ is set to $len/2$, and σ is set to 1.

From a content perspective, sentiment center sentences not only exhibit strong emotional intensity, but the sentiment polarity should also be unambiguous. Therefore, the definition of the content feature function is as shown in Eq. (2):

$$f_2(s) = \frac{\sum_{t \in s} l(t)}{\sum_{t \in s} |l(t)|} \quad (2)$$

In Eq. (2), $l(t)$ represents that the word “ t ” is a sentiment word and indicates its sentiment polarity. When ‘ t ’ is a positive word, $l(t) = 1$; conversely, when ‘ t ’ is a negative word, $l(t) = -1$. From the content feature function, it can be observed that only sentences containing sentiment words with the same polarity receive higher scores, while sentences with no sentiment words or a mixture of negative and positive sentiment words receive lower scores.

From an expression style perspective, sentiment center sentences often include summarizing words or phrases like “In conclusion” and “All in all “. Therefore, the definition of the expression style function is as shown in Eq. (3)

$$f_3(s) = \sum_{t \in s} \text{conclusive_Expressions}(t) \quad (3)$$

In Eq. (1), $\text{conclusive_Expressions}(t)$ represents that “ t ” is a summarizing expression. If a sentence contains summarizing words or phrases, the score of that sentence will be higher.

Taking into consideration the three feature functions of sentiment center sentences, the summation of feature functions in Eq. (1), (2), and (3) is performed. The top N sentences with the highest scores are selected (typically N is set to 1 or 2) as sentiment center sentences. If the total number of sentences in a text is less than N , all sentences are considered sentiment center sentences.

After conducting complex semantic analysis on the SET data following data preprocessing, sentiment center sentences from each SET comment are subjected to sentiment classification using SL-TeaE. This approach is referred to as SL-TeaE (CSA).

Input the SET comments after data preprocessing.

Perform complex semantic analysis on the SET comments, and output the sentiment center sentence for each SET comment.

4 Experimental Results Analysis

4.1 SET Data

The data originated from the SET data in our university’s teaching system, including end-of-semester total evaluation data, mid-teaching phase data and online course real-time data. In total, there are 519 pieces of evaluation data from 4 teachers, and after data preprocessing, 508 pieces of valid data remain as corpus.

4.2 Experimental Metrics

We utilize standard evaluation metrics commonly employed in sentiment analysis models, including precision (P), recall (R), and F1 score (F1), and utilize the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) as evaluation metrics for the quantitative scoring experiments in course evaluations.

4.3 Performance Comparison

Comparison of Sentiment Classification Performance. Teaching comments encompass both Positive Teaching Comments (PTC) and Negative Teaching Comments (NTC). By conducting comparisons with a General Sentiment Lexicon (GSL) and an Expanded General Sentiment Lexicon (EGSL), this study validates the effectiveness of the proposed domain-specific sentiment lexicon construction. Simultaneously, comparative experiments were carried out with common supervised learning algorithms such as K-Nearest Neighbors (KNN), Naïve Bayes algorithm, Maximum Entropy model (ME), and Support Vector Machine (SVM). The results of the comparative experiments on PTC and NTC are illustrated in Fig. 2 and Fig. 3.

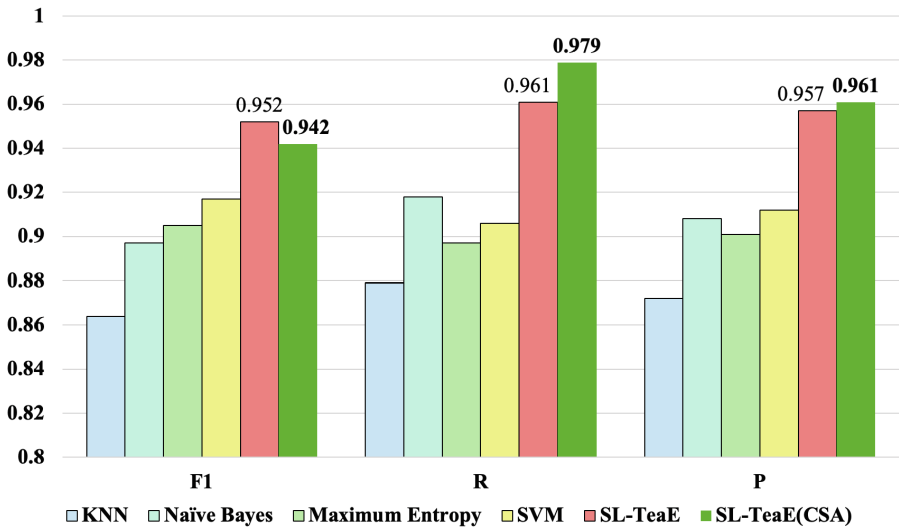


Fig. 2. The Performance of Sentiment Classification on PTC

From Fig. 2 and Fig. 3, the following conclusions can be drawn: 1) SL-TeaE(CSA) Improvement over SL-TeaE: SL-TeaE(CSA), which incorporates complex semantic analysis into SL-TeaE, demonstrates enhanced sentiment classification performance. Specifically, for negative teaching comments, SL-TeaE(CSA) outperforms SL-TeaE with notable improvements in precision, recall, and F1 score by 11.4%, 4.4%, and 7.6%, respectively. For PTC, while precision slightly decreases by 1%, both recall and F1 score show improvement. 2) Comparison with Common Supervised Learning

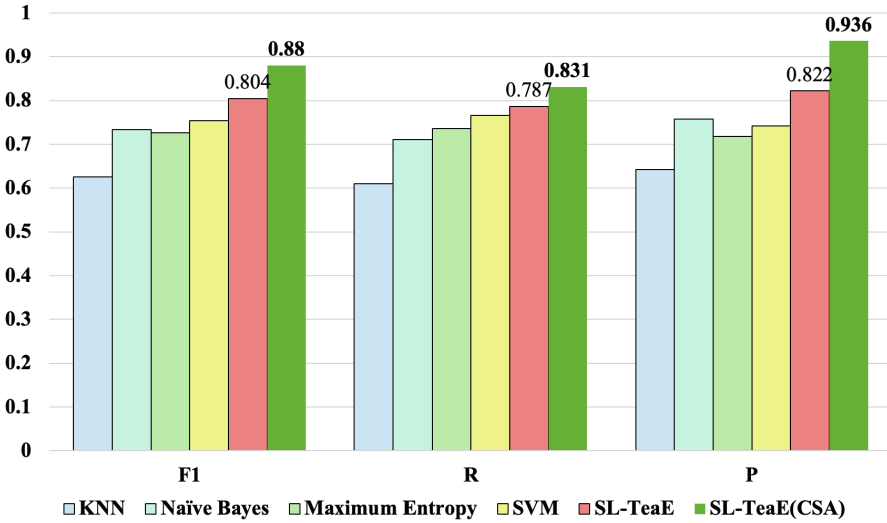


Fig. 3. The Performance of Sentiment Classification on NTC

Algorithms: Compared to common supervised learning algorithms such as KNN, Naïve Bayes, Maximum Entropy, and SVM, SL-TeaE(CSA) exhibits superior sentiment classification performance. It achieves better precision, recall, and F1 score, especially in positive teaching comments where precision increases by 7.8%, recall by 10.0%, and F1 score by 8.9%. For NTC, the highest improvements are observed in precision (19.4%), recall (34.0%), and F1 score (34.9%).

In conclusion, the comparative experiments indicate that SL-TeaE(CSA) excels in sentiment classification performance within the SET domain.

Comparison of Quantitative Evaluation Scores. Through the comprehensive teaching evaluation scores of each course provided by students on the school’s academic administration system, this study compared the scores of the four courses taught by four teachers with the extended general sentiment based on the General Sentiment Lexicon (GSL). Comparative experiments were conducted with Expanded General Sentiment Lexicon (EGSL) and SL-TeaE to verify the accuracy of this model in quantitative teaching evaluation scores. The specific comparison results are shown in Table 1 and Table 2.

Table 1. Comparison of quantitative scores calculation performance

Teacher	GSL	EGSL	SL-TeaE	SL-TeaE(CSA)	Actual scores
Teacher1	73.54	73.96	94.49	96.34	97.54
Teacher2	82.70	83.92	96.56	98.19	98.77
Teacher3	81.73	82.09	93.38	94.27	96.45
Teacher4	83.59	85.45	96.15	97.84	98.10

Table 2. Quantitative scores calculation error comparison

Sentiment computing methods	MAE	RMSE
GSL	17.32	17.76
EGSL	16.36	16.90
SL-TeaE	2.57	2.62
SL-TeaE(CSA)	1.06	1.28

Analysis of Table 1 and Table 2 reveals notable disparities in the course composite assessment scores derived from the GSL when compared to the actual course assessment scores. Notably, these calculated scores exhibit a significant deviation from the correct relative order, displaying the largest mean absolute and root mean square errors. While the course composite assessment scores generated by the EGSL show a relative improvement compared to those derived from the general sentiment lexicon, they still exhibit inaccuracies in their ordering. Conversely, the course composite assessment scores obtained through SL-TeaE and SL-TeaE(CSA) are closer to the actual scores and follow the correct order. Specifically, SL-TeaE(CSA) demonstrates even greater proximity to the true values, boasting the smallest MAE and RMSE of 1.06 and 1.28, respectively.

5 Conclusion

In conclusion, this study explores a specialized emotion lexicon in the field of teaching, enhancing the generalization of the model. By combining complex semantic analysis, the emotional analysis performance of the model has significantly improved. Compared to a general emotion lexicon, the F1 values for positive and negative teaching comments have increased by 7.3% and 34.9%, respectively. The emotional classification performance is also superior to common supervised learning algorithms. Additionally, SL-TeaE(CSA) demonstrates greater accuracy in quantifying evaluation scores, with minimal error in comprehensive course evaluation scores. The model's scores closely align with actual course evaluations, exhibiting consistent ranking with the actual scores. The effectiveness of SL-TeaE(CSA) has been demonstrated.

The experimental results of this article have practical significance for the evaluation of teachers' teaching level, which can help teachers carry out personalized teaching evaluation analysis, such as comparing the teaching levels of different teachers, different courses, and different periods horizontally, and vertically discovering the teaching transitions of teachers in different periods, including the upward period, the teaching pause period, or the teaching decline period. In order to promote teacher teaching reform through student evaluation and improve teaching quality through teaching reform.

Acknowledgements. This study was supported by the Key Project of Regional Innovation and Development Joint Fund of National Natural Science Foundation of China (Grant No. U22A2025).

References

1. Wang, B.: Research on sentiment analysis of student evaluations of teaching based on deep learning. Chongqing University of Posts and Telecommunications (2021)
2. Balahadia, F.F., Fernando, M., Juanatas, I.C.: Teacher's performance evaluation tool using opinion mining with sentiment analysis. In: Region 10 Symposium. IEEE (2016)
3. Lin, Q., Zhu, Y., Zang, S., et al.: Lexical based automated teaching evaluation via students' short reviews. *Comput. Appl. Eng. Educ.* **27**(1), 194–205 (2019)
4. Wang, B., Gao, L., An, T., et al.: A Method of Educational News Classification Based on Emotional Dictionary. In: 2018 Chinese Control and Decision Conference, pp. 1–5. IEEE Press, Shenyang, China (2018)
5. Taboada, M., Brooke, J., Tofiloski, M., et al.: Lexicon-based methods for sentiment analysis. *Comput. Linguist.* **37**(2), 267–307 (2011)
6. Hatzivassiloglou, V., Mckeown, K.R.: Predicting the semantic orientation of adjectives. In: Proceedings of the ACL (2002)
7. Huang, S., Niu, Z., Shi, C.: Automatic construction of domain-specific sentiment lexicon based on constrained label propagation. *Knowl.-Based Syst.* **56**(Jan.), 191–200 (2014)
8. Liu, W., Zhu, Y., Li, C., et al.: Research on building Chinese basic semantic lexicon. *J. Comput. Appl.* **29**(10), 2875–2877 (2009)
9. Yang, C., Feng, S., Wang, D., et al.: Analysis on web public opinion orientation based on extending sentiment lexicon. *J. Chin. Comput. Syst.* **31**(4), 691–695 (2010)
10. Zhou, Y., Yang, J., Yang, A.: A method on building Chinese sentiment lexicon for text sentiment analysis. *J. Shandong Univ. (Eng. Sci.)* (6), 27–33 (2013)
11. Zhang, J., Ning, J., Li, Y.: Data analysis of students' subjective evaluation of teaching based on cluster analysis and its application. *China Educ. Light Ind.* **25**(2), 21–28 (2022)
12. Cai, Y., Yang, K., Zhou, Z., et al.: A hybrid model for opinion mining based on domain sentiment dictionary. *Int. J. Mach. Learn. Cybern.* **10**(8), 2131–2142 (2019)
13. Bollegala, D., Weir, D., Carroll, J.: Using multiple sources to construct a sentiment sensitive thesaurus for cross-domain sentiment classification. In: 49th Annual Meeting of the Association for Computational Linguistics 2011, vol. 1, pp. 132–141. Association for Computational Linguistics (2011)
14. Wawer, A.: Mining co-occurrence matrices for SO-PMI paradigm word candidates. In: 13th Conference of the European Chapter of the Association for Computational Linguistics 2012 (EACL 2012), pp. 74–80. Association for Computational Linguistics (2012)
15. Yang, A., Lin, J., Zhou, Y.: Method on building Chinese text sentiment lexicon. *J. Front. Comput. Sci. Technol.* **11**, 1033–1039 (2013)
16. Gao, H., Zhang, J.: Sentiment analysis and visualization of hotel reviews based on sentiment dictionary. *Comput. Eng. Softw.* **42**(01), 45–47 (2021)
17. Huang, X., Chen, L., Zheng, Y., Guo, H., Shen, F., Gao, H.: SL-TeaE: an efficient method for improving the precision of teaching evaluation. In: Yang, X., et al. (eds.) ADMA 2023. LNCS, vol. 14179, pp. 3–17. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-46674-8_1
18. Liu, Q., Shen, W.: Research of keyword extraction of political news based on Word2Vec and TextRank. *Inf. Res.* **6**, 22–27 (2018)