Introduction

Interactive narrative often involves interacting with virtual agents. Most character dialogue are handcrafted → authoring bottleneck
- LA Noire (2011) ~2200 pages
- Star Wars: The Old Republic: ~ 40+ novels

Basis of our work: To help authors with their creative process by using Expressive Natural Language Generation [ENLG]. This process focuses on stylistic, social aspects of the linguistic behavior. We would like to learn character models through film dialogue.

Why film screenplays?
- Authored to deliberately convey the feelings, thoughts, and perceptions of the film character.
- Can examine an operationalization of archetypes by looking at dialogue of heroes, villains, wise men, etc.

The Corpus

- 7,400 film characters
- 644,000 lines of dialogue
- 9,599,000 tokens
- Ontological Features from IMDB (Internet Movie Database, www.imdb.com)

Feature Categories
- Genre: drama, thriller, crime, comedy, action, romance, adventure
- Gender: male, female

Feature (Set) Feature Description
- Basic Number of sentences, sentences per turn, number of verbs per sentence, etc.
- Polarity Use SentiWordNet 3.0 on all available words
- Dialogue Act Trained with NPS Corpus with 15 dialogue acts types (e.g., Accept, Bye, Clarify)
- Passive Sentence 3rd party software to detect passive sentences
- LUC Word Categories Use Linguistic Inquiry and Word Count (LIWC) text analysis software to categorize words (e.g., positive emotion category words: love, sweet, nice).
- Tag Question Regular expression
- Verb Strength Averaged sentiment values of verbs

Future Augmentation
- TV series: investigate characters scripted by different authors
- More thoroughly evaluate the accuracy of our automatically generated annotations

Film Character Dialogue Stylistic Differences

Annie Hall: Sports Club

Theme:
How to get a Ride

Genre:
Drama vs. Action

The Terminator 2: Cigar Biker scene

Theme:
I need your clothes, your boots, and your motorcycle.

Fight scene!

Film Corpus (original character dialogue)

Extract

Learned Character Models

Map significant features

ENLG Engine Input Parameters

Apply story domain (e.g., SpyFeet)

Generated Dialogue

Evolution (see Ref. 3)

Personage parameters for Annie

PERSONAGE parameters for

SpyFeet

Exercise Role-Playing Game

Generated dialogue

- Come on, I don’t know, you’ll People say Carrell is strange while I don’t rush to um, judgment.
- I don’t know, I think that you brought me cabbage, so I will tell something to you, alright?
- I see, I am not sure. Obviously, I respect Wał. However, he isn’t my close friend, is he?

Corpus Application: Learning Character Models

Learning Character Models

Map significant features

ENLG Engine Input Parameters

Apply story domain (e.g., SpyFeet)

Generated Dialogue

Evolution (see Ref. 3)

Personage parameters for Annie

PERSONAGE parameters for

SpyFeet

Exercise Role-Playing Game

Generated dialogue

- Come on, I don’t know, you’ll People say Carrell is strange while I don’t rush to um, judgment.
- I don’t know, I think that you brought me cabbage, so I will tell something to you, alright?
- I see, I am not sure. Obviously, I respect Wał. However, he isn’t my close friend, is he?

References

Annotated Corpus of Film Dialogue for Learning and Characterizing Character Style

Marilyn Walker, Grace Lin, Jennifer Sawyer
{maw, glin, jsawyer}@soe.ucsc.edu

Natural Language and Dialogue Systems Lab
University of California, Santa Cruz, U.S.A.
http://nlds.soe.ucsc.edu/software