Context in Humor Detection Models

Isaac Ehrlich
Department of Cognitive Science, University of Toronto, Toronto, ON
isaac.ehrlich@mail.utoronto.ca

Joon Park
Department of Cognitive Science, University of Toronto, Toronto, ON
taejoon.park@mail.utoronto.ca

Abstract

Despite its common presence in everyday language and the direct practical implications it has on human cognition, such as memory, humor lacks a formal, computational definition. In an attempt to formalize humor, computational approaches have been taken using natural language processing models known as Support Vector Machines (SVM), a supervised machine learning algorithm. However, current models for automatic humor detection are exclusively focused on language, and do not account for the multisensory nature of human humor perception. Consequently, we expect context-dependent humor such as cartoons to be frequently misclassified by traditional humor-detection models. Motivated by work done by Mihalcea & Strapparava (2006), we first implement a standard content and syntax-based model of humor detection, and then extend this model with an additional context parameter. In doing so, we are able to increase SVM classification accuracy for context-dependent humorous data.

Keywords: Text classifier; SVM; Content; Context; Humor; Non-Humor; Syntax; Decision tree

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 contradictive to serious topics, others classify any event or object that elicits laughter as humorous (Warner & Mcgraw, 2014). Deepening 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 multi-sensory 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 specify relevant context.

To study this, we 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 is 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. We examine the success of traditional syntax and content based models on both context and non-context-dependent humor, and then apply this novel context-based humor detection model to visually-dependent humor, demonstrating that classification accuracy increases for context-dependent humor with this method.

Methods - Automatic Humor Detection

Using decision trees and SVMs, we analyzed automatic humor classification techniques based on the following features: (a) heuristic based syntactic analysis (alliteration and antonymy); (b) combined syntax and content-based features and; (c) combined syntax, content and context based features, integrated in a machine learning framework. Access to all of these models, the original data, and their results can be found at the following link: `https://github.com/imehrlich/humor_detect`.

Humor Identification – Stylistic and Syntactic Features

Previous studies have shown that syntactic features have significant impact in one's ability to identify humor (Ruch, 2002). There are several syntactic features of a sentence that may contribute to humor which includes puns, use of slang words, alliterations, and antonymy. For the purposes of this paper, we limited our search space to alliteration and antonymy in order to replicate the findings made by (Mihalcea & Strapparava, 2006).

Alliteration: The use of alliteration is seen in a variety of contexts, such as newspaper headlines and poetry. However, despite its commonplace, detecting alliteration in a sentence was suggested as a reliable heuristic to measure humor (Ruch, 2002), and shown effective in predicting humor in several experiments (Bucaria, 2006; Mihalcea & Strapparava, 2006). The following example is a humorous one-liner that contains an alliteration chain (underlined):

You can <u>tun</u>e a piano but you can't <u>tun</u>a fish.

In order to extract this alliteration measure from a sentence, each sentence was processed to filter out special characters such as quotations and punctuation. Furthermore, semantically cheap words, such as conjunctions and articles, were removed during the filtration process.

After processing, our algorithm searched for alliteration chains by transcribing all remaining words into the phonemes defined by the pronunciation dictionary. For example, the one-liner shown above, would be considered as having a count of one alliteration chain. To find this count, the algorithm would transcribe the words, "tune" and "tuna" to 'T UW1 N' and 'T UW1 N AH0' respectively. Since both words begin with 'T UW1 N,' an alliteration-detection function highlights this as a case of alliteration.

Antonymy: Antonyms are found when two words in opposite meaning occur within a sentence. Humor is often identified with contradictions and incongruity such as sarcasms and puns, and therefore the presence of antonyms can also be indicative of humor (Mihalcea & Strapparava, 2006). Some features that account towards contradictions might involve the tone of voice, or the context in which the sentence is spoken in. However, in the initial model, a simple approach was taken, focused on quantifying contradictions by identifying antonyms within a sentence. The following example is a one-liner that contains a set of antonyms (underlined):

Always try to be <u>modest</u> and be <u>proud</u> of it!

Before quantifying this measure, the sentences were once again reduced to exclude special characters, punctuation, and words without clear antonyms such as conjunctions and articles. After this filtering, we then referred to the WORDNET corpus (Miller, 1995) to search for antonyms of each words in the sentence. The total number of antonyms was then found by calculating each antonym of a given word found within the same sentence.

Once each of the phrases were processed, they were given a label for the total number of occurrences of antonyms and alliterations. With the labels attached to each of the phrases, the algorithm used a decision tree to predict whether the one-liner was from a humorous data set or a non humorous data set. In accordance to the decision tree implemented by Mihalcea & Strapparava (2006), if a phrase contained either alliteration or antonymy (or both), it was classified as humorous. The following is a graphical representation of a decision tree used:

```
alliteration = 0
    antonymy <= 1 : no
    antonymy > 1 : yes
alliteration > 0: yes
```

Figure 1: Sample decision tree for the application of stylistic features in understanding automatic humor recognition

To extend from the original work of the study, we wanted to confirm the efficacy of the binary classifier model, Support Vector Machines (SVM). Rather than using a heuristic approach, we used the features derived from the syntax model (alliteration and antonym) as an input into our SVM. With the total number of counts of stylistic features as our independent variable, we set aside 20% of the data and trained our algorithm with the remaining. Once the algorithm was trained, the efficacy of the algorithm was tested by feeding in the batch of test data into our model.

Humor Identification – Content and Stylistic Features

The content of the words themselves also play a major role in understanding humor. For a cognitive agent to understand words in a sentence, they need to be able to accurately conceptualize the ordering and meaning of words and characters. Once this correct interpretation is made, it is possible to accurately determine whether or not a specific content of the phrase is humorous.

For the purposes of accurate humor detection, we included content features in our model, and used SVMs, a common language processing model in order to develop an algorithm for humor classification.

Support Vector Machine is a supervised machine learning algorithm that is used as 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 algorithm plots the data in n-dimensional space (according to the number of features) to generate a hyper-plane that best differentiates the positive and the negative examples.

For our study, we applied a Support Vector Machine (SVM) to solve the problem of differentiating between the humorous and non-humorous data. Although alternative methods such as a Naives Bayes approach may offer reduced complexity and competence in categorizing text, we found SVMs to be applicable to the current context. This was due to SVM models showing higher accuracy in several text categorization tasks and humor detection experiments (van den Beukel & Aroyo, 2018; Mihalcea & Strapparava, 2006).

To achieve appropriate input for a language processing model, we processed the one-liners with the same process used in the heuristic model. Refer to the figure below:

```
> "That piece of cheese has not moved
in over an hour"
> ['piece', 'cheese','move', 'hour']
```

Figure 2: Sample distillation strategy

Once we distilled the phrase, the data was fed into the SVM as input and the data was trained with 80% of the humorous and the non-humorous data. Then, with the remaining data, the text classification model was tested to find for its classification accuracy.

In order to increase the accuracy of the humor detection model, the content of the words and the stylistic features were fed in as inputs to the SVM. As more inputs were given, it was hypothesized that the resulting accuracy of the model was to increase.

Humor Identification – Content, Stylistic Features, and Context

As an extension to the original work by Mihalcea & Strapparava (2006), we added another feature, context, into the SVM model to further the accuracy of our humor detection algorithm. To use context as an input to the SVM model, we manually annotated the contexts for each images. From there, we used a symmetry matrix to identify synonyms between the one-liner and the annotated context words. These synonym counts were then used as additional inputs to the SVM model.

Data

In order to accurately discriminate between humour and non-humor, our humor recognition models were trained using two humorous and two non-humours data set. The non-humorous data sets used in the study are 'ABC Headlines,' a collection of newspaper headlines, and 'Proverbs,' a collection of proverbs. The humorous data sets involved one context-independent data set, 'One-Liners'[1], and one context-dependent data set, 'New Yorker Caption Contest Data' (NYCC). The NYCC data set was contextually-dependent, as the text in this data set took the form of humorous captions, specifically written to particular image, also provided in the data set. Aditionally, we used Python's 'nltk' package to import a list of non-interesting, conjunction words that were used to process the data prior to extracting humorous features. Below are listed additional analysis-specific methods and data used to investigate different features of our data sets.

Alliteration: Once the data was filtered, we used CMU data, a database of phonemes, to extract phonemes for each individual words in a sentence.

Antonymy: For antonymy, we used WORDNET[2] in order to look for antonyms between a pair of words.

Content: In order to use a text classifier to accurately recognize humour, we imported numerous data sets from "nltk" and "sklearn" packages. Detailed imports can be seen by referring to the github link provided earlier on. Additionally, it is important here to note that since the SVM model takes numerical inputs as arguments, the features of the content data set were extracted into a numerical form using the 'TfidVectorizer' function.

Context: Contextual model used the Wordnet to search for synonyms between a list of words used to describe the context of an image and the content of the one-liner. From the number of synonym count generated, this was used as the context feature in the SVM model.

Gaussian SVM Simulation

Prior to fitting SVMs to language data, we explore the mechanisms of basic SVM models by applying it to two, two dimensional Gaussian clusters. Figure 3 displays the data used for this simulation.

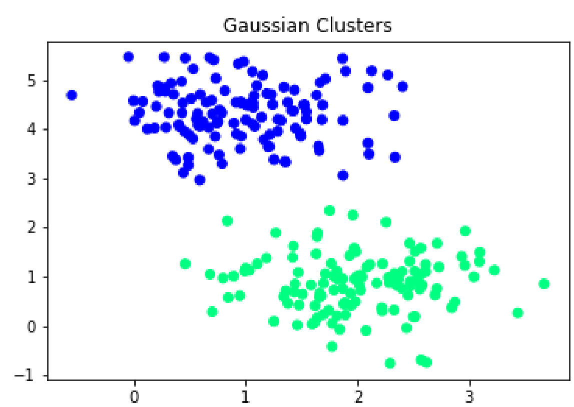

Figure 3. Easily separable Gaussian clusters used as proof-of-concept in an SVM simulation.

Since data used in most language-based SVM analyses is high-dimensional and difficult to visualize, this simulation provides a clear example of the mechanism underlying SVMs. In this example, we fit an SVM model to 500 normally sampled points (250 per group). Figure 4 visualizes the learned category boundary between the two Gaussian clusters.

[1]All data sets and original sources are available at `https://github.com/imehrlich/humor_detect/tree/master/Data`

[2]Available at: http://www.speech.cs.cmu.edu/cgi-bin/cmudict

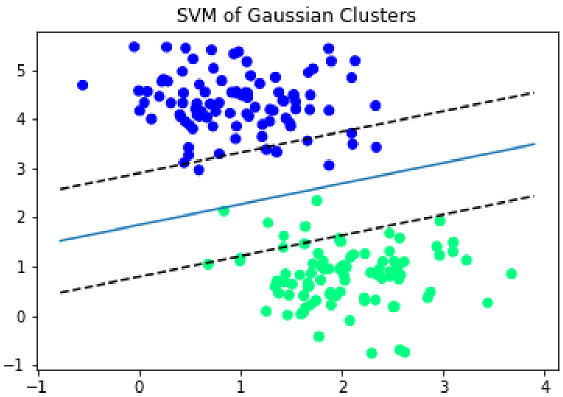

Figure 4. The classification boundary learned by the SVM, successfully discriminating between two Gaussian clusters.

Results

Syntax Analysis

In order to affirm the value of alliteration and antonymy in humor detection, the initial decision tree and SVM models were tested using only these features. Table 1 shows the resulting classification accuracy based on the syntactic features of alliteration and antonymy. Classification accuracy here is defined as the proportion of correctly assigned labels. SVMs outperformed Decision Trees for all data set pairings, however, all classifiers performed significantly above chance level (0.5). As expected, accuracy for One-Liners is also greater than accuracy for the NYCC data set. The SVM model also significantly outperformed the Decision Tree in all data set pairings.

Table 1: Classification Accuracy Based on Syntactic Features.

Data Sets	Decision Tree	SVM
One-Liner/ABC	0.6830	0.7800
One-Liner/Proverbs	0.6933	0.7600
NYCC/ABC	0.5770	0.6250
NYCC/Proverbs	0.5942	0.6300

Plotting the two-dimensional coordinates based on alliteration and antonymy also allows for further insight on the decision boundary derived by the syntactic-based SVM. The corresponding classification boundary can be seen in Figure 5. Agreeing with past research and intuition, the proximity of the boundary to the axes suggests that alliteration and antonomy are syntactic features more prevalent in humorous than non-humorous text.

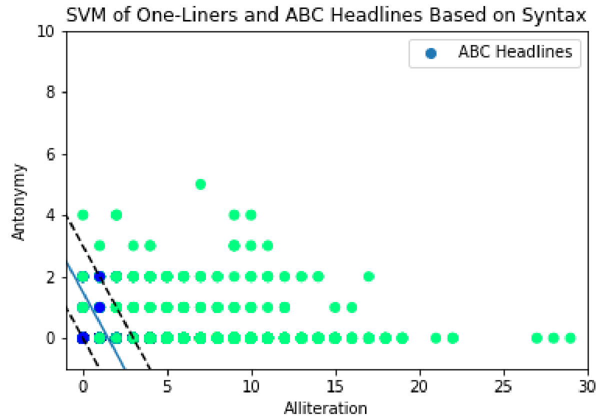

Figure 5. The SVM decision boundary for One-Liners and ABC headlines shows that alliteration and antonymy are more prevalent in humorous text.

Content Analysis

The addition of content as supplementary input to the original syntactic SVM model provided further increase in classification accuracy. The accuracy for the combined content and syntax SVM is provided in Table 2.

Table 2: Classification Accuracy Based on Content and Syntax.

Data Sets	Content SVM
One-Liner/ABC	0.9187
One-Liner/Proverbs	0.8382
NYCC/ABC	0.8908
NYCC/Proverbs	0.8072

A direct comparison between the two syntax models and combined content and syntax model reaffirms the increasing trend in classification accuracy, as well as more nuanced trends within the data. One-Liners continue to outperform the NYCC data set in their respective non-humorous pairings, and classification accuracy is higher for humorous data sets paired with ABC Headlines than humorous data sets paired with Proverbs. However, the most prevalent trend is the increase in accuracy between the syntax and content models.

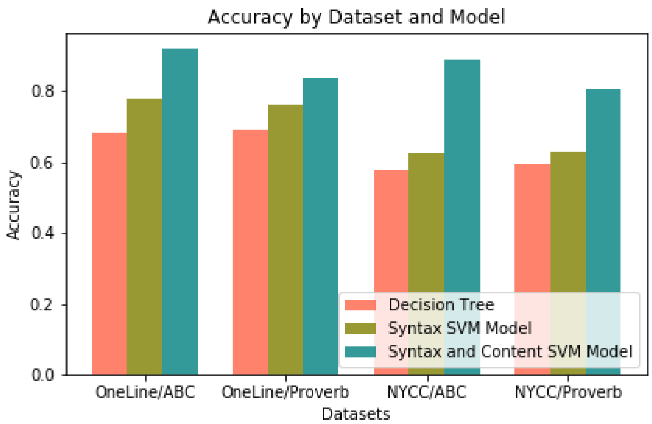

Figure 6. Comparison of content and syntax models across identical data set pairings show the improved accuracy offered by the content model.

Context Analysis

To examine the effect of an additional context parameter on context-dependent humorous data, we compare the classification accuracy of the context model on NYCC data to the classification accuracy of the content model on NYCC data. Figure 7 illustrates the slight increase in accuracy achieved by the context model.

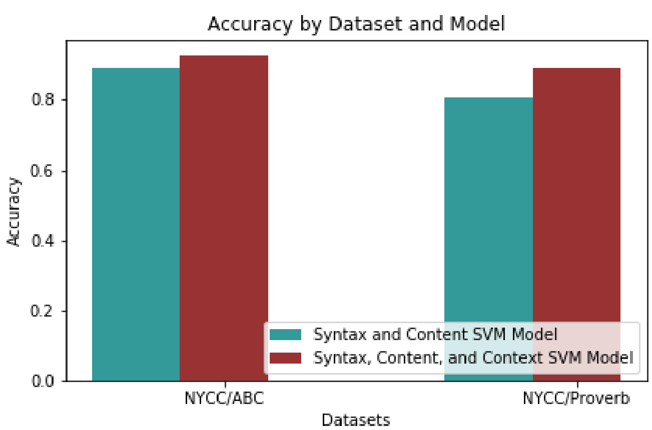

Figure 7. Comparison of the context and content models across identical data set pairings show the a slight improvement in accuracy offered by the context model.

Discussion

The high classification accuracy achieved from content-based SVMs reaffirms the conclusions made by Mihalcea & Strapparava (2006); both in terms of the value in syntactic parameters such as alliteration and antonymy, as well as the effectiveness of content-based humor classification. Despite slight differences in data sets (Mihalcea and Strapparava use Reuters Headlines instead of ABC), we were nevertheless also able to reproduce the differences in accuracy between non-humorous data sets, as both experiments observed higher accuracy when classifying headlines. This perhaps confirms the notion that non-humorous text corpuses exhibit substantial variation between themselves as well. Furthermore, similar high accuracy levels across both our studies suggest that (a) a computational approach to automatically identifying humor is appropriate, and (b) a content-based SVM model is able to correctly classify text across many different contexts with high efficacy.

However, while the content-based model was successful for all humor data set combinations, the classification accuracy for One-Liners remained higher than the accuracy for the NYCC data for all models in corresponding pairings. Despite the high classification accuracy, this disparity reaffirmed our hypothesis that a purely content-based model may not be optimal for contextual humor. This concept was further reinforced by the increased accuracy observed after implementing the context model. Ultimately, these findings suggest that environmental cues and contextual evidence may play an important role in certain types of humor, and these external factors should be considered when implementing automatic humor detection models.

However, though our results seem promising, there are still inherent drawbacks to our experimental design. Foremost, our investigation into possible syntactic predictors of humor is by no means extensive. Puns, for example, are a significant category of one-liners that we do not fully examine. While the play on words and similar sounds found in puns may be partially captured by our model for alliteration, the possible space of potential syntactic tendencies in puns is left largely unexplored. For example, our syntactic model did not capture the pun, 'What do you call a bee that lives in America? A U.S. Bee,' as there are no examples of alliteration or antonymy within the sentence. While we make progress by involving more external parameters to identify humor, there are still substantial syntactic features, such as puns, as well as additional multisensory inputs that are not considered in the model.

Another limitation to our study is found in our restriction to analysis of humor that is represented as one-liners or short text. While one-liners are a common form of humor, in natural settings, humor takes different forms, and can additionally manifest as a steady escalation through different environmental cues over long periods of time. Since our current model does not weigh results based on length of input, any long piece of text would likely be classified as humorous, as it has increased opportunity for more instances of syntactic predictors, as well as variation in content likely to be labeled as humorous.

Finally, extensions to our research may also be made by applying automated image recognition, or alternative models for text classification. While we annotated image contexts manually, a fully automated model would be able to identify context on its own merits. Our study does not make use of word-embeddings commonly employed in semantic analyses, nor do we explore how alternative text classifiers compare in this setting experimentally, such as Naive Bayes or multisensory neural networks, which could potentially offer similar or improved results. However, despite these limitations, by expanding traditional humor detection models to include external contextual information, we offer an improved model of humor detection for context-dependent data, as well as a new perspective on traditionally syntax and content oriented natural language processing tasks.

References

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

Bucaria, C. (2006). Lexical and syntactic ambiguity as a source of humor. *Humor*, *17*(3), 279–309.

Eco, U. (1985). *Innocation and repetition: Between modern and post-modern aesthetics*. Daedalus.

Hasan, M. K., Rahman, W., Zadeh, A., Zhong, J., Tanveer, M. I., Morency, L.-P., ... Hoque (2019). *Ur-funny: A multimodal language dataset for understanding humor.*

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

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

Miller, G. A. (1995, November). Wordnet: A lexical database for english. *Commun. ACM*, *38*(11), 39–41. Retrieved from `https://doi.org/10.1145/219717.219748` doi: 10.1145/219717.219748

Ruch, W. (2002). Computers with a personality? lessons to be learned from studies of the psychology of humor. In *Proceedings of the april fools day workshop on computational humour* (pp. 57–50).

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* (Vol. 9, pp. 286–291).

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