Influence of Personality-based Features for Dialogue Generation in Computational Narratives

Weilai Xu[1] and **Fred Charles** and **Charlie Hargood** and **Feng Tian** and **Wen Tang**

Abstract. In this paper, we present an approach for generating dialogues for characters within the context of computational narratives using personality-based features for deep neural networks. The approach integrates the requirements of both narrative genres and personality traits for the definition of character-based stylistic models. The modelling of characters' features from existing datasets of complete stories permits the generation of personality-rich character dialogues. We present early results from an evaluation based on a sample of characters' personality traits across different narrative genres, demonstrating variability in the resulting dialogues.

1 Introduction

Dialogues in computational narratives are a strong mean of communicating author-driven story elements through the characters themselves. Computational narratives have been well researched over the last 20 years, but mostly around exploring narrative structure(s) (e.g. [6]), whilst the dialogue-driven narratives are still hindered due to the limitations of dialogue generation. In recent years, research in natural language generation has focused on content rather than on *style*, even though style is a crucial part of narrative dialogue and its rich contents [3, 8]. Existing works incorporate stylistic information by using additional features, such as gender [2], sentiment [1], and theme [1], as well as controlling the style of generated sentences by altering these features. Most of these works only consider the local features, i.e. the features pertaining to individual sentences affecting expression alteration in the scope of each individual sentences. However, it is necessary to use higher level knowledge for generating narrative-based dialogues, which represents the authorial intent and provides consistency over the story generated. Based on the consideration above, our motivation is to generate dialogues with features derived from the story, with which the generated dialogues can reflect some styles in the context of story design and authoring.

2 Methodology

McKee [5] points out that characters are assigned various characteristics based on their functions within a story. This was the basis of our approach where the variety of characters' personality traits has the potential to generate impact on the utterances whilst retaining the authorial story intents. The main processes of the methodology are listed as following:

1. **Data collection and process:** The dialogues are collected from IMSDb since [7] demonstrated that rich narrative knowledge can be extracted from screenplay. All screenplays are processed by removing all elements other than dialogue, and the dialogues are parsed into utterance pairs (partial dialogue) according to character (speaker) names by movie and genre.
2. **Label:** We specifically select three primary traits (extraversion, emotional stability, and agreeableness, or abbr. Ex, Em and Ag) using principal component analysis of training datasets. Each response of the utterance pairs is labelled with *High*, *Medium*, or *Low* on these three traits according to their character's personality trait score using Mairesse's personality tool [4] based on the Big-Five personality model. This score is here calculated from all the sentences uttered by a single character from a complete movie, representing an *overall* personality score of this character. This score is in the range 1 to 7, which is then divided into 3 sub-ranges: *Low* in the range lower than 3.5, *Medium* in 3.5 - 4.5, and *High* in the range greater than 4.5. For instance, for the extraversion trait, the label *Low* denotes introvert and the label *High* denotes extravert.
3. **Train Seq2Seq dialogue system:** The LSTM Seq2Seq model is pre-trained on Cornell's[2] and DailyDialog[3] corpora, which only contain dialogue pairs. The pre-trained model then is fine-tuned on the IMSDb corpus with the personality labels described above. During the encoding phase, the unlabelled utterances are processed through the encoder for generating the context vectors. Then the corresponding responses with the labels representing the personality of characters are used as input to the decoder for the supervised learning stage.
4. **Generate responses:** The fine-tuned model tends to generate possible responses in variety according to different target personality trait combinations. We randomly selected 30 utterances from the test dataset for each genre out of 4 major narrative-rich genres (Drama, Romance, Action, and Thriller). Therefore, each test utterance with 28 different combinations ($3^3 + 1$ blank control) is processed and 100 possible responses for each combination (2,800 in total) are generated using beam search with beam size 100.

Table 1: Generated samples with personality traits combinations based on an input utterance collected from a movie of Drama genre.

Trait Combination	**Input:** Man, I just been sitting here thinking.
Ex: *H*, Em: *H*, Ag: *H*	**Output:** Oh, my God, what are we doing here?
Ex: *H*, Em: *H*, Ag: *M*	**Output:** You know, it 's been a long time since we got out of the car.
Ex: *H*, Em: *L*, Ag: *M*	**Output:** How do you know I 'm here?
Ex: *M*, Em: *L*, Ag: *H*	**Output:** How long does it take?
Ex: *M*, Em: *L*, Ag: *L*	**Output:** No, we 're not.
Ex: *L*, Em: *H*, Ag: *L*	**Output:** No way!
Ex: *L*, Em: *M*, Ag: *H*	**Output:** You 're right, sir.

[1] Bournemouth University, Faculty of Science and Technology, UK, email: wxu@bournemouth.ac.uk

[2] https://www.cs.cornell.edu/~cristian/Cornell_Movie-Dialogs_Corpus.html
[3] http://yanran.li/dailydialog

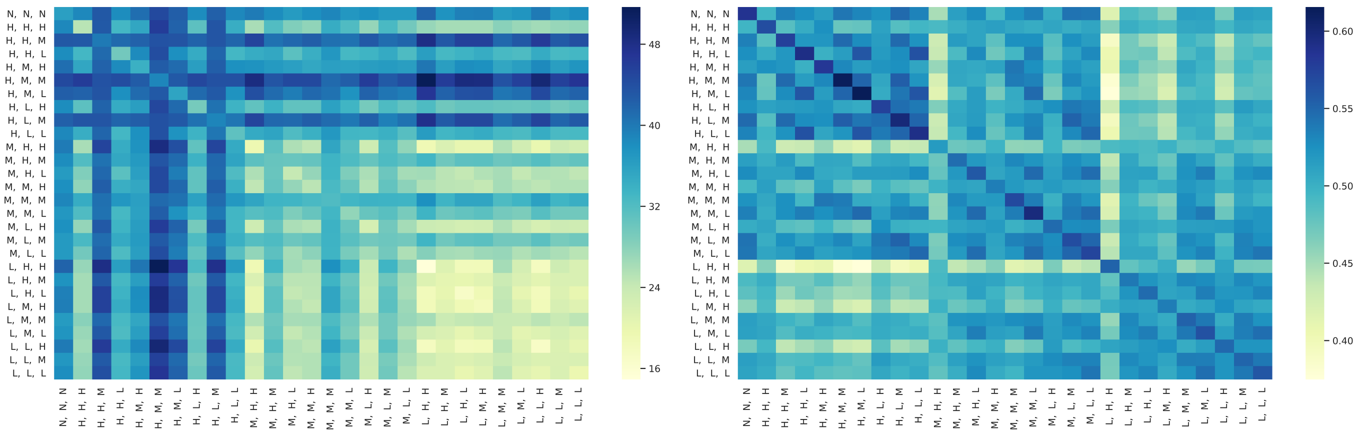

Figure 1: Mean edit distance (left) and mean cosine similarity (right) between two personality trait combinations pair-wisely of 30 test utterances in the Drama genre. Each row and column labels contain the labels of three personality traits (extraversion, emotional stability, and agreeableness). These labels are ordered by the label *High*, *Medium*, and *Low* for each trait. A blank control combination without trait label is placed on the first row and column, labelled *None*.

3 Results

In order to evaluate how various personality trait combinations affect the variety of generated responses, we use two criteria for evaluating the generated responses between two combinations pair-wisely, within surface text and semantic level respectively for 4 major genres. An example of generated responses is presented as Table 1.

1. **Surface Text Variety:** The impact on surface text variety is quantified by calculating the edit distance (specifically, we use Levenshtein distance here, the same below) of the responses between two combinations. For instance, the greater distance in generated utterances length denotes a higher variety.
   The overall surface variety between two personality trait combinations pair-wisely is shown in Figure 1 (left). Combinations with *High* label of trait extraversion (more extravert) tend to have a greater surface difference to most of the other combinations. And combinations with *Low* label (more introvert) show a smaller surface difference to others.
   To quantify how each trait affects the surface variety, we swap the label of each trait for each combination individually and calculate the change of the edit distance. This evaluation shows that swapping the label of trait extraversion has a significantly greater distance change than swapping the value of the other two traits, reaching 6.024, and the values of emotional stability and agreeableness are 1.763 and 3.259 respectively. This observation indicates extraversion trait affects the surface variety more than the other two traits, i.e. this trait has explicit and direct impact on surface text. We also have similar observations from the other three narrative genres, which indicates impact is genre independent.
2. **Semantic Variety** The cosine similarity of the responses between two combinations is calculated to quantify the impact on semantic variety, using Universal Sentence Encoder[4] which provides semantic representation by word embeddings. By this criterion, smaller distance means greater variety.
   Figure 1 (right) demonstrates, for most input utterances, that the cosine similarity scores between every two trait combinations have average distribution and fall into the range from 0.4 to 0.6, showing the model with personality features is capable of generating sentences with various semantic meaning. We also notice that swapping the label of all three personality traits has a similar impact on the semantic variety (values are 0.013, 0.012, and 0.013 respectively). Similarly to the surface text variety, we also have similar observations from the other three narrative genres on the aspect of semantic variety.

[4] https://tfhub.dev/google/universal-sentence-encoder/2

4 Discussion

In this paper, we present an early approach for generating character-based dialogues in the context of narratives, showing that dialogue style can be altered using characters' personality features across different narrative genres. Future work will include exploring how the importance of the narrative role of the character in the narrative can affect the dialogue styles by incorporating more story authoring relevant features, as well as how suited these generated dialogues are in narrative scenes. The results presented are the foundation of further research on character-based stylistic dialogue generation, which will benefit interactive digital narrative entertainment systems.

REFERENCES

[1] Jessica Ficler and Yoav Goldberg, 'Controlling linguistic style aspects in neural language generation', in *Proceedings of the Workshop on Stylistic Variation*, pp. 94–104, (2017).

[2] Jiwei Li, Michel Galley, Chris Brockett, Georgios Spithourakis, Jianfeng Gao, and Bill Dolan, 'A persona-based neural conversation model', in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 994–1003, (2016).

[3] Grace Lin and Marilyn Walker, 'Stylistic variation in television dialogue for natural language generation', in *Proceedings of the Workshop on Stylistic Variation*, pp. 85–93, (2017).

[4] François Mairesse, Marilyn A Walker, Matthias R Mehl, and Roger K Moore, 'Using linguistic cues for the automatic recognition of personality in conversation and text', *Journal of artificial intelligence research*, **30**, 457–500, (2007).

[5] Robert McKee, 'Substance, structure, style, and the principles of screenwriting', (1997).

[6] Julie Porteous, Fred Charles, and Marc Cavazza, 'Plan-based narrative generation with coordinated subplots', in *Proceedings of the Twenty-second European Conference on Artificial Intelligence*, pp. 846–854. IOS Press, (2016).

[7] David R Winer and R Michael Young, 'Automated screenplay annotation for extracting storytelling knowledge', in *Thirteenth Artificial Intelligence and Interactive Digital Entertainment Conference*, (2017).

[8] Weilai Xu, Charlie Hargood, Wen Tang, and Fred Charles, 'Towards generating stylistic dialogues for narratives using data-driven approaches', in *International Conference on Interactive Digital Storytelling*, pp. 462–472. Springer, (2018).