WriteRec Final Presentation

by: (pat[ri)el(la]y)

(WriteRec Airing of Grievances)

by: (pat[ri)el(la]y)

Outline

- Review on task goals and annotation specification
- Characteristics of the data set
- Difficulties collecting the dataset (solved and unsolved)
  - Possible improvements in the next iteration
- IAA at different aspects/levels and their interpretation
  - Metrics used and justification behind them
  - Interpretation and error analysis of the agreements
- Machine learning experiment
  - Experiment design
  - Features extracted from the annotation
  - Experiment results

Task Goals (Review)

1. To acquire plot summaries of uniform size and format for one genre

2. To identify entities as well as the events, properties, and relationships that affect the entities and show the interactions between the entities

3. To recommend books based on similarities in plots, characters, and character development

Annotation Specification (Review)

- Identifying entities
  - Protagonist, antagonist, unknown, other, object
- Identifying triggers
- Identifying relations between entities via triggers

Annotation Instructions: Character Types

- Tag the first full-name mention of the character; else
- Tag the first named mention of the character; else
- Tag the first nominal descriptor of the character; else
- Tag the first pronominal descriptor of the character; else
- Character is implied by the conflict (non-consuming tag)

Annotation Instructions: Conflict Identification

- Tag verbs and events representing interactions **between** characters
  - directionally:

    from agent/source

    to theme/goal/malefactive/benefactive/experiencer

- Tag verbs and events representing struggles and victories by characters **against their environment**, non-characters, etc
  - directionally:

    character is the goal, but what is the source?

    sometimes explicit non-characters. An arc from implicit Environment “character”?

- Tag verbs and events representing **internal struggles** by characters
  - directionally:

    character is both the source and goal

Characteristics of the Data Set

- Got summaries from GoodReads
- Weeks 1, 2, 3 with 50 docs each
  - 150 total plot “summaries” (final: ~115)
- Nonstandardized length
- Nonstandardized content
- Some more relevant information, some less relevant information

Characteristics of the Data Set

The Sacred Band of Thebes lives on, a world away, in this mythic novel of love in war in ancient times. In 338 BCE, during the Battle of Chaeronea that results in the massacre of the Sacred Band of Thebes, the legendary Tempus and his Stepson cavalry rescue twenty-three pairs of Theban Sacred Banders, paired lovers and friends, to fight on other days. These forty-six Thebans, whose bones will never lie in the mass grave that holds their two hundred and fifty-four brothers, join with the immortalized Tempus and his Sacred Band of Stepsons, consummate ancient cavalry fighters, to make new lives in a faraway land and fight the battle of their dreams where gods walk the earth, ghosts take the field, and the angry Fates demand their due.

Characteristics of the Data Set

From New York Times Bestselling author Brent Weeks...
For Durzo Blint, assassination is an art-and he is the city's most accomplished artist.

For Azoth, survival is precarious. Something you never take for granted. As a guild rat, he's grown up in the slums, and learned to judge people quickly - and to take risks. Risks like apprenticing himself to Durzo Blint.

But to be accepted, Azoth must turn his back on his old life and embrace a new identity and name. As Kylar Stern, he must learn to navigate the assassins' world of dangerous politics and strange magics - and cultivate a flair for death.

Difficulties Collecting Data Set (solved + unsolved)

- GoodReads doesn't make it easy to find books in the 'Heroic Fantasy' category
- Span issues
  - Hopefully fixed by MAE versions
- Changed the DTD halfway through and had difficulty updating old xmls to match new DTD
  - Can be fixed with more time
- Low Annotator agreement
  - Fixed with more annotators
- Abjudications were the worst
  - Mehhhhr MAE
  - See following slides.

All Of The Thoughts About MAE*, pg. 1

- Database exception
  - woops may have deleted something important
- If you click things too quickly in argselection mode, you get a database exception
- If you open the wrong kind of doc, everything messes up File IO
  - Read: DO NOT OPEN A PDF IN MAE
- Quote handling
  - Spans with double quotes break the xml
- You have to save everything before you close it, otherwise MAE gets sick

*raise your hand if you have been personally victimized by MAE

All Of The Thoughts About MAE*, pg. 2

- The last two of any equal discontiguous 1-char spans are considered identical
- MAE requires a certain kind of header to read files that it already created
  - Missing headers in preprocessed files
- MAE reads newlines as a character in span assignment but nowhere else
- MAE doesn't close document databases so you have to restart the task to start a new one
- If you reorder candidates in adjudication all of the pointers become broken
- If we'd understood MAE before we made the spec, all characters would be a single extent type
- MAE was a little finicky when loading files

*raise your hand if you have been personally victimized by MAE

Effects of Annotators on Dataset

1. Annotator opinions
    a. Perceptions of those who benefit or are harmed
    b. Passive relationships: agents harder to identify
    c. Guidelines for relations not specific enough
2. Annotator motivation
    a. Time, effort required for annotation is high

IAA

- Metrics used and justification behind them
  - Krippendorff’s alpha
    - Good for sparse agreement
    - Implemented (partially) in MAE
  - Fleiss’s kappa
    - Good fallback, doesn’t force binarized scales
    - Implemented (partially) in MAE

- Interpretation and error analysis of the agreements
  - Predicted based on adjudication:
    - Nodes okay, edges bad
    - Labels: added Unknown after feedback, but inconsistently used

Machine Learning

- Experiment design
  - k-Means Clustering
    - 5 clusters
- Features extracted from the annotation

  - (Clay draws a graph on the board and asks a dramatic rhetorical question)

Machine Learning

- Experiment design
  - k-Means Clustering
    - 5 clusters
- Features extracted from the annotation
  - (Overlapping) **local** features of the graph (Gibert 2012)
  - Character type, character type and number, character type and subtype
  - Relationships caused by pro/antagonists, others/unknowns,
    - Benefitting/harming self, others, neither
  - By objects
    - Benefitting/harming pro/antagonists, others/unknowns, neither
- Experiment results
  - Difficult to evaluate
  - Clusters are somewhat consistent across multiple trials though!

Evaluation

1. Vs. baselines:
    a. KMC using Bag of Words
        i. Yes

    b. Tf-idf and PPMI weighting to the BoW vector input
        i. Coming soon to a theater near you

2. Extrinsically:
    a. Have guinea pigs rate a cluster exemplar (1), a random cluster-mate (2), and a random non-mate (3); (1) and (2) will be closer than either is to (3) (if we succeeded)
        i. Try with humans also

Bibliography

- Jaume Gibert. (2012)." Vector Space Embedding of Graphs via Statistics of Labelling Information". Ph.D. thesis, UABarcelona

Thanks!