

Title	Estimating Text Concreteness in Online Discussions
Author(s)	胡, 明熹
Citation	
Issue Date	2025-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/19794
Rights	
Description	Supervisor: 長谷川 忍, 先端科学技術研究科, 修士 (情報科学)

This study explores the estimation of abstraction-concreteness (AC) scores in online discussion texts, aiming to improve the detection of high-quality contributions. With the increasing prevalence of online forums as platforms for discussion, understanding and evaluating the quality of user-generated content has become a significant challenge. One crucial aspect of discussion quality is concreteness, which influences readability, engagement, and effective communication. Concreteness refers to the extent to which a text includes specific, detailed, and vivid information, as opposed to abstract and generalized statements. However, prior research has primarily focused on sentence- or word-level concreteness, with limited exploration of how entire comments in online discussions exhibit concreteness. Online forums facilitate knowledge exchange but vary significantly in content quality. Understanding how concreteness impacts discussion effectiveness is critical for enhancing content filtering and recommendation systems. Prior research has examined various linguistic factors affecting text quality, yet a systematic approach to estimating AC in user-generated content remains underdeveloped. This research seeks to bridge this gap by constructing predictive models for AC scores and investigating five supporting dimensions—Actionability, Clarity, Orientation, Relevance, and Specificity—as key factors influencing AC.

To achieve these objectives, a dataset was constructed by collecting comments from the ExplainLikeImFive (ELI5) subreddit, a platform where users explain complex topics in simple terms. The dataset includes human-annotated AC scores and supporting dimensions, obtained through Amazon Mechanical Turk (MTurk). A rigorous filtering process was applied to improve inter-rater consistency, ensuring that the annotations reflect a shared understanding of the abstract-concrete spectrum. The study employed a combination of human annotations, dictionary-based estimates, and machine learning models to analyze and predict AC scores. The research methodology involved three primary modeling approaches: (1) rule-based weighting using Pearson correlations, (2) linear regression with five supporting dimensions as predictors, and (3) feature extraction models based on TF-IDF regression and GPT-4 few-shot prompting. Additionally, a baseline estimate was derived using word concreteness scores from psycholinguistic dictionaries to provide a reference for model performance. The dataset underwent pre-processing, including filtering to remove inconsistencies and ensuring that only high-agreement annotations were retained for model training.

Experimental results indicate that models leveraging structured feature inputs outperform text-based approaches. The linear regression model demonstrated the highest accuracy, achieving the lowest Mean Absolute Error (MAE = 0.34) and Root Mean Squared Error (RMSE = 0.41), while also being the only model with a positive R-squared value ($R^2 = 0.24$). This suggests that integrating structured linguistic features significantly enhances predictive performance. In contrast, text-based models, including TF-IDF regression and GPT-4 predictions, exhibited significantly higher errors, with GPT-4 generating particularly inconsistent scores. The findings confirm that concreteness perception extends beyond word-level features and is better captured through structured linguistic dimensions rather than raw textual analysis alone.

Further analysis of the supporting dimensions revealed that Clarity and Specificity exhibited stronger correlations with AC, suggesting that well-defined and detailed content is perceived as more concrete. Actionability, while relevant, showed weaker correlations with AC, indicating that while providing actionable advice contributes to text quality, it does not necessarily align with higher concreteness perceptions. The study also assessed the feasibility of automatically extracting these dimensions from raw text. TF-IDF regression consistently outperformed GPT-4 across all dimensions, particularly in Clarity and Specificity. However, GPT-4 showed substantial inconsistencies, especially in Relevance and Orientation, indicating that while large language models capture general linguistic patterns, they struggle with nuanced text quality assessments without domain-specific fine-tuning.

The implications of this research extend beyond theoretical linguistics, offering practical applications for improving online discourse. By integrating AC estimation models into discussion platforms, moderators can enhance content ranking mechanisms, prioritizing well-structured and informative comments. Additionally, automated concreteness scoring could benefit educational applications, helping instructors assess the clarity and specificity of student responses in discussion-based learning environments. Content creators and writers can also leverage AC scoring to refine their writing, ensuring that their work is engaging and easy to understand. Furthermore, AC scoring could be integrated into automated writing assistants and feedback systems, improving real-time text suggestions and readability evaluations.

One of the broader applications of AC estimation lies in misinformation detection and fact-checking. More concrete statements often contain verifiable information, whereas highly abstract statements may be more prone to misinterpretation or fabrication. By assessing concreteness levels in social media discourse, AC scoring could serve as an additional layer of verification for content credibility assessments. Similarly, in corporate and legal communication, ensuring high concreteness can aid in drafting clearer policies and legal documents, reducing ambiguities that may lead to misinterpretation. Another key application is in search engine optimization and content recommendation, where highlighting concrete and informative content can improve user engagement and knowledge retention.

Future research should refine the definitions of AC and its supporting dimensions to improve annotation consistency and model interpretability. Expanding the dataset to include diverse text genres, such as scientific literature, journalistic writing, and instructional materials, would enhance model generalization. Additionally, analyzing cross-linguistic differences in AC perception could provide valuable insights into how concreteness varies across cultures and languages. Investigating how contextual features, such as the discourse structure of comments, affect AC ratings could also improve model robustness. Further, deep learning techniques, including fine-tuned transformer models like BERT, could be explored to improve prediction accuracy by leveraging contextual embeddings. Hybrid models that integrate linguistic features with neural networks may provide a balanced approach, combining interpretability with predictive power.

Another promising avenue for future work is refining automated annotation techniques to reduce reliance on human labeling. Leveraging active learning strategies, where models select the most uncertain samples for human review, could improve annotation efficiency while maintaining high-quality data. Additionally, evaluating how AC interacts with sentiment, engagement metrics, and user trustworthiness could further enhance our

understanding of how concreteness contributes to effective communication in online discussions. Another direction worth exploring is how AC levels correlate with engagement metrics such as comment popularity, upvotes, and response rates, which could provide additional insights into the impact of concreteness on online interactions.

Finally, the study highlights the need for a broader discussion on how concreteness influences communication effectiveness across different domains. Future research could examine its role in persuasive writing, policymaking, and legal discourse, where clarity and specificity are crucial for effective information dissemination. By further refining computational methods for AC estimation, this research contributes to the broader field of natural language processing, fostering more structured and meaningful interactions in digital communication environments. Additionally, practical implementations of AC scoring in education, journalism, and content moderation could be explored to create user-friendly tools that assist in generating more effective and engaging textual communication. Such advancements would reinforce the importance of concreteness in knowledge dissemination and digital interaction, paving the way for further research on optimizing communication strategies through computational analysis.