A robot and comedian walk into a bar, and… AHA!

Jamie Brew$^\zeta$
brew.jamie@gmail.com

Joseph Holt$^{\beta*}$
jpholt2@wisc.edu

Lalit Jain$^\delta$
lalitj@uw.edu

Robert Mankoff$^\epsilon$
bob@bobmankoff.com

Liam Marshall$^\alpha$
limarshall@wisc.edu

Robert Nowak$^\alpha$
rdnowak@wisc.edu

Rahul Parhi$^\alpha$
rahul@ece.wisc.edu

Scott Sievert$^{\alpha*}$
stsievert@wisc.edu

$^\alpha$Department of Electrical and Computer Engineering, University of Wisconsin–Madison
$^\beta$Department of Computer Sciences, University of Wisconsin–Madison
$^\delta$Department of Computer Sciences, University of Washington
$^\epsilon$Cartoon Collections, $^\zeta$Botnik Studios

Abstract

In New Yorker Cartoon Caption Contest (CCC), people compete to write the funniest caption for a given cartoon each week. We've been running this contest for several years, and have recorded people's ratings on the funniness of nearly 900,000 captions. Using these data, we propose the AI Humor Assistant (AHA) to help people write funnier captions. AHA does this by automatically judging which captions are funnier based on our historical rating data. AHA mirrors human performance in determining the funnier of two cartoon captions.

Introduction

Coming up with a joke is difficult for most people, and strategies for constructing jokes are often too general and difficult to apply to a specific context [2, 11, 4, 6]. Recent progress in NLP and machine learning promise to create computer-assisted tools for writers, but the applications of such tools have so far been limited.[2] More specialized suggestions have been developed for particular contexts. For example, Shahaf et. al. suggest how to make cartoon captions funnier; however these suggestions do not provide much guidance in terms of true creativity (e.g., it is suggested that captions with simple words in a unusual combinations are typical funnier) [13]. Can AI help people to be funnier?

Problem statement

Each week, The New Yorker runs a cartoon caption contest where users submit (supposedly) funny captions for a particular cartoon. To find the funniest captions, The New Yorker uses crowdsourcing to collect multiple ratings for each of the caption submissions (typically thousands of submissions).[3] We have collected over 100 million ratings on nearly 900,000 unique captions in over 150 contests.[4] With all this data, how can AI help people write funnier captions?

Contributions

To help humans write funnier captions in The New Yorker Cartoon Caption Contest (CCC) we have

---

[*]These authors implemented AHA. Authors are listed alphabetically by last name.

[2]e.g., https://www.grammarly.com

[3]The ratings for each caption are "unfunny", "somewhat funny", or "funny".

[4]http://github.com/nextml/caption-contest-data

- collected a **dataset** that has recorded over 100 million ratings on the funniness of nearly 900,000 unique captions. Anyone can rate captions at http://nextml.org/captioncontest.
- developed AHA, an interactive **AI web application** to help people write funnier captions for future caption contests. The backend uses a model trained on the dataset above to predict the funnier of any two captions given the context of a cartoon.

Further Details

Dataset The dataset was collected as part of a crowdsourcing project based on the NEXT system [5, 14] for interactive machine learning. The crowdsourcing employs multi-armed bandit algorithms to adaptively focus the data collection toward funnier captions [15]. This enables the crowdsourcing to quickly identify the funniest captions with high confidence.

This unique dataset contains over 100 million ratings and is available at http://github.com/nextml/caption-contest-data. Other humor datasets this large usually do not explicitly contain human judgments (e.g., scraping hashtags from Twitter) [10, 9, 3, 7] *or* are small and have human judgments (typically through a crowdsourcing service like Mechanical Turk) [8, 13].

AI web application Our web application AHA is a machine learning based tool designed to help users write interactively write funny captions. In our user interface, we show

- a cartoon image
- a text box for the user to write their own captions
- a recording of all captions the user has written in their session
- a ranking according to AI-predicted funniness of the captions[5]

This web application could be used during a brainstorming session to generate a funny caption for each week's contest. A beta version of AHA is visible at http://nextml.org/aha.

The backend model is a decision tree classifier [1] trained to predict the funnier of two captions given the context of a cartoon. This decision tree is trained on similar features to those used by Shahaf et. al. [13]. We validated our model against human performance by collecting human pairwise ratings for certain contests. Our model mirrors human performance, as shown in Figure 1.
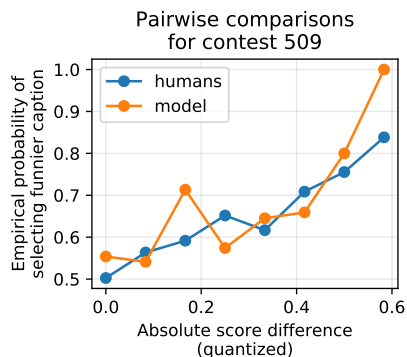

Figure 1: A comparison of human and model pairwise ratings. For contest 509, pairwise comparisons from humans were collected on the top 25 funniest captions. Our model is applied to these same captions. The ratings of "unfunny", "somewhat funny" and "funny" have scores of 1, 2 and 3 respectively.

Conclusion

We have developed AHA, an interactive AI tool, that helps users improve the quality of their humor. This tool compares two captions and selects the funnier one, allowing the user to interactively write funnier captions.

Future work involves providing more explanatory AI feedback and adding suggestions on what particular features make a caption funny. With our unique dataset, other natural language processing (NLP) tasks are feasible such as fully-automatic caption generation.

---

[5]via Borda scores [12]

References

[1] *Lightgbm: A highly efficient gradient boosting decision tree*, volume Advances in Neural Information Processing Systems, 2017.

[2] Salvatore Attardo. *Linguistic theories of humor*, volume 1. Walter de Gruyter, 2010.

[3] Luke de Oliveira and Alfredo Láinez Rodrigo. Humor detection in yelp reviews.

[4] Joyce Oramel Hertzler. Laughter: A socio-scientific analysis. Exposition Press, 1970.

[5] Kevin G Jamieson, Lalit Jain, Chris Fernandez, Nicholas J Glattard, and Rob Nowak. NEXT: A system for real-world development, evaluation, and application of active learning. *Advances in Neural Information Processing Systems*, page 2656–2656, 2015.

[6] Rod A Martin and Thomas Ford. *The psychology of humor: An integrative approach*. Academic press, 2018.

[7] Rada Mihalcea and Carlo Strapparava. Making computers laugh: Investigations in automatic humor recognition. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 531–538. Association for Computational Linguistics, 2005.

[8] Dragomir Radev, Amanda Stent, Joel Tetreault, Aasish Pappu, Aikaterini Iliakopoulou, Agustin Chanfreau, Paloma de Juan, Jordi Vallmitjana, Alejandro Jaimes, Rahul Jha, et al. Humor in collective discourse: Unsupervised funniness detection in the new yorker cartoon caption contest. *arXiv preprint arXiv:1506.08126*, 2015.

[9] Yishay Raz. Automatic humor classification on twitter. *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Student Research Workshop*, page 66–66, 2012.

[10] Antonio Reyes, Paolo Rosso, and Davide Buscaldi. From humor recognition to irony detection: The figurative language of social media. *Data & Knowledge Engineering*, 74:1–12, 2012.

[11] Willibald Ruch. The perception of humor. In *Emotions, qualia, and consciousness*, pages 410–425. World Scientific, 2001.

[12] Nihar B Shah and Martin J Wainwright. Simple, robust and optimal ranking from pairwise comparisons. *The Journal of Machine Learning Research*, 18(1):7246–7283, 2017.

[13] Dafna Shahaf, Eric Horvitz, and Robert Mankoff. Inside jokes: Identifying humorous cartoon captions. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 1065–1065, 2015.

[14] Scott Sievert, Daniel Ross, Lalit Jain, Kevin Jamieson, Rob Nowak, and Robert Mankoff. NEXT: A system to easily connect crowdsourcing and adaptive data collection. 2017.

[15] Ervin Tanczos, Robert Nowak, and Bob Mankoff. A kl-lucb algorithm for large-scale crowdsourcing. Advances in Neural Information Processing Systems:5894–5903, 2017.