Jacob Reske

05/30/13

Walt Disney Studios

NLP Research Notes

Scratch Notes

* Purpose: to use NLP/Sentiment analysis to predict overall emotional gauge of media
* What things DO we want to infer from social media—and why?
  + Interest in a film by user type—likelihood of seeing a film.
    - Added value: financial predictions more accurate. Can use interest to influence prediction models on orders, deployment of products, etc.
  + Influence of specific users on topics pertaining to WDS media. Influence of topics relevant to the media on opinion of the media itself
    - Added value: Awareness of 3rd party marketing. Putting a face to the most influential people/users/topics. Could inform marketing a film/show/product, response, could be a valuable tool for agencies. (You just can’t know everything.)
  + Predict the likelihood of a 1st party post to be well-received by fans
    - Added value: Marketing insights. Influences direct interaction with customers 🡪 increased value of social media presence. Could also inform decisions on advertising, market research, etc.
  + Find correlations between online user interest in films and critical reception
    - Added value: Awareness on the influence of professional informers on market reception before and as a product is released, and how that influence either grows or diminishes over time, and how that information is disseminated.
  + Infer the parts of a product that viewers like/will like
    - Added value: market analysis
* What is “interest” in a product? How do you measure it? What can humans get out of a single Twitter post?
  + Willingness to make a connection with a product
    - Wants/desires: “I want to go see…” “I don’t need another film that is…”
      * Scale: [-10, 10], where positive values represent
    - Intent (the probability of an event occurring): “I’m going with my friend to see…” “I will not watch…”
      * Scale: [-10, 10], where positive values represent certainty of an event occurring, and negative values represent certainty of an event not occurring, for a given event.
    - Reason for willingness: “I want to participate *because*…” “I don’t want to see this because of…”
      * This is harder. Need NLP, separate data set for reasoning alone.
    - Examples:
      * A customer really wants to purchase a video game but cannot because it is too expensive. (Want index is high, intent index at -9, reasoning sorted to price category)
      * A customer doesn’t want to see a film but feels obligated to see it with a friend. (Want index is low, intent index is a 10, reasoning is in social/friendship category)
  + Word-based analysis:
    - Intensity of emotion describing object: ”It was *horrible*”
    - Positive versus negative valence
      * Scale: [-10, 10] (similar to Jeff Breen’s scale)
    - Common versus uncommon words
      * Index: probably an open source db
    - Words by demographic
      * ‘’
  + Grammatical
    - Tense (past, present, future): “Going to see…” “I watched,” “I will watch”
      * Index: string (past, present, future)
    - Capitalization…?
  + Vocabulary
    - Industry-specific, product-specific words
    - Data-rich vocabulary (demographic-based, word-based… similar to above)
  + Reliability
    - Rhetorical devices (?????)
    - Punctuation (quantity??)
* Given a single piece of social media. By platform, what content is readily available?
  + Text [string]
  + Share/approval system [[value]]
  + Time posted [[date]]
  + Language of content posted [[string]]
  + Tagged content in post [[string]]
  + Username of poster [[string]]
  + *Possible:* location posted. [[string]]
  + *Possible:* Age/gender of poster. [[string]]
  + *Possible:* Platform/browser of poster [[string]]
* First level of inference:
  + Text: Parse text for ontological association, key words [[string]]
  + Username: Other posts by user on content [[string]]
  + User: Other posters or reposters of the same content [[string]]
  + Length: avg margin of repost over time [[float], [date]]
* Second level of inference:
  + Text: positive/negative associations
  + Text: contextualization, cross-referencing with similar topics to inform opinion