

MMPD: A Multimodal Personality Dataset

Anonymous ACL submission

Abstract

This document is a supplement to the general instructions for *ACL authors. It contains instructions for using the L^AT_EX style files for ACL conferences. The document itself conforms to its own specifications, and is therefore an example of what your manuscript should look like. These instructions should be used both for papers submitted for review and for final versions of accepted papers.

1 Introduction

Personality refers to the individual differences in characteristic patterns of thinking, feeling, and behaving. The study of personality has a broad and varied history in psychology with an abundance of theoretical traditions. The major theories include dispositional (trait) perspective, psychodynamic, humanistic, biological, behaviorist, evolutionary, and social learning perspective. At the very beginning, the personality prediction problem is viewed simply as a text classification, which aims to identify one’s personality traits from online corpus he/she creates.

With the advent of advanced computational techniques and the burgeoning field of multimodal deep learning, using multimodal models to predict personality types becomes an hot topic in user profile research. But most of the existing datasets are text-based scraping from the social media blogs or comments. To meet the requirement of lacking other modality information, a lot of multimodal dataset were released in recent years. Jiang et al. (2019) published a multimodal dataset called FriendsPersona which concludes more than 700 dialogues and 7 characters based on famous TV series *Friends*. There are also many multimodal datasets used to perform other tasks and some of the personality prediction works will modify their datasets to adapt the personality context. For instance, TVQA is a large dataset which is initially designed to do the vi-

sual question answering task. It is used frequently in our research field because of the its large scale.

However, on one hand, whatever the datasets are used for, they have always had a problem: the **accuracy** of the personality traits annotations. They just invited three or four people who are interested in that program to annotate the personality manually and the results are highly depend on the volunteers’ comprehension of the characters. Zhu et al. (2023) found that personality database website has marked thousands of virtual characters in movies and TV shows and they scraped the personality data from it to annotate TVQA dataset. On the other hand, most of the multimodal personality prediction works focus on images rather than videos because of the roughly division of the scenes and absence of annotations for each utterance in existing datasets. From the psychological perspective, people’s personality is always changing based on the environment. So we should not consider the personality prediction in isolation and relationship networks are integral to personality prediction because they provide a rich context in which individual behaviors, preferences, and traits are manifested and observed.

In this work, we aim to provide a brand new multimodal dataset MMPD that contains enough dialogues and videos, accurate annotations and relationship networks between characters as well as focus on multimodal compositionality. We collected the personality annotations from a popular voting website, because each personality type are determined by thousands of people which is more reliable than manually annotating. MMPD is built on 298 movies and 10 TV series in different genres, including more than 35k dialogues, 170k utterances, 4016 characters and 963 hours videos. With the rich annotation, our dataset supports 4 personality traits models (MBTI, Big Five, Enneagram and Instinctual Variant), 7 kinds of social relationship and 8 attitudes for the emotional relationship.

Our contributions are as follows:

081	• We propose the largest and diverse multimodal dataset MMPD comparing the existing works.	126
082		127
083	• We divide the scenes according to the original scripts and design a method to annotate for each utterance reaching 87% accuracy.	128
084		129
085		130
086	• We firstly define several relationship types to represent the interaction between the character scene by scene.	131
087		132
088		
089	2 Dataset Design	
090	This paper introduces a new multimodal personality dataset, MMPD, consisting of 303 movies and 18 TV shows, which is the largest of existing multimodal datasets. In this section, we provide a specific description about our dataset in terms of design principles and the structure in details.	133
091		134
092		135
093		136
094		137
095		138
096	2.1 Design Principles	
097	In constructing such a dataset for personality prediction, incorporating four distinct personality models, provides a comprehensive framework for undersatanding the multifaceted nature of human Personality:	139
098		140
099		141
100		142
101		143
102	• Myers–Briggs Type Indicator (MBTI): The MBTI is utilized for its popular four dimensional categorization of personality types, offering a straightforward way to undersatand how individuals engage with the environment and make decisions.	144
103		145
104		146
105		
106		
107		
108	• Big Five: The Big Five personality trait model is included for its empirical support and broad acceptance within the psychological community. It covers a range of personality dimensions (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism) that are universally recognized and have been linked to various outcomes in personal and professional contexts.	147
109		148
110		149
111		150
112		151
113		152
114		153
115		154
116		155
117	• Enneagram: The Enneagram adds depth to this dataset by introducing a typology of nine interconnected personality types, offering a dynamic perspective on individual differences. Its inclusion is strategic, as it provides insights into core motivations, fears, and desires that underpin behavior, thus allowing for a more detailed exploration of personality dynamics and potential growth paths for individuals.	156
118		157
119		158
120		159
121		
122		
123		
124		
125		
	• Instinctual Variant: The concept of Instinctual Variants (or Subtypes) within the Enneagram framework enriches the dataset by addressing the fundamental survival drives—Self-Preservation, Social, and Sexual (One-to-One)—that influence an individual’s priorities and interactions.	160
		161
		162
		163
		164
		165
		166
		167
		168
		169
		170
		171
		172
	Each model offers unique insights and covers different aspects of personality making them collectively valuable for a multidimensional approach personality prediction. Besides these personality models, we introduce two main categories of relationship among characters.	
	The first one is social relationship, which provides a comprehensive canvas on which to observe and interpret the nuances of personality in action. We conclude 7 social relationship from the perspective of psychology and sociology, which recognizes that personality is not only a matter of internal traits and instincts but also fundamentally shaped and expressed through interactions with others in various domains of life.	
	Figure 1: A sample of MMPD	
	The social relationships above are relatively non-changable, not depicting the attitudes towards someone else. So we define another 8 types for the emotional relationships, as the aid for the comprehension of personality.	
	We choose affection, jealousy, dislike, pity, respect, hostility, envy and gratitude as our annotators for the emotional relationship, which concludes the diverse attitudes in human’s daily life.	
	Thus, we select a binary tuple to annotate the pair of characters for each scene, as well as emotional relationship tag for each utterance.	
	2.2 Structure of MMPD	
	MMPD has a very large scale for the three modalities: video, audio and text. We built a fine grained structure describing the interactions and corresponding personality traits for each utterance based on the original scripts.	
	Firstly, we divide the scripts into several scenes according to the coherence in language of camera instead of ramdonly clipping in a certain time period. For each scene, we match the time stamps in the subtitle files to the raw scripts’ utterances in dialogues. Additionally, ChatGPT model is used to determine the relationship types of characters for	

Dataset	Field	Dialogues	Characters	Source
MEmoR	Emotion Recognition	8.53k	7	The Big Bang Theory
FriendsPersona	Personality Recognition	0.71k	7	Friends
CPED	Emotion and Personality Recognition	12k	392	40 TV shows
UDVIA	Social Interaction Analysis	188	147	Dyadic Interaction
The ChaLearn FI	Personality Recognition	10k	3000	Youtube
TVQA	Temporal and Spatial Localization	29.4k	Unknown	6 TV shows
Our Dataset	Personality Recognition	70k	4000	300+ Movies and 10 TV Shows

Table 1: Comparison of different datasets and our MMPD, there should be utterances of each dialogues

Relationship type	Description
Family Relationship	Parents (grand parents) and children, siblings, etc.
Friendship	Based on common interest, mutual respect and affection, but not related to the blood
Romantic Relationship	Based on emotional attraction and include dating, marriage, etc.
Professional Relationship	Formed in a work environment, such as colleagues, superiors and subordinates, etc.
Social Relationship	Formed in a broader social context, such as neighbors, club members.
Academic Relationship	Formed in an educational setting, such as between teachers and students, classmates.
Online Relationship	Established in online spaces or through social media platforms.

Table 2: This is the chart to represent the social relationship

Relationship type	Description
Affection or Fondness	Parents (grand parents) and children, siblings, etc.
Jealousy	Based on common interest, mutual respect and affection, but not related to the blood
Dislike or Aversion	Based on emotional attraction and include dating, marriage, etc.
Pity or Sympathy	Formed in a work environment, such as colleagues, superiors and subordinates, etc.
Respect	Formed in a broader social context, such as neighbors, club members.
Hatred or Hostility	Formed in an educational setting, such as between teachers and students, classmates.
Envy	Established in online spaces or through social media platforms.
Gratitude	This is the emotion of being thankful for someone else's help or kind actions.

Table 3: This is the chart to represent the emotional relationship

each scene. And then we clip the corresponding movie and TV show according to the time stamps that we matched before. Finally we put the four types of personality models to the corpus, as shown in Fig. 1.

Aiming to deliver a tidy and readable structure, there is no more suitable file types than json format. We distribute different scenes in a single json file with index. For each movie or TV show, the video clips and corresponding json files are stored in the same directory.

Example illustrating in a picture

3 Methodology

3.1 Source of Data

Considering the unreliable labeling method of existing works, we collect the personality annotations from personality database website as well as the voting distribution that indicates the credibility of current personality type. We used some python scripts to scrape the personality data from the website and annotate them to the corresponding scripts. As for the scripts and subtitles, we also find some open-source websites for research offering the free scripts and subtitles of many famous movie and television programs. To represent the diversity of the real world scenarios, we select various genres of the movies and TV series which includes action, thriller, romance, comedy, science fiction, etc.

Explain where and how to collect the data.

3.2 Data Alignment Process

As subtitle contain temporal information and original scripts associate utterances with characters, we are supposed to align them properly as efficient as possible. However, most of the existing multimodal datasets annotate the timestamps manually with taking up a great deal of time. There are also some works which utilize different automatic tools to align the utterances with their corresponding information. For instance, Shen, et al (2020) get the

So we designed a fuzzy matching algorithm to match utterances with their corresponding timestamps and speakers.

Based on the subtitle file we collected, our main idea is that we already have the precise timestamp for each utterance so we just need to put the corresponding character's name into the right position. At the beginning, we set a threshold like 0.8 to represent the similarity between the utterances in

scripts and subtitles. And then we compare each utterance from scripts and subtitle to figure out whether their similarity reach the threshold. If it does, the corresponding character will be marked to the utterance.

3.3 Annotation Process

`\documentclass[11pt]{article}`

To load the style file in the review version:

`\usepackage[review]{acl}`

For the final version, omit the review option:

`\usepackage{acl}`

To use Times Roman, put the following in the preamble:

`\usepackage{times}`

(Alternatives like txfonts or newtx are also acceptable.)

Please see the \LaTeX source of this document for comments on other packages that may be useful.

Set the title and author using `\title` and `\author`. Within the author list, format multiple authors using `\and` and `\And` and `\AND`; please see the \LaTeX source for examples.

By default, the box containing the title and author names is set to the minimum of 5 cm. If you need more space, include the following in the preamble:

`\setlength\titlebox{<dim>}`

where `<dim>` is replaced with a length. Do not set this length smaller than 5 cm.

4 Evaluation

accuracy of methodology

Provide statistics about the dataset, such as size, number of instances, diversity of languages or dialects, and any other relevant metrics.

4.1 Footnotes

Footnotes are inserted with the `\footnote` command.¹

4.2 Tables and figures

See Table 1 for an example of a table and its caption. **Do not override the default caption sizes.**

¹This is a footnote.

4.3 Hyperlinks

Users of older versions of L^AT_EX may encounter the following error during compilation:

```
\pdfendlink ended up in different
nesting level than \pdfstartlink.
```

This happens when pdfL^AT_EX is used and a citation splits across a page boundary. The best way to fix this is to upgrade L^AT_EX to 2018-12-01 or later.

4.4 Citations

5 Conclusion

6 Copyright

Table 4 shows the syntax supported by the style files. We encourage you to use the natbib styles. You can use the command `\citet` (cite in text) to get “author (year)” citations, like this citation to a paper by ?. You can use the command `\citep` (cite in parentheses) to get “(author, year)” citations (?). You can use the command `\citealp` (alternative cite without parentheses) to get “author, year” citations, which is useful for using citations within parentheses (e.g. ?).

A possessive citation can be made with the command `\citeposs`. This is not a standard natbib command, so it is generally not compatible with other style files.

6.1 References

The L^AT_EX and BibT_EX style files provided roughly follow the American Psychological Association format. If your own bib file is named `custom.bib`, then placing the following before any appendices in your L^AT_EX file will generate the references section for you:

```
\bibliography{custom}
```

You can obtain the complete ACL Anthology as a BibT_EX file from <https://aclweb.org/anthology/anthology.bib.gz>. To include both the Anthology and your own `.bib` file, use the following instead of the above.

```
\bibliography{anthology,custom}
```

Please see Section 7 for information on preparing BibT_EX files.

6.2 Appendices

Use `\appendix` before any appendix section to switch the section numbering over to letters. See Appendix A for an example.

7 Related Works

Unicode cannot be used in BibT_EX entries, and some ways of typing special characters can disrupt BibT_EX’s alphabetization. The recommended way of typing special characters is shown in Table 1.

Please ensure that BibT_EX records contain DOIs or URLs when possible, and for all the ACL materials that you reference. Use the `doi` field for DOIs and the `url` field for URLs. If a BibT_EX entry has a URL or DOI field, the paper title in the references section will appear as a hyperlink to the paper, using the `hyperref` L^AT_EX package.

Acknowledgements

This document has been adapted by Steven Bethard, Ryan Cotterell and Rui Yan from the instructions for earlier ACL and NAACL proceedings, including those for ACL 2019 by Douwe Kiela and Ivan Vulić, NAACL 2019 by Stephanie Lukin and Alla Roskovskaya, ACL 2018 by Shay Cohen, Kevin Gimpel, and Wei Lu, NAACL 2018 by Margaret Mitchell and Stephanie Lukin, BibT_EX suggestions for (NA)ACL 2017/2018 from Jason Eisner, ACL 2017 by Dan Gildea and Min-Yen Kan, NAACL 2017 by Margaret Mitchell, ACL 2012 by Maggie Li and Michael White, ACL 2010 by Jing-Shin Chang and Philipp Koehn, ACL 2008 by Johanna D. Moore, Simone Teufel, James Allan, and Sadaoki Furui, ACL 2005 by Hwee Tou Ng and Kemal Oflazer, ACL 2002 by Eugene Charniak and Dekang Lin, and earlier ACL and EACL formats written by several people, including John Chen, Henry S. Thompson and Donald Walker. Additional elements were taken from the formatting instructions of the *International Joint Conference on Artificial Intelligence* and the *Conference on Computer Vision and Pattern Recognition*.

A Example Appendix

This is an appendix.

Output	natbib command	ACL only command
(?)	\citep	
?	\citealp	
?	\citet	
(?)	\citeyearpar	
?’s (?)		\citeposs

Table 4: Citation commands supported by the style file. The style is based on the natbib package and supports all natbib citation commands. It also supports commands defined in previous ACL style files for compatibility.