Joon Park, Isaac Ehrlich
1003470870, 1004502102
COG403

Project Proposal: Humor as a Function of Visual Context

Introduction

Humor is a common, naturally-occurring feature of language and communication. However, despite its commonplace in everyday language, both the definition of humor, as well as the 'algorithm' for what makes something funny, are ill-defined and highly disputed (Attardo, 2010). While some linguists argue that humor is any content that is contrastive to serious topics, others classify any event or object that elicits laughter as humorous (McGraw & Warner, 2014). Furthering the complexity of defining humor is the question of whether or not humor is formalizable at all, given that comicality is subjective, and that humor changes historically (Attardo, 2010; Eco, 1985).

The study of humor detection in natural language processing attempts to resolve some of this debate by uncovering a model capable of distinguishing humorous and non-humours text. For example, using support vector machines (SVMs) and decision trees as classifiers, a model successfully differentiated between humorous and non-humorous texts by examining content and syntactic style (Mihalcea & Strapparava, 2006). Beyond accurately labeling humorous data, these models also contribute to a possible "recipe" for humor by distilling features important to performing this classification, such as alliteration, antonymy, and homophones. However, these models are unable to represent the full complexity of humor, and are limited to one-liners, humorous statements told in a single sentence.

While humor detection models have provided valuable insight to humor, their exclusive focus on language is a problem when compared to human humor judgement. Although some

forms of humor may be sufficiently described through text, humor is naturally a multisensory experience (Hasan et al., 2019), and is often dependent on visual cues (such as accompanying photographs or facial expressions), or environmental context. Therefore, content-dependent models that focus exclusively on text, such as the SVM model proposed by Mihalcea & Strapparava (2006), may fail to capture humor where content is inherently linked to other external cues. Consequently, we hypothesize that in order to accurately label humor that is dependent on visual cues, an additional parameter must be added to content-based models that specifies relevant context.

To study this, we will implement both a content-dependent and a context-dependent model of humor detection, and apply them to non-contextual humor (one-liners), as well as contextual humor (New Yorker cartoon captions). The context-dependent humor detection model will be a novel extension of existing content-dependent models with an added 'context' parameter based on the subject of an image in a cartoon/caption pairing. For example, while a sentence about a vacuum cleaner may not inherently be labeled as funny, given an image of a vacuum cleaner (and a context label for vacuum), an accompanying caption may become funny, and therefore a content-dependent and context-dependent model may have different results. This extended model will be trained using human-labeled caption data from the New Yorker Caption Contest.

Computational Model

In this project, we are going to apply a support vector machine (SVM) to solve the problem of differentiating between the humorous and non-humorous data. SVM is a binary classifier (in this case the classes are humorous and non-humorous) that has been shown to be successful in text categorization (Joachims, 1998; Mihalcea & Strapparava, 2006).

The ability of SVM models to predict humorous and non-humorous one-liners is demonstrated by Mihalcea & Strapparava (2006), correctly distinguishing between humorous one-liners and non-humorous sentences. Mihalcea & Strapparava (2006) also demonstrated how SVMs could be combined with other text categorization methods, such as stylistic examinations, in order to improve classification accuracy. For example, while the SVM model was sufficient in distinguishing proverbs and one-liners due to content differences, adding checks for stylistic differences improved accuracy when comparing news headlines and one-liners. For these reasons, the hybrid SVM model developed by Mihalcea & Strapparava (2006) is a reasonable basis for an extended model including context.

Two alternatives to this computational model are decision trees and the Naive Bayes model. Decision trees and Naive Bayes models are advantageous in some cases as they offer reduced complexity (and therefore are faster), while still demonstrating competence as text categorization models. However, SVM models showed higher accuracy in several text categorization tasks and humor detection experiments (van den Beukel & Aroyo, 2018), and therefore are desirable for this task.

Datasets

New Yorker Caption Contest Dataset (Link)

- Cartoon/caption pairings
- Human humor judgements on submitted captions

UR-FUNNY Dataset (Link)

- Dataset of one-liners

A Million News Headlines (Link)

- To be used as non-humorous text

Open Mind Common Sense (Link)

- To be used as non-humorous text

Predicted Outcomes

The initial step of the study will involve training a content-dependent model on one-liner humorous data (UR-FUNNY Dataset) and non-humorous (A Million News Headlines & Open Mind Common Sense Dataset) single sentences. The classification error will be measured for these judgements, and based on previous studies we expect to have low error on this dataset for the content-dependent model (Mihalcea & Strapparava, 2006).

The second part of the experiment involves training both the content and context-dependent model on the New Yorker Caption Contest (NYCC) data set. In the NYCC, human humor judgements were made on a scale of 1-3, 1 representing non-humorous and 3 representing humorous. All captions with more 'funny' than 'not funny' ratings will be considered humorous, and otherwise considered non-humorous. Classification error in this experiment is the proportion of misclassified humorous and non-humorous data. We expect a context-dependent model to outperform a content or style-dependent model for cartoon/caption pairings. Practically, this means that the context-dependent model will agree with human labels, labeling humorous captions as humorous, and non-humorous captions as non-humorous (and will have a low classification error), whereas the content-dependent model will have high-classification error, disagreeing with humans on NYCC data.

References

Attardo, S. (2010). *Linguistic theories of humor* (Vol. 1). Walter de Gruyter.

Eco, U. (1985). Innovation and Repetition: Between Modern and Post-Modern Aesthetics. *Daedalus, 114*(4), 161-184.

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. In *Proceedings of the European Conference on Machine Learning.* (137–142).

McGraw, P., & Warner, J. (2014). *The humor code: A global search for what makes things funny.* Simon and Schuster.

Mihalcea, R., & Strapparava, C. (2006). Learning to laugh (automatically): Computational models for humor recognition. *Computational Intelligence*, *22*(2), 126-142.

van den Beukel, S., & Aroyo, L. (2018). Homonym Detection For Humor Recognition In Short Text. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis.* (286-291).