

NOVEL2GRAPH

Visual Summaries of Narrative Text Enhanced by Machine Learning

Vani K and Alessandro Antonucci

Istituto Dalle Molle di Studi sull'Intelligenza Artificiale - IDSIA (Switzerland)



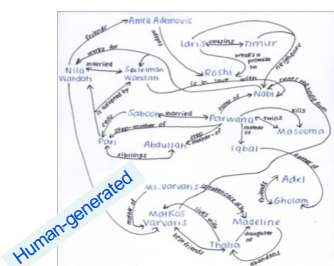
The Problem(s)

NLP, DL & NARRATIVE TEXT

- NLP is mostly used with non-narrative texts (medical, news, etc.)
- Narrative text has peculiar features (style variance, frequent use of neologisms, diversity in figures of speech, etc.)
- Deep Learning (DL) is transforming NLP!
- Can DL provide the necessary flexibility for NLP on narrative texts?

THE GOAL

- A reader (or a computer) coping with a complex plot of a novel or a story with lots of characters connected by relations of different type
- Visual summaries represented using graphs over the characters, whose edges are relations can explain the plot!
- Humans are already doing that manually!



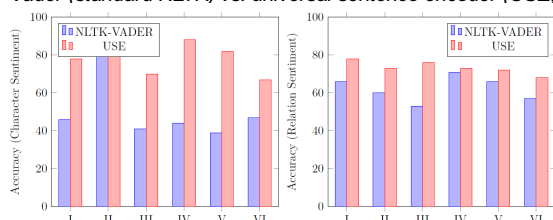
CHARACTERS (AND THEIR ALIASES) DETECTION

- Detecting Characters = Named Entity Recognition (Stanford NER)
- Aliases? Clustering based on similarity between detected entities
- DBSCAN can do clustering with pairwise similarities (fuzzywuzzy)
- Heuristic rules added to improve the clustering



SENTIMENT ANALYSIS

- Sentiment analysis to identify characters (and relation) traits
- Characters = phrase level f (Lord **voldemort** **mystique**)
- Relations = sentence level (**Harry** remembered his first nightmarish trip into the forest the first time he had ever encountered the thing was then **voldemort**.....)
- Vader (standard NLTK) vs. universal sentence encoder (USE, DL)



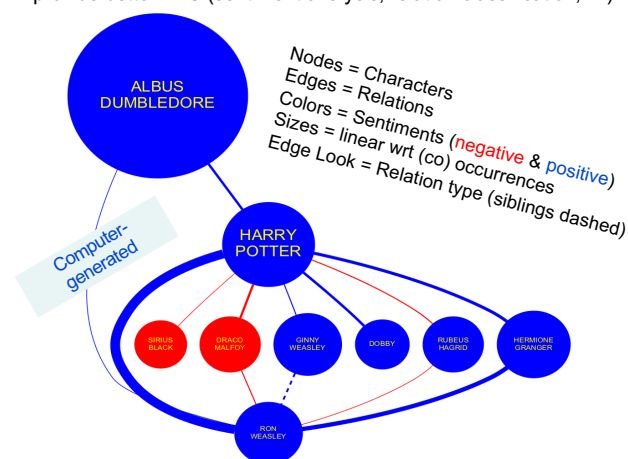
RELATION ANALYSIS & GENDER DETECTION

- Gender detection is a standard task (70% acc)
- Relation analysis, e.g., siblings detection
- Can be based on statistical ML tools (e.g., Naïve Bayes)
- With standard features (e.g., BoW)
- (Few) manually annotated sample are enough

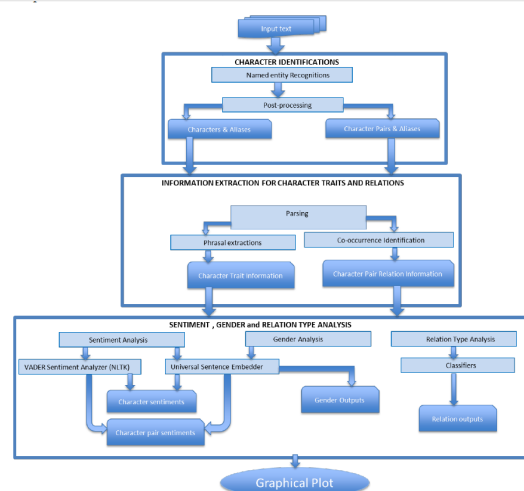
The Solution(s)

THE SOFTWARE

- A (Python) software library for automatic creation of visual summaries of narrative text by NLP
- A tool to improve readability of complex stories
- Not only based on descriptive statistical analysis (e.g., number of occurrences of characters or their co-occurrences)
- Deep Learning along with statistical machine learning and NLP can provide better NLU (sentiment analysis, relation classification, ...)



NOVEL2GRAPH WORKFLOW



CONCLUSIONS & OUTLOOKS

- A first attempt to apply the synergy of NLP and DL for the extraction and representation of information from narrative text
- Promising preliminary results, but lots of dimensions to be explored
- Dynamical plots ("animations" over the book) could be done in an incremental way or based on a time-window (chapter)
- Challenge is the temporal positioning of the facts (our analysis is based on the page position not on the time position)