Constructing Intensity Scales for Emojis

Pulkit Singh (Advisor: Christiane Fellbaum)

1   Motivation and Goal

Expressing the degree of an emotion or judgment is an intuitive part of communication. We naturally express whether we disliked, hated or despised the film we saw last week, or whether we like, love or adore a book. It's no surprise then, that our languages have a number of words that belong to the same semantic category, but denote (and connote) different intensities. Linguists have done considerable work in identifying and encoding intensity relations among adjectives and verbs, as this forms a crucial part of enabling machines to understand natural language.

Emojis have become a crucial part of computer-mediated communication in the relatively recent past. Since emojis are often used to express emotions and opinions, it makes intuitive sense that there would also be gradation in their usage. Given this context, the goal of my independent work is to extend the work of intensity encoding to emojis. This will be done by using a corpus of web-based communication from Twitter, in conjunction with a survey of lexical-semantic patterns filled out by respondents.

2   Problem Background and Related Work

Sheinman, Tokunga et al. (2013) constructed intensity scales for adjectives by using lexical-semantic relation in a web-based corpus. They used the AdjScales method and searched for expressions of the form " A1 but not A2 " and " if not A1, at least A2", among others, to identify the semantic-strength relation between two adjectives. For example, a fragment of the form "if not terrified, at least scared" indicates that terrified is a more intense adjective than scared. They then used pairwise results in groups of adjectives to construct a linear scale of intensity.[1]

This work was extended to verbs by Fellbaum and Mathieu (2014). The AdjScales method was used for Experiencer verbs. Expressions of the form "V1 (perhaps) even V2" and "V1, not to say V2" were used to identify the relative ordering of verbs groups expressing fear, surprise, etc. This work is extremely relevant to the task at hand, as people often supplement descriptions of their own emotions or actions with emojis, and the verbs in these descriptions can be used as indicators of emotional intensity. [3]

Novak et al (2015) conducted a cross-linguistic study regarding the sentiment (positive, neutral or negative) of emojis. They computed the sentiment of the emojis from a dataset of over 1.6 million tweets in 13 different languages, that were hand-labelled by human annotators. Using these sentiment scores, they constructed a sentiment map of the 761 most commonly used emojis. [4]

Wijeratne et al. (2016) built a machine-readable sense inventory for emoji by integrating four openly available emoji resources into a single, query-able, dictionary of emoji definitions and interpretations. The web resource they compiled, called EmojiNet, contains sense-disambiguated emojis in a standardized format.

3   Approach

This project plans to address the task of constructing intensity scales for emojis in two ways - with a twitter corpus-based tabulation of the co-occurence of graded verbs and emojis, and

with a human survey of lexical-semantic patterns with emojis. This is novel because emojis are not generally treated as words. The first part enables us to aggregate the kinds of emojis people use with different intensities of words. The survey enables us to evaluate and enrich this corpus-based understanding by studying how humans respond to lexical-semantic pattern filling with emojis.

- **Twitter-Corpus Based Analysis**

  - Find groups of emojis that belong to the same semantic group
    * Decide set of semantic categories - like happiness, sadness, anger, fear etc
    * Use EmojiNet to find emojis that correspond to each category (perhaps in conjunction with BabelNet)
  - Find scaled words in the categories decided earlier. For example in the fear category, find a scale of the form "intimidate < scare < frighten < alarm < terrify" (as in Fellbaum Matthieu)
  - Use the twitter API to find groups of tweets that contain each of the verbs in the scale, and aggregate the different emojis used in conjunction with each word.
  - Assign intensity scores to each word on the scale, and according to the co-occurences of emojis with the verb, sum and normalize scores for each emoji. Use the calculated emoji scores to find the scale

- **Survey with Lexical-Semantic Patterns of Emojis**

  - This approach will extend the work of Sheinman Tokunaga, and Fellbaum and Matthieu to emojis
  - Lexical-semantic patterns will be designed, such as "This movie had some scary parts, I was *slightly scared face emoji* but not * blank *, where respondents will fill in the blank with the emoji they see fit.
  - The Amazon Mechanical Turk service will be used to get human intelligence to fill out this survey.
  - These survey results will indicated the semantic strength of the emojis being used, and help us evaluate the scale that was built from the text corpus.

4 Plan

- October

  - Apply for twitter developer account
  - Play with the twitter API to ensure that construction of dataset is viable
  - Design methodology for collecting and storing data
  - Build dataset by the end of the month
  - Design survey about intensity score of emojis using lexical-semantic patterns

- November

  - Implement intensity score calculations
  - Collate intensity scores onto a scale
  - Collect data from humans with designed survey (Amazon Mechanical Turk)

- December

  - Analyze survey data
  - Evaluate model against results of survey data

5   Evaluation

The models will be evaluated against human survey data about the intensity of emojis as indicated above. Since emojis are used concurrently with natural language, humans would be the best standard for semantic comparison of emoji intensity.

References

[1] Sheinman, Vera, Christiane Fellbaum, Isaac Julien, Peter Schulam, Takenobu Tokunaga. "Large, Huge or Gigantic? Identifying and encoding intensity relations among adjectives in WordNet". Springer Science+Business (2013).

[2] Fellbaum, Christiane and Yvette Mathieu. "A corpus-based construction of emotion verb scales" (2014)

[3] Novak, Petra, Jasmina Smailovic, Borut Sluban, and Igor Mozetic. "Sentiment of Emojis." CoRR1509 (2015).

[4] Wijeratne, Sanjaya, Lakshika Balasuriya, Amit Sheth, and Derek Doran. "EmojiNet: Building a Machine Readable Sense Inventory for Emoji." ArXiv.