氏　　名	Wei Wenqing		
学位の種類	博士（情報科学）		
学位記番号	博情第 548 号		
学位授与年月日	令和 7 年 3 月 21 日		
論文題目	STUDY ON MULTIMODAL USER IMPRESSIONS RECOGNITION FOR NON-TASK DIALOGUE SYSTEMS		
論文審査委員	岡田　将吾	北陸先端科学技術大学院大学	教授
	長谷川　忍	同	教授
	井之上　直也	同	准教授
	白井　清昭	同	准教授
	熊野　史朗	NTT 基礎科学研究所	主任研究員

論文の内容の要旨

With the continuous development of human-computer interaction technology, the application of dialogue systems in various fields is becoming increasingly widespread. Among them, non-task-oriented dialogue systems have shown great potential in fields such as chatbots and open-domain dialogue systems. While improving the quality of non-task-oriented dialogue systems is crucial for enhancing user interactions. A high-quality dialogue system must not only understand user intent but also generate accurate and natural responses. This necessitates a robust evaluation framework to assess the system's capabilities. Additionally, evaluation serves as the foundation for system improvement and optimization. Through evaluation of the dialogue system, weaknesses in the system can be identified, making it easier to fine-tune models, data, or algorithms to improve overall performance.

With the rise of multimodal dialogue systems, the demand for their evaluation has also increased. However, existing evaluation methods for dialogue systems often focus solely on text-to-text interactions, neglecting the importance of multimodal data in dialogue systems. In contrast, unimodal systems typically rely only on language content or speech intonation, which may lead to neglect or misinterpretation of users' emotions. Multimodal information, such as speech intonation, facial expressions, and body movements, can better capture users' emotional changes. Therefore, utilizing multimodal information to evaluate multimodal dialogue systems is crucial.

Moreover, text-based evaluation metrics, such as BLEU and ROUGE, are insufficient for assessing multimodal dialogue systems. At the same time, existing multimodal databases face limitations in data collection, particularly in collecting speech and image data, resulting in incomplete and limited evaluation methods. Motivated by these challenges, this research aims to address data collection issues in the evaluation of multimodal non-task-oriented dialogue systems and propose innovative evaluation methods. By collecting, organizing, and utilizing multimodal data, we aim to evaluate dialogue system performance more comprehensively and accurately, thereby enhancing user experience and impressions. Therefore, this research has significant

theoretical and practical implications and will make important contributions to the development of the field of multimodal dialogue system evaluation.

Above all, to establish an automated, robust, and accurate model for evaluating multimodal dialogue systems. Firstly, we introduce a method for identifying user satisfaction at the dialogue level, filling a gap in previous research. We use a method based on multimodal modeling, which comprehensively considers various information such as text, speech, and images to evaluate dialogue system performance more comprehensively. Then, we utilize deep learning models to comprehensively analyze user satisfaction at the dialogue level and user impressions at the exchange level, enhancing the accuracy and reliability of the evaluation methods. Through experimental evaluation, we confirm the effectiveness and feasibility of the proposed methods in the field of multimodal dialogue system evaluation, providing new insights and methods for further research in this area.

The user impression can be analyzed and evaluated at two levels: the exchange level and the dialogue level. These two levels are closely interconnected but differ in focus, making them well-suited for capturing information at different hierarchical levels. While the relationship between user impressions at the dialogue level and user sentiments at the exchange level is secondly explored, which proposes a multi-task learning model that comprehensively considers information from both levels. By analyzing the relationship between 18 dialogue labels and user sentiment during dialogue exchanges and utilizing multi-task learning models, we successfully achieve accurate identification of user impressions at the dialogue level, bringing new insights and methods to the field of dialogue system evaluation.

Lastly, we address the issue of existing methods neglecting the influence of users' personal information which included age, gender, and personality on their impressions by proposing a model based on adversarial learning. By reversing the gradient direction during training, our network learns adversarial features that remain consistent across different users' personal information domains, effectively mitigating the influence of users' personal information and making the model applicable to evaluate non-task-oriented dialogue systems. Through experimental validation, we confirm the effectiveness and feasibility of the proposed method, providing new insights and methods for further research in the field of dialogue system evaluation.

In conclusion, this study proposes novel approaches to addressing the evaluation challenges encountered by multimodal non-task-oriented dialogue systems. The proposed methods improve the accuracy and comprehensiveness of dialogue system evaluation, offering valuable insights for enhancing user experience and satisfaction in various applications.

Keywords: Dialogue system, Multimodal, Evaluation, User impression, User traits adaptation.

論文審査の結果の要旨

Wei 氏は知的対話システムのユーザ評価の自動獲得手法に関して先駆的な研究を行った．近年，ユーザへの情報提供・推薦といった特定のタスクを指向する対話に加え，高度な言語処理能力の向上により，対話自体を楽しむオープンドメイン（特定のタスクに特化しない雑談などの自由対話）の対話システムに関する研究が盛んに行われている．一方で，オープンドメイン対話は明確な対話ゴールがないためシステムの質を評価することが難しく，対話評価はユーザによる主観評価に依存していた．そこで，Wei 氏はユーザの対話中に表出する表情・音声などの非言語情報と，発話言語情報からユーザの対話システムに対する主観評価を獲得する方法に着眼した．まず，第一に Wei 氏はユーザの主観評価の推定モデルの定式化を行った．システムとの対話中に観測される言語，音声，表情・動作の時系列データから対話後に取得されたユーザの主観評価を精緻に推定する機械学習モデル手法を探究した．設計された LSTM モデルは「よく調整された対話である」という評価の高・低を 75%で推定できることが示された．第二に，主観評価スコアの推定精度を向上させるために，発話ごとに変化するセンチメント（心象）ラベルを援用するマルチタスク学習手法を新規に提案し，全ての評価項目に対して推定精度を向上させることに成功した（平均 9%，最高 15.7%の改善）．

最後に，ユーザの個性が対話に与える影響を追究した．話者の年齢・性別・性格等は対話の仕方に影響を与える．これらの個人差要因により汎用的な主観評価推定が困難になる課題に対して，敵対的学習手法を導入することで精度低下を緩和することができることを示した．以上 3 つの研究成果で校正された本論文は，マルチモーダル対話システムの自動評価モデルの機械学習モデルの設計方法を提案し，さらに主観評価推定における特有の課題である個人差，時系列ラベルが不足することに起因する精度低下の課題を解決する方法を提案しており，学術的に貢献するところが大きい．

総合的には，他の関連手法との比較が不十分なことや，いくつかの限界があるが，研究業績も十分にあり，当該分野の発展に寄与する成果と認められるため，合格水準に達していると総合的に判断する．よって博士（情報科学）の学位論文として十分価値あるものと認めた．