Research on Intelligent Evaluation and Service Optimization of Q&A Quality in Academic Social Networks (Postprint)

Authors: Yan Weiwei, Huang Wei, warm

Date: 2023-04-01T16:02:48+00:00

Abstract

[Purpose/Significance] The Q&A services provided by academic social networks have become an important channel for scholars to quickly obtain academic information and solve academic problems. Implementing intelligent evaluation of Q&A quality and service optimization based on machine learning is of great significance for the dissemination of high-quality content in academic social networks. [Method/Process] Taking ResearchGate's Q&A service as the research object, this study constructs an answer quality evaluation system from four dimensions: structural features, content features, other features, and answerer features, and utilizes machine learning methods and data augmentation techniques for answer quality classification prediction. [Results/Conclusion] The SMOTE algorithm demonstrates effectiveness in handling imbalanced samples; Support Vector Machine achieves excellent classification performance in single-model prediction; ensemble models further improve prediction accuracy, with the ensemble model constructed based on Random Forest, Support Vector Machine, and BP Neural Network exhibiting the best classification performance. Based on this, optimization of academic social network Q&A services can be achieved by building an intelligent Q&A quality evaluation system.

Full Text

Preamble

Scholars often evaluate answer quality by transforming sentiment polarity into quality assessment features based on data quality frameworks [8] and information quality evaluation criteria [9], ultimately employing RankSVM algorithms to fuse multiple features for predicting answer quality in Chinese Q&A communities [10]. Through a review of relevant research, it becomes evident that when exploring factors influencing answer quality in Q&A communities, domestic and international scholars have not only continued to apply classical theories such as data quality frameworks but have also attempted to integrate different theories and develop novel perspectives to expand and refine answer quality evaluation standards. For instance, Sun et al. [10] empirically constructed a social search answer quality evaluation model from four dimensions: content quality, contextual quality, source quality, and sentiment quality, while D. Ishikawa et al. [11] developed a 12-dimensional evaluation index system for social Q&A community answer quality encompassing respondent experience, evidence sources, politeness, and detail level. Additionally, some scholars have introduced user perception and external cues into answer quality factor research to improve evaluation standards. From the user perception perspective, Wu et al. [12] utilized information architecture theory to evaluate user-generated answer quality in academic Q&A communities from answer, user, and community viewpoints, explaining indicators through emotion theory, cognitive theory, and user behavior. From the external cues perspective, Zhang [13] combined cue theory with user perspectives to identify seven types of external cues—including information utilization cues, information identification cues, and information reporting cues—that influence users' perceived judgment of information quality in social Q&A communities.

Regarding answer quality evaluation research, scholars generally treat answer quality assessment as a machine learning classification problem [4], focusing primarily on "selection and combination of answer quality influencing factors" and "selection and optimization of prediction algorithms." In factor selection and combination, Cai and Kong et al. [14-15] proposed temporal features reflecting answer dynamic characteristics and found that time-based features better predict best answers compared to traditional features. Jiang et al. [16-17] argued that answers in Q&A communities carry sentiment, introducing sentiment-annotated features and social-emotional features into automated answer quality evaluation, with results showing that sentiment features improve classification accuracy. In algorithm selection and optimization, Li et al. [18] studied answer quality on the RG platform and found that optimized SVM algorithms had overwhelming advantages in precision over other models. Guo et al. [16] employed a GA-BP neural network model for automated evaluation, achieving high accuracy and low average error. Beyond classical machine learning methods, some scholars have used deep learning models for answer quality evaluation, such as Vekariya et al. [20] who utilized a global max pooling layer to compress parameters and reduce computation, combined with a DeepLSTM model to predict best answers for given questions, and He et al. [21] who employed BTM topic models to calculate semantic similarity of "question-answer pairs" and BERT models to incorporate answer comments into quality evaluation.

However, existing research has primarily focused on comprehensive Q&A community contexts (e.g., Yahoo! Answers, Quora, Zhihu). While these studies have produced numerous results, the validity of their proposed answer quality evaluation systems and automated evaluation models in various vertical domain Q&A

communities remains to be verified. Users in comprehensive Q&A communities are mostly anonymous with complex backgrounds, whereas academic social networks require users to provide real names and institutional affiliations during registration, ensuring that answers on academic Q&A services are provided by research personnel. This differentiated user group leads to distinct answer characteristics, with academic Q&A services typically containing more complex and professional content that may elaborate on fundamental theories, methods, connotations, and extensions of academic problems [22-23]. Researchers have found that answer quality evaluation criteria for academic Q&A services differ significantly from those for comprehensive Q&A communities [24], necessitating exploration of new features for evaluating answer quality in academic contexts.

Furthermore, current research on answer quality evaluation has concentrated on "features and algorithms," with few scholars exploring prediction accuracy improvement from a data augmentation perspective. Additionally, relevant research lacks guidance for academic social networking platform operators on how to implement automated answer quality evaluation in Q&A services to optimize platform services—issues that require further investigation and discussion.

3 Research Design

ResearchGate (RG) is currently the most popular academic social networking site [25]. Since its launch in 2008, the platform has accumulated over 135 million publications, Q&A posts, and research projects, with more than 17 million users. Compared to other academic social networks, RG has leveraged its large user base and Q&A functionality to accumulate rich Q&A interaction data resources that can support intelligent Q&A quality evaluation research. Given that Q&A quality evaluation typically assesses both answering behavior and content, this study takes RG as its object to conduct empirical research on optimizing academic social network Q&A services based on intelligent Q&A quality evaluation.

Automated evaluation offers comprehensive advantages over manual evaluation in terms of speed, accuracy, and cost, meeting the needs of academic social networking platforms to identify high-quality answers and optimize Q&A services. Therefore, this study employs machine learning methods to implement intelligent Q&A quality evaluation on the RG platform. The research framework consists of three components: evaluation system construction and model selection, data preparation, and modeling/prediction. Evaluation system construction and model selection includes building the answer quality evaluation system with indicator quantification and selecting evaluation models. Data preparation includes data collection, data cleaning (handling missing and abnormal values), data transformation (automated feature extraction), and data annotation (manual labeling to construct a supervised learning problem). Modeling and prediction includes model evaluation and comparison, model optimization, and classification prediction using combined models (see Figure 1 [Figure 1: see original paper]).

3.1 Answer Quality Evaluation System

To ensure the scientific validity and completeness of answer quality evaluation indicators, this study conducted extensive literature review and selected indicators previously confirmed to influence answer quality, constructing a preliminary evaluation system comprising three dimensions: answer structural features, answer content features, and answer other features. The respondent characteristic dimension indicators were further supplemented based on RG platform functionality to build a complete RG answer quality evaluation system (see Table 1).

Table 1 RG Answer Quality Evaluation System

Dimension	Indicator	Explanation	Primary References
Answer Structural Features	Answer character count	Number of characters in answer	[16,26]
	Answer keyword count	Number of keywords in answer	[16-17]
	Answer sentence count	Number of sentences in answer	[16-17]
	Long sentence ratio	Ratio of long sentences to total sentences	[16,26]
	Punctuation ratio	Ratio of punctuation marks to characters	[16-17]
Answer Content Features	Question-answer length ratio	Ratio of question length to answer length	[26-27]
	Question-answer topic similarity	Thematic similarity between question and answer	[15,17]
	Text diversity	Lexical diversity in answer text	[17,26]
	Answer information entropy	Information volume contained in answer	[15,17]

Dimension	Indicator	Explanation	Primary References
	Answer sentiment attitude	Sentiment orientation expressed by respondent	[15,17]
	Answer subjectivity	Degree of subjectivity in answer	[15,26]
Answer Other Features	Answer response order	Position of answer in chronological sequence	[16-17]
	Answer recommendation count	Number of recommendations answer received	[16-17]
Respondent Characteristics	RG-score	RG's metric for evaluating users	[15,17]
	Personal likes	Total likes received by respondent	[15,17]
	Citation count	Total citations of respondent's publications	[15,17]
	Read count	Total views of respondent's publications	[16-17]
	Research project count	Number of research projects posted	[16-17]
	Answer count	Number of questions answered	[16-17]
	Question count	Number of questions asked	[16-17]
	Profile picture	Whether respondent has profile picture	[16-17]
	Publication count	Number of research materials posted/claimed	[16-17]

3.1.1 Answer Structural Features Structural features refer to characteristics that can be directly obtained through answer statistics. Similar to traditional web resources, answers on academic social networking platforms are primarily presented in text format. Therefore, relevant indicators and methods applicable to traditional web quality analysis can be applied to answer quality evaluation in academic Q&A services, including text length, keyword count, and punctuation ratio. Quantification of structural features is relatively simple; using segmentation tools and text processing techniques, Python programs can directly extract and quantify these indicators from answer texts.

3.1.2 Answer Content Features Content features refer to characteristics embedded in text that require natural language processing to manifest. Quantification methods include:

(1) **Question-Answer Topic Similarity**: High-quality answers should belong to the same topic as their corresponding questions [26]. LDA topic models are used to calculate topics for both questions and answers separately, followed by measuring topic similarity.

(2) **Text Diversity**: Text diversity is quantified by lexical diversity in answer text [27]. Fewer average occurrences of vocabulary indicate stronger lexical diversity. The calculation method is shown in Formula (1), where Ti represents the occurrence count of each word in the answer.

$$D_{\text{diversity}} = \frac{\text{Total word count}}{\sum_{i=1}^{n} T_i^2} \quad \text{Formula (1)}$$

(3) **Answer Information Entropy**: From an information dissemination perspective, information entropy represents information value. High-quality answers contain valuable information, making information entropy a proxy for answer quality [26]. Answer information entropy is calculated using Formula (2), where Pi represents the probability of each letter's occurrence.

$$H_{\text{information entropy}} = -\sum_{i=1}^{n} P_i \cdot \log_2 P_i \quad \text{Formula (2)}$$

(4) **Answer Sentiment Attitude and Subjectivity**: High-quality answers exhibit stronger sentiment tendencies than average answers [16], and respondent attitudes influence answer recognition [26]. Sentiment analysis and subjectivity judgment are performed using the TextBlob Python library, with numerical outputs quantifying these features.

3.1.3 Answer Other Features Q&A services involve time-sensitive questions; questioners typically expect prompt responses, making answer value negatively correlated with response time. This study introduces "answer response order" as an indicator reflecting timeliness, with all answers sorted chronologically in ascending order. Answer recognition (recommendation count) reflects answer value—higher recommendation counts indicate greater likelihood of high quality. "Answer recommendation count" is included as an indicator of user recognition, obtainable via web scraping from RG pages. Since neither "answer response order" nor "answer recommendation count" fit into other dimensions, they are grouped as answer other features.

3.1.4 Respondent Characteristics Respondent characteristics focus on RG platform users' own attributes in Q&A interactions. Typically, opinion leaders or more professional expert users provide higher-quality answers that receive greater user support [15,17]. To extract features reflecting respondent influence, this study combines available RG user metrics and references the opinion leader identification model ENIA [28], using RG-score, RI value, citation count, personal likes, read count, and project count as indicators of respondent professionalism, activity, and influence. These metrics can be directly obtained and quantified from respondent profile pages.

3.2 Model Selection

This study selects ID3 Decision Tree (Iterative Dichotomiser 3), Random Forest (RF), Support Vector Machine (SVM), and BP Neural Network—common models in automated answer quality evaluation tasks—as baseline candidate models. ID3, as a classic classification algorithm, offers moderate performance but fast execution and strong interpretability, meeting baseline prediction requirements. When a classification model outperforms ID3, it is considered valuable and eligible for combined model prediction. Considering RF's suitability for relatively low-dimensional data (dozens of dimensions) with high accuracy requirements, and SVM's strong performance on small-sample machine learning problems, both are selected as classification prediction models, aligning with this study's low-dimensional (23 features) and small-sample Q&A data characteristics. Additionally, BP, as one of the most common deep learning algorithms, possesses powerful feature fitting capabilities [29] suitable for situations with intrinsic feature relationships. Since this study retained as many variables as possible during evaluation system construction to avoid information loss from dimensionality reduction—resulting in some redundancy and correlation among evaluation indicators—BP is selected as a classification model, matching the characteristic of intrinsic relationships among Q&A data features.

4 Empirical Study on Intelligent Evaluation of Academic Social Network Q&A Quality

4.1 Data Preparation

4.1.1 Data Acquisition and Preprocessing This study collected all answers and respondent data before July 6, 2020, for the question "Can technology replace a teacher?" under the AI theme on RG Q&A service, totaling 3,873 entries. Considering manual annotation costs, systematic sampling was employed—starting from the first answer with a sampling interval of 200 entries—to select 2,000 answers as raw data. After data cleaning to remove missing values, abnormal values, and garbled data, 1,670 usable entries remained for model construction. Python automation programs were then used to quantify each indicator.

4.1.2 Data Annotation Based on existing Q&A community answer quality factor research [30-33], answers were evaluated and labeled for quality across five dimensions: practicality, relevance, completeness, readability, and persuasiveness. Four experienced RG Q&A users were invited as annotators, trained to establish consistent evaluation standards before manually performing binary labeling (1 for high quality, 0 for low quality). The final labeled dataset contained 251 high-quality samples and 1,419 low-quality samples—a ratio approaching 1:6.

4.2 Model Evaluation and Comparison

This study implemented RG Q&A service answer quality automated evaluation models using Python 3.7 and libraries including keras, sklearn, and imblearn. The dataset was automatically partitioned into training (80%) and testing (20%) sets. After training ID3, RF, SVM, and BP algorithms on the training set and optimizing parameters, model performance was compared using precision, recall, and F1-score on the test set. Precision represents the proportion of correctly classified positive samples among all predicted positive samples, while recall represents the proportion of correctly classified positive samples among all actual positive samples. Both metrics are positively correlated with classifier performance. F1-score is the harmonic mean of precision and recall, providing a more comprehensive evaluation [34].

As shown in Table 2 , all four models exhibited class imbalance bias when classifying RG answer quality (classifiers tended toward severely imbalanced accuracy—high for the majority class but extremely low for the minority class), resulting in prediction failure and inability to detect high-quality answers (low precision, recall, and F1-score). This issue critically undermines intelligent Q&A quality evaluation in academic social networks, as the ultimate goal is to screen and display high-quality answers. Typically, class imbalance is addressed through undersampling or oversampling. Oversampling increases minority class samples to approximate the majority class size [35], suitable for small, imbalanced datasets like this study's data. To enable high-quality answer screening, this study employs oversampling techniques to address label imbalance before prediction.

4.3 Model Optimization Based on SMOTE Algorithm

This study adopts the SMOTE (Synthetic Minority Oversampling Technique) algorithm [36] to resolve label imbalance. SMOTE improves upon random oversampling by using KNN (K-Nearest Neighbor) technology to generate new samples rather than simply replicating original samples, producing more representative synthetic samples. Table 3 compares training/test sets before and after SMOTE oversampling.

Table 3 Training and Test Sample Partitioning

Dataset	Before SMOTE Balancing	After SMOTE Balancing
	High-quality	Low-quality
Training Set	201	1,135
Test Set	50	284

After retraining the four classification models on the SMOTE-balanced dataset, performance analysis (Table 4) shows significant improvements in precision and recall for high-quality answer classification, meeting screening requirements. Specifically, ID3 performs poorly compared to other algorithms, while RF and BP neural networks show similar and relatively strong performance. SVM demonstrates the best performance with the highest precision, recall, and F1-score.

Table 4 Performance Evaluation of Four Classification Models (After Optimization)

Model	Precision	Recall	F1-score	Support
ID3	0.72	0.68	0.70	334
Random Forest	0.89	0.91	0.90	334
SVM	0.94	0.95	0.94	334
BP Neural Network	0.90	0.92	0.91	334

4.4 Combined Model Prediction

To leverage different models' strengths, researchers often combine them for prediction—a concept proposed by J.M. Bates and C.W.J. Granger in 1969 and widely applied in machine learning. Since RF, SVM, and BP neural networks all outperform ID3, they participate in combined prediction through pairwise combinations (RF+BP, SVM+BP, SVM+RF) and full-model combination (SVM+RF+BP).

Combination prediction uses "dynamic weighting." For pairwise combinations, given Model 1's prediction $M1$ and Model 2's prediction $M2$, dynamic weights

$w1$ and $w2$ are assigned ($w1 + w2 = 1$, with $w1$ taking nine incremental values from 0.1 to 0.9). The combined prediction Y is calculated as:

$$Y = w_{1M}1 + w_{2M}2 \quad \text{Formula (3)}$$

Similarly, full-model combination uses dynamic weighting across three models. As weighting proportions change, optimal weighting yields the best prediction results. Table 5 shows optimal weighting results for four combination models.

Table 5 Performance Evaluation of Combined Models

Combination Model	Optimal Weighting	Precision	Recall	F1-score	Support
RF + BP	0.6RF + 0.4BP	0.91	0.93	0.92	334
SVM + BP	0.7SVM + 0.3BP	0.94	0.95	0.94	334
SVM + RF	0.5SVM + 0.5RF	0.93	0.94	0.93	334
SVM + RF + BP	0.4SVM + 0.2RF + 0.4BP	0.96	0.96	0.96	334

Results demonstrate that combined prediction significantly outperforms single-model prediction. The SVM+RF+BP combination at a 4:2:4 ratio achieves the best performance with an F1-score reaching 96%, showing the strongest generalization capability. Thus, combined answer quality prediction models effectively screen high-quality answers and can be applied to intelligent Q&A quality evaluation in academic social networks.

5 Q&A Service Optimization Based on Intelligent Quality Evaluation

Building on empirical results, this study proposes an intelligent Q&A quality evaluation system architecture for RG (Figure 2 [Figure 2: see original paper]) to optimize academic social network Q&A services. The system comprises three components: Q&A data collection and processing, answer quality evaluation, and high-quality answer display. The first two components form the foundation for automated answer quality evaluation, mining quality responses, while the third component focuses on high-quality answer display functionality, directly driving user experience improvements.

Figure 2 [Figure 2: see original paper] RG Q&A Quality Intelligent Evaluation System Architecture

5.1 Q&A Data Collection and Processing

The system connects to RG's Q&A database to automatically acquire answer texts, response times, respondent information, and other Q&A data. Scattered and disordered data undergo cleaning, transformation (automated feature extraction), and other processing steps to form raw data suitable for quality evaluation.

5.2 Answer Quality Evaluation

The optimal combined model validated earlier serves as RG's automated answer quality evaluation model. Q&A data exported from the database is input into the combined prediction model, which automatically predicts and labels answer quality through computer programs.

5.3 High-Quality Answer Display

High-quality answer display is subdivided into three sub-components: "quality answer display," "quality content push mechanism," and "respondent incentive mechanism," focusing on improving user experience for both information recipients (lurkers and questioners) and information providers (respondents).

The system architecture can similarly guide other academic social networking platforms' Q&A service optimization. Particularly for domestic academic social networks still in early development, implementing intelligent Q&A quality evaluation systems helps control information quality issues early, prevent negative impacts, mine quality content to enhance service quality, and promote platform development. Specifically, platforms can create a "quality answers" section on each question page to display automatically screened high-quality answers, enabling lurkers and questioners to access potentially useful responses immediately, saving search time and improving browsing experience. Since individuals inherently desire recognition and appreciation [37], publicly displaying users' quality answers affirms their contributions on academic social networks, stimulating creation enthusiasm and increasing participation in academic Q&A services. Building on quality content mining, platforms can further establish a "quality content recommendation mechanism" that regularly pushes interesting questions with quality answers to users based on their professional backgrounds and research interests, facilitating rapid quality content dissemination. Additionally, since attention or recognition received after answering positively influences respondents' willingness to participate [38], implementing a "quality answer incentive mechanism" using community points or honor badges when answers are identified as high-quality can stimulate respondents' sense of identity and achievement, promoting sustained participation.

6 Conclusions and Limitations

This study employs machine learning for automated answer quality evaluation to advance intelligent Q&A quality assessment and promote information quality improvement and service optimization in academic social networks. Empirical results demonstrate the feasibility and rationality of screening quality content (high-quality answers) in academic social network Q&A services from four perspectives: answer structural features, content features, other features, and respondent characteristics. Using machine learning methods combined with data augmentation techniques like SMOTE and combined model construction effectively screens high-quality answers. Additionally, this study proposes a holistic approach for designing and implementing intelligent Q&A quality evaluation systems, providing references for RG and other platforms to optimize Q&A services. The optimization solution meets users' information quality expectations through quality content display, builds community identity and achievement for respondents, incentivizes sustained participation, and ultimately enhances user stickiness on both ends of Q&A services to achieve a virtuous cycle.

This research has limitations. First, it only uses AI-themed Q&A data to validate the evaluation system and automated model's effectiveness and rationality; the relatively single topic requires further validation across broader themes. Second, conclusions and optimization solutions based on RG platform data need validation on other academic social networks or Q&A communities. Third, this study aims to reveal Q&A service optimization paths from the platform perspective; future research could further investigate answer quality from the respondent's perspective.

References

[1] Li Lei. Research on Answer Quality Assessment in Academic Social Q&A Platforms [D]. Nanjing: Nanjing University of Science and Technology, 2018.

[2] Shah C, Oh JS, Oh S. Exploring characteristics and effects of user participation in online social Q&A sites [J/OL]. First Monday, 2008, 13(9). [2020-09-25]. https://firstmonday.org/article/view/2182/2028.

[3] Zhang Ning, Yuan Qinjian. Governance and improvement of information quality in academic social networks [J]. Library and Information Service, 2019, 63(23): 79-86.

[4] Jiang Wen, Xu Xin. Review of online Q&A community information quality evaluation research [J]. New Technology of Library and Information Service, 2014(6): 41-50.

[5] Yao Y, Tong H, Xie T, et al. Detecting high-quality posts in community question answering [J]. Information Sciences, 2015, 302(1): 70-82.

[6] Wang RY, Storey VC, Firth CP. A framework for analysis of data quality research [J]. IEEE Transactions on Knowledge & Data Engineering, 1995, 7(4): 623-640.

[7] Tenopir C, Levine K, Allard S, et al. Trustworthiness and authority of scholarly information in a digital age: Results of an international questionnaire [J]. Journal of the Association for Information Science & Technology, 2016, 67(10): 2344-2361.

[8] Wang RY, Strong DM. Beyond accuracy: What data quality means to data consumers [J]. Journal of Management Information Systems, 1996, 12(4): 5-33.

[9] American Public Health Association. Criteria for assessing the quality of health information on the Internet [J]. American Journal of Public Health, 2001,

91(3): 513-514.

[10] Sun Xiaoning, Zhao Yuxiang, Zhu Qinghua. Construction of answer quality evaluation indicators for social search based on SQA systems [J]. Journal of Library Science in China, 2015, 41(4): 65-82.

[11] Daisuke Ink, Sakai OT. What makes a good answer in community question answering? An analysis of assessors' criteria [EB/OL]. [2021-01-05]. https://www.researchgate.net/publication/228449185.

[12] Wu Yawei, Zhang Xiangxian, Tao Xing, et al. Construction of answer quality evaluation indicators in academic Q&A communities based on user perception [J]. Information Science, 2020, 38(10): 141-147.

[13] Zhang Yuxuan. Research on answer information quality perception in social Q&A platforms based on external cues [D]. Wuhan: Central China Normal University, 2016.

[14] Cai YZ, Chakravarthy S. Predicting answer quality in Q/A social networks: Using temporal features [R]. Arlington: University of Texas at Arlington, 2011.

[15] Kong Weize, Liu Yiqun, Zhang Min, et al. Answer quality evaluation method in Q&A communities [J]. Journal of Chinese Information Processing, 2011, 25(1): 3-8.

[16] Jiang Wen, Xu Xin, Wu Gaofeng. Automated evaluation of online Q&A community information quality with added emotional features [J]. Library and Information Service, 2015, 59(4): 100-105.

[17] Guo Shunli, Zhang Xiangxian, Tao Xing, et al. Automated evaluation of user-generated answer quality in social Q&A communities: A case study of Zhihu [J]. Library and Information Service, 2019, 63(11): 118-130.

[18] Li L, He D, Jeng W, et al. Answer quality characteristics and prediction on an academic Q&A site: A case study on ResearchGate [C]//24th International Conference on World Wide Web. Florence: ACM, 2015: 1453-1458.

[19] Le LT, Shah C, Choi E. Assessing the quality of answers autonomously in community question-answering [J]. International Journal on Digital Libraries, 2019, 20(4): 351-367.

[20] Vekariya DV, Limbasiyan R. A novel approach for semantic similarity measurement for high-quality answer selection in question answering using deep learning methods [C]//6th International Conference on Advanced Computing and Communication Systems (ICACCS). Coimbatore: IEEE, 2020: 518-522.

[21] He Xun. Research on answer quality prediction in online Chinese Q&A communities [D]. Jinan: Qilu University of Technology, 2020.

[22] Goodwin S, Jeng W, He D. Changing communication on ResearchGate through interface updates [EB/OL]. [2021-01-05]. https://www.researchgate.net/publication/273664849.

[23] Li L, Zhang C, He D. Factors influencing the importance of criteria for judging answer quality on academic social Q&A platforms [J]. Aslib Journal of Information Management, 2020, 72(6): 887-907.

[24] Li L, Zhang C, He D, et al. Researchers' judgment criteria of high-quality answers on academic social Q&A platforms [J]. Online Information Review, 2020, 44(3): 603-623.

[25] Ren Pingping. ResearchGate's internationalization of academic social networking [J]. International Talent Exchange, 2020(5): 52-53.

[26] Wang Wei, Ji Yuqiang, Wang Hongwei, et al. Evaluation of answer quality in Chinese Q&A communities: A case study of Zhihu [J]. Library and Information Service, 2017, 61(22): 36-44.

[27] Li Zhan, Chao Wenhan, Chen Xiaoming, et al. Question and answer quality evaluation and prediction in Chinese community Q&A [J]. Computer Science, 2011, 38(6): 230-236.

[28] Li Y, Ma S, Zhang Y, et al. An improved mix framework for opinion leader identification in online learning communities [J]. Knowledge-Based Systems, 2013, 43(2): 43-51.

[29] Liu Yongheng. Research on automobile sales forecasting based on neural networks and time series [D]. Nanchang: Nanchang University, 2019.

[30] Zhu ZM, Bernhard D, Gurevych I. A multi-dimensional model for assessing the quality of answers in social Q&A sites [EB/OL]. [2020-09-30]. https://tuprints.ulb.tu-darmstadt.de/1940/.

[31] Jia Jia, Song Enmei, Su Huan. Answer quality assessment in social Q&A platforms: A comparative study of Zhihu and Baidu Knows [J]. Journal of Information Resources Management, 2013, 3(2): 19-28.

[32] Liu Yirong, Liu Yun. Answer aggregation and optimization in Q&A platforms: A case study of Quora [J]. Research on Library Science, 2017(6): 48-56, 13.

[33] Yuan Hong, Zhang Ying. Answer quality evaluation in Q&A communities: A comparative study of Baidu Knows and Zhihu [J]. Digital Library Forum, 2014(9): 43-49.

[34] Zhou Zhihua. Machine Learning [M]. Beijing: Tsinghua University Press, 2016.

[35] Li Xiaoxiao. Application of random forest method in personal credit risk analysis [D]. Kunming: Yunnan University, 2019.

[36] Zhou Qi. Research on personal credit risk assessment algorithms for imbalanced data [D]. Baoding: Hebei University, 2020.

[37] McLaughlin G. SMOG grading—a new readability formula [J]. Journal of Reading, 1969, 12(8): 639-646.

[38] Zhang Haitao, Sun Tong, Zhang Xinrui, et al. Research on the dynamic mechanism of user role transformation in social Q&A communities [J]. Journal of Modern Information, 2020, 40(9): 32-41.

Author Contributions: Yan Weiwei: Proposed research ideas, determined topic, revised paper; Huang Wei: Proposed research framework, data analysis, wrote paper; Wen Xin: Responsible for data acquisition and programming.

Author Affiliations: School of Information Management, Wuhan University, Wuhan 430072

Abstract: [Purpose/significance] The Q&A service provided by academic social networking sites has become an important way for scholars to access academic information quickly and solve academic problems. It is of great significance for the dissemination of high-quality content in academic social networking sites to implement the intelligent evaluation of Q&A quality and the service optimization based on machine learning. [Method/process] This paper took ResearchGate as the research object, constructed an answer quality evaluation system based on four dimensions of structural features, content features, respondent characteristics and other characteristics of answers, and then used machine learning methods and data augmentation technology to perform the automatic answer quality classification prediction. [Result/conclusion] The results show that SMOTE algorithm is effective in dealing with unbalanced samples; In the first mock exam, support vector machine (SVM) achieves excellent classification performance; The combined model can further improve the prediction accuracy, and the combined model based on random forest, SVM and BP neural network has the best classification performance. On this basis, the academic social network Q&A service can be optimized by building the intelligent quality evaluation system.

Keywords: answer quality evaluation; Q&A service; academic social networking site; machine learning

Note: Figure translations are in progress. See original paper for figures.