



TITLE:

Task and User Adaptation based on Character Expression for Spoken Dialogue Systems(Abstract_要旨)

AUTHOR(S):

Yamamoto, Kenta

CITATION:

Yamamoto, Kenta. Task and User Adaptation based on Character Expression for Spoken Dialogue Systems. 京都大学, 2023, 博士(情報学)

ISSUE DATE:

2023-03-23

URL:

<https://doi.org/10.14989/doctor.k24728>

RIGHT:

(続紙 1)

京都大学	博士 (情報学)	氏名	山本 賢太
論文題目	Task and User Adaptation based on Character Expression for Spoken Dialogue Systems (音声対話システムのためのキャラクタ表現に基づくタスク・ユーザ適応)		
(論文内容の要旨)			
<p>Many spoken dialogue systems such as smart speakers and conversational robots are used in human society. Conversational robots or agents are given various social roles such as a museum guide and an interviewer. It is important for the system to establish a relationship with the user in order to be used continuously over a long time. One of the approaches to this problem is to define a character of the spoken dialogue system. If the user feels a desirable character impression from behaviors of the system, they would be attached to the system. However, the desirable character expression should be dependent on the dialogue task and the user. The goal of this study is to make the spoken dialogue system express the appropriate character according to the dialogue task and the user.</p> <p>This thesis presents a character expression model for spoken dialogue systems based on the following behaviors specific to the spoken dialogue: utterance amount, backchannels, fillers, and switching pause length. The model controls these behaviors of the spoken dialogue system to give the intended character impression. This study investigates two approaches to the question of the appropriate character for the spoken dialogue system. One is task adaptation, where the system expresses the appropriate character for the task. The other is user adaptation, where the system expresses the appropriate character for the user personality.</p> <p>Chapter 1 describes the background of this study, which includes spoken dialogue systems and human-robot interaction. In Chapter 2, related works on character and user adaptation of the spoken dialogue systems are reviewed. Chapter 3 describes the platform of the spoken dialogue system for an android ERICA used in this study. The robot has a physical body and generates various human-like dialogue behaviors, which are important for the character expression.</p> <p>In Chapter 4, the baseline character expression model using spoken dialogue behaviors is designed. It starts with data collection for training the model, in which subjects listened to the dialogue data and evaluated their character impressions of the dialogue robot. The statistical analysis showed correlation between the character impression and spoken dialogue behaviors. The baseline character representation model was constructed by using this dataset. The proposed model is given three character traits (extrovert, emotional instability, and politeness) and controls the four dialogue behaviors: utterance amount, backchannel frequency, filler frequency, and switching pause length. Subject experiments confirmed that the proposed model expressed the characters as intended. The corpus analysis also showed that the proposed model can predict various speakers' characters in the corpus.</p> <p>Chapter 5 addresses the enhancement of the character expression model</p>			

using neural networks, which represent the three character traits for model training. It is necessary to collect many pairs of data on the character impressions and behaviors. Therefore, semi-supervised learning is proposed based on a Variational Auto-Encoder (VAE) which combines the limited amount of the labeled pair data with unlabeled corpus data. It is shown that the proposed model can express given characters more accurately than the baseline model with only supervised learning. Using the proposed learning method, moreover, the spoken dialogue system can express an additional dialogue behavior that is not included in the labeled data.

Chapter 6 addresses task adaptation using the character expression model. The corpus analysis shows that the tendency of characters is different among dialogue tasks. The character expression model is implemented for the tasks of job interview and laboratory guide. Subjective evaluations of the dialogue video show that expressing a character in accordance with the dialogue task by the proposed model improves the user impression of the appropriateness in formal dialogue such as job interview. In the real dialogue experiment, it is confirmed that characters designated to the task received higher evaluation scores than the baseline system.

Chapter 7 addresses user adaptation using the character expression model. The analysis of the speed dating corpus shows that the combination of the subjects' personality affects the favorable impressions. Based on the analysis, a character adaptation model that controls spoken dialogue behaviors is designed and developed. This model classifies the user personality and the system character into four classes and determines the most appropriate character of the system according to the user personality. We conducted real dialogue experiments, in which subjects talked with a robot as a laboratory guide (task-oriented dialogue) and chit-chat (non-task-oriented dialogue) in four different character conditions. It is confirmed that the extrovert character was preferred for items on the laboratory guide's skill and that the character matched to the user personality was preferred for items on how easy to talk with the robot.

This thesis is concluded in Chapter 8.

(論文審査の結果の要旨)

音声対話システムや会話ロボットが様々な社会的場面で、人間に受け入れられ、関係を構築するためには、適切なキャラクタを有し、表現することが必要と考えられる。本研究はこのような仮説の下で、音声対話固有のふるまいに着目したキャラクタ表現モデル、及びその効率的な機械学習の方法を提案し、研究紹介・就職面接・雑談などを行うシステムにおいて、タスク適応及びユーザ適応の効果を実験的に評価した結果をまとめたもので、主な成果は以下の通りである。

1. キャラクタを表現可能な音声対話固有のふるまいとして、発話量、相槌の頻度、フィラーの頻度、交替潜時の長さの4つを考慮して、外向性・情緒不安定性・丁寧さの3次元のキャラクタ特性と有意な相関があることを示した。これに基づいて、ロジスティック回帰に基づくキャラクタ表現モデルを構築し、キャラクタからふるまいを生成することで、意図した外向性と丁寧さを人間が知覚できることを示した。
2. 上記のふるまいからキャラクタ特性を認識するエンコーダと、キャラクタ特性からふるまいを表出するデコーダからなるVariational Auto-Encoder (VAE)を構成し、再構成損失を導入することで、キャラクタ特性のラベルがないコーパスデータを活用できる半教師あり学習を提案した。このキャラクタ表現モデルと学習法により、人間が表出するふるまいの予測精度が改善することを示した。
3. 上記のモデルを用いて、傾聴・就職面接・お見合い練習などの対話を分析した結果、タスクによって人間が表現するキャラクタの傾向が異なっていることを明らかにした。また、お見合い練習対話においては、対話者間のパーソナリティの組合せが好感度に影響することを確認した。これらの知見に基づいて、就職面接・研究紹介・雑談を行うロボットにおいて、キャラクタ表現によるタスク適応とユーザ適応を行った結果、適切なキャラクタを用いることによりユーザの印象が向上する傾向を確認した。

以上のように本論文は、音声対話システムにおいてキャラクタ表現を行う新たなモデル化と学習法を提案するとともに、人間型ロボットを用いて複数のタスクで実装・評価を行っており、学術上・実用上寄与するところが少なくない。よって、本論文は博士（情報学）の学位論文として価値あるものと認める。また、令和5年2月15日に論文とそれに関連した内容に関する口頭試問を行った結果、合格と認めた。また、本論文のインターネットでの全文公開についても支障がないことを確認した。