# Joke Recommender System Using Humor Theory

**Soumya Agrawal (agraw105@purdue.edu)**
**Julia Taylor Rayz (jtaylor1@purdue.edu)**
Purdue University
West Lafayette, IN 47906 USA

**Abstract**

In this paper, we propose a methodology that aims to develop a recommendation system for jokes by analyzing its text. This exploratory study focuses mainly on the General Theory of Verbal Humor and implements the knowledge resources defined by it to annotate the jokes. These annotations contain the characteristics of the jokes and hence are used to determine how alike the jokes are. We use Lin's similarity metric and Word2vec to calculate the similarity between different jokes. The jokes are then clustered hierarchically based on their similarity values for the recommendation. Finally, for multiple users, we compare our joke recommendations to those obtained by the Eigenstate algorithm which does not consider the content of the joke in its recommendation.

**Keywords:** Computational humor; General theory of verbal humor; Clustering; Joke similarity

## Introduction

Humor is an interesting phenomenon that can be identified most of the time but is very difficult to 'define' (McGhee & Pistolesi, 1979). Yet, its importance becomes more evident with humorless technological advances. Humor is much more than just a source of entertainment; it is an essential tool that aids communication. Various empirical findings have confirmed that stress and depressing thoughts can be regulated with the help of humor (Francis, Monahan, & Berger, 1999). Positive psychology, a field that examines what people do well, notes that humor can be used to reduce tension, make friends, make others feel good, or to help buffer stress (Lurie & Monahan, 2015) (Ruch & Heintz, 2016).

The need for humor in a computerized setup is often discussed and many researchers have presented their findings. Some of the applications of computational humor are human-computer interfaces (Morkes, Kernal, & Nass, 1998), education (McKay, 2002), edutainment (Stock, 1996), understanding how human brain works (Binsted et al., 2006; Ritchie, 2001), etc.

The advancements in AI have sowed the seeds of the idea that computers can understand the human language. Since humor is a ubiquitous aspect of the human experience, it is fair to expect the computers to take into consideration the humorous facet. Almost two decades ago, it was pointed out that if computer systems can incorporate humor mechanisms, then these systems would appear to be more user-friendly hence less alien and intimidating (Binsted, 1995). This statement still holds and to achieve this, one of the key things to consider is that different people find different things funny which makes research in this field both challenging and interesting.

Verbally expressed or verbal humor is a common form of humor, and one of the subclasses of verbal humor is the joke. A joke can be defined as "a short humorous piece of literature in which the funniness culminates in the final sentence" (Hetzron, 1991). This paper focuses on verbally expressed humor with the help of jokes.

The motivation for this research comes from the observation that the smart assistants like Alexa and Siri recite the same jokes to all the users without considering their humor preferences. The idea behind this research is to come closer to understand human humor preferences and recommend jokes based on it. We propose a framework to recommend jokes to the users by taking into account the text of the joke as well as the liking of the users. Our assumption is that individuals like certain categories or types of jokes. These types can be identified through the individual's funniness ratings.

This framework is centered on the identification and quantification of similarity between jokes. The General Theory of Verbal Humor states that jokes can be represented and compared with the help of six knowledge resources (Attardo & Raskin, 1991). We use these knowledge recourses to find joke similarity in the Jester Dataset. Once similar jokes are identified, we explore whether subject ratings confirm the similarity.

There exists a joke recommendation system, Jester, (Goldberg, Roeder, Gupta, & Perkins, 2001) but it considers the users and the text of the joke as a black box and relies solely on the user ratings for the recommendation. It works as a baseline model to our proposed model and we compare the joke recommendations to the same user by both the models. We also analyze the ratings given by the users to the jokes that are considered similar to our model.

## Humor Theories

Humor studies date back to the era of Plato (*Philebus*) and Aristotle (*Poetics*). There are three major classes of humor theory: superiority theories, release/relief theories, and incongruity theories. The general idea behind superiority theories was that people laugh at other people's misfortunes since it makes them feel superior to them (Attardo, 1994) (Raskin, 1985). Release/relief theories assert that humor and laughter are a result of the release of nervous energy (Meyer, 2000). The family of incongruity theories states that humor arises when something which was not anticipated happens (Raskin, 1985). There has been a debate among various thinkers if incongruity alone can be considered to be sufficient enough to be able to mark something as funny (Suls, 1977).

This gave birth to the Incongruity-Resolution theories which focused not only incongruity but also on its realization and resolution. Suls (1972) proposed a two-stage model that stated that when there is some incongruity in the text, if one can resolve it then it's a joke otherwise the text leads to puzzlement and no laughter (Ritchie, 1999). Another model to resolve incongruity was summarized by Ritchie (1999) as the surprise disambiguation model which states that the setup of the joke has two different interpretations out of which one is more obvious than the other. The hidden meaning of the text is triggered once the punchline is reached.

Raskin's Script-based Semantic Theory of Humor (Raskin, 1985) is the first linguistic theory of humor. It is regarded as neutral concerning the three classes of humor theories. SSTH states that a joke carrying text should be fully or partially compatible with two scripts and these scripts must oppose. Raskin introduced several types of script oppositions, such as real/unreal, actual/non-actual, good/bad, life/death, sex/non-sex. The following joke is analyzed in Raskin (Raskin) with the scripts of Doctor and Lover[1] being the two scripts that overlap and oppose.

Joke$_1$*: 'Is the doctor at home?' the patient asked in his bronchial whisper. 'No,' the doctor's young and pretty wife whispered in reply. 'Come right in.'* (Raskin, 1985)

The joke evokes the script of a Doctor due to the words "doctor", "patient" and "bronchial". The second script, Lover, is triggered by the words "no" as well as the description of the doctor's wife. The wife's reply is incongruous to the first script, and thus the second script emerges, which makes the punchline, "come right in" explainable. The joke is said to have a partial script overlap between Doctor and Lover – both scripts contain a person that comes to the doctor's house for a visit – and since these scripts are opposing each other based on sex/non-sex, the text is considered a joke (Attardo, 1994) (Raskin, 1985).

Attardo and Raskin (1991) revised the SSTH into General Theory of Verbal Humor which stated that the jokes can be described using six knowledge resources (KRs) which are ordered hierarchically: script overlap/opposition (SO), logical mechanism (LM), situation (SI), target (TA), narrative strategy (NS), and language (LA). Upon empirical verification of the KR hierarchy, LM was found to behave differently than predicted (Ruch, Attardo, & Raskin, 1993). GTVH also made the comparison of jokes possible with the KRs. The higher the number of common parameters in jokes, the higher is joke similarity. Additionally, jokes that differ only in SO are less similar than the jokes that differ only in LM, than the jokes that differ only in SI and so on. For example, the following jokes are introduced in Attardo & Raskin (1991) to illustrate the comparison:

Joke$_2$: *"How many Irishmen does it take to screw in a light bulb? Five. One to hold the light bulb and four to turn the table he's standing on."*

Joke$_3$: *"How many Poles does it take to wash a car? Two. One to hold the sponge and one to move the car back and forth".*

Joke$_4$: *"Do you think one Pole can screw in a light bulb?" "No." "Two?" "No." "Three?" "No. Five. One to screw in a light bulb and four to turn the table he's standing on."*

The KRs representing these jokes are represented in Table 1:

Table 1: Joke Comparison (Attardo & Raskin, 1991)

| KR | Joke$_3$ | Joke$_4$ | Joke$_5$ |
|---|---|---|---|
| SO | Dumbness | Dumbness | Dumbness |
| LM | Figure-Ground Reversal | Figure-Ground Reversal | Figure-Ground Reversal |
| SI | Light Bulb | Car Wash | Light Bulb |
| TA | Irish | Poles | Poles |
| NS | Riddle | Riddle | Ques -Ans |
| LA | LA 1 | LA 1 | LA2 |

Here, jokes 3 and 4 differ in three of the parameters, namely, LA, NS, and SI; jokes 2 and 3 differ in two of them, namely TA and SI; and jokes 2 and 4 in three of them, namely LA, NS and TA. Jokes 2 and 3 are the most similar since they differ in only two knowledge resources. Since SI is placed at a higher level in the hierarchy, jokes 3 and 4 are the least similar even though they have the same number of different KRs as jokes 2 and 4. This paper will rely on this theory to process humor computationally.

## Methodology

We assume that previously unseen jokes should be recommended to users as well as jokes that have been rated by others (and thus, have been seen by the system). This means that the content of the jokes, not just the user ratings, has to be taken into consideration. To do so, we develop a methodology to compare jokes based on their content, find their similarity, and then cluster them accordingly. The jokes which are clustered together -- and have at least one highly rated joke – serve as the recommendations for the users.

### Corpus

This paper adopts jokes from the Jester dataset. We use version 3[2] of the dataset which is an updated dataset of the previous versions. Version 1 has rating values from -10 to +10 of 100 jokes collected between April 1999 to May 2003 and the version 2 has 50 more jokes with 115,000 new ratings collected between November 2006 to May 2009. Overall, the version 3 of the dataset has over 1.8 million continuous

[1] The naming of the scripts has been debated in various humor papers. The Ontological Semantic Theory of Humor (Raskin, Hempelmann, & Taylor, 2009) can be used to identify the scripts without committing to their naming.

[2] http://eigentaste.berkeley.edu/dataset/

ratings of 150 jokes from 54,905 anonymous users which were collected from November 2006 to March 2015. It should be noted that many jokes in the dataset are no longer relevant (out of date), but they can nevertheless be used to test the methodology. The dataset consists of a set of 8 jokes termed as *gauge set,* as these jokes are rated by all the users. The remaining non-gauge jokes have a very sparse rating matrix since around 82% of the user ratings are null.

All jokes from the dataset have been annotated with the six knowledge resources as defined by GTVH by the domain knowledge experts. We wish to point out that two pairs of jokes in the dataset are identical and we decide to remove the duplicate from measuring joke similarity.

## Baseline Model

The joke recommendation system (Goldberg et al., 2001) is based on a constant-time collaborative filtering algorithm that recommends jokes to the users based on their rating of the gauge set jokes. To overcome the problem of the sparse rating matrix, the model is built on the ratings of gauge set jokes only. The algorithm uses Principal Component Analysis (Pearson, 1901) to optimally reduce the dimension of the data to two. Since the projected data had a high concentration around the origin, a clustering algorithm was developed which recursively bisected the data near the origin into rectangle-shaped clusters. Whenever a user enters the system the ratings of the gauge set are collected which helps the algorithm to determine which cluster to place the user in. For each cluster, the mean of the non-gauge jokes ratings is calculated which are sorted in the decreasing order and this yields a lookup table. The lookup table is referenced every time a joke is recommended to the user.

## GTVH-based Framework for Joke Similarity

We analyze the text of the jokes based on the GTVH knowledge resource (KR) annotations done by the domain knowledge experts. We focused on SO, LM, SI and TA, as LA value should differ for every joke and most jokes in the dataset have the same NS value. To find the pairwise similarity of the jokes we compare the instances of the corresponding KRs. Attardo and Raskin (1991) do not define the hierarchy of each of the KRs, however, a sketch of SO hierarchy can be reconstructed from Raskin (Raskin), and a partial hierarchy of LMs can be found in (Attardo, Hempelmann, & Di Maio, 2002). We extended the hierarchies of SOs, LMs, and SIs based on the information from the jokes, using the methodology for ontology construction from the Ontological Semantic Technology – a foundation of the Ontological Semantic Theory of Humor.

To construct a hierarchy, each of the entities are described by their properties. The properties and their values serve the guiding principle for hierarchy construction (Taylor & Raskin). Each of the children differ from the parent by a property, and the siblings should differ from each other only by the values of the chosen property. Once the hierarchy is constructed, all descendants that do not have siblings are collapsed into a single node. In other words, no non-leaf node can have less than two children.

### Joke Similarity

The joke similarity metric for each of the resources is motivated by Resnik (1995) model, that proposed to estimate the common amount of information by the information content of the least common subsumer of the two nodes. Lin (1998) extended this concept by adding that the similarity metric must also take into account the differences between the two entities. To compare each instance of SO, LM and SI, the following function is used:

$$\text{Similarity}\,(kr_a, kr_b) = \begin{cases} 1 & if\ kr_a = kr_b \\ 0 & if kr_a\ or kr_b is\ null \\ sim_{Lin}\,(kr_a, kr_b) & all\ other\ cases \end{cases}$$

where $kr_a$ and $kr_b$ are the instances of the same KRs and $sim_{Lin}$ is Lin's similarity measure (Lin, 1998), adapted from Jurafsky and Martin (2018) used for word similarity:

$$sim_{Lin}\,(c_1, c_2) = \frac{2 * logP(LCS(c_1, c_2))}{logP(c_1) + logP(c_2)}$$

where P(c) is defined by as the probability that a random word selected in a corpus is an instance of concept c and $LCS(c_1, c_2)$ is the lowest node in the hierarchy that subsumes both $c_1$ and $c_2$. In our case, $c_1$ and $c_2$ are instances of a hierarchy of SO, LM, or SI.

To compare TA instances, we use word embeddings. Recent advancements in NLP research has seen the popularity of word embedding models which represent the words as vectors in a predefined vector space. One such word embedding model is word2vec (Mikolov, Chen, Corrado, & Dean, 2013) which is a shallow neural network that takes a text corpus as input and returns the vector representations of the words. To compare TAs, we used a pre-trained googlenews model which has word vectors for 3 million words, obtained by training on a google news dataset of around 100 billion words. There were some TA annotations in our dataset that were not present in the word2vec-based model. To overcome this problem, we made appropriate replacements of those annotations, ensuring that the new annotations preserve the context.

$Joke_5$ and $Joke_6$ illustrate joke annotation and calculation of joke similarity. We provide a modified version of the $Joke_5$ in this paper due to a potentially offensive nature of the original and replace $Joke_6$ with a very close joke taken from another source:

$Joke_5$: *A guys walks into a bar and tells the bartender that he has the best Polish joke. "I am Polish," responds the bartender. "Don't worry, I will tell it slowly."*

$Joke_6$: *"What did the liberal arts major say to the engineering grad?" "Do you want fries with that?"*[3]

[3] https://upjoke.com/liberal-art-jokes

The GTVH-based annotations for $Joke_5$ and $Joke_6$ are shown below:

$$\begin{matrix}\mathbf{SO}\\ \mathbf{LM}\\ \mathbf{SI}\\ \mathbf{TA}\end{matrix}\left\{\begin{matrix}\mathbf{actual/non-actual}\\ \mathbf{faulty\ reasoning}\\ \mathbf{going\ to\ bar}\\ \mathbf{poles}\end{matrix}\right\}\left\{\begin{matrix}\mathbf{actual/non-actual}\\ \mathbf{faulty\ reasoning}\\ \mathbf{intellectual\ discussion}\\ \mathbf{graduates}^{4}\end{matrix}\right\}$$

Since the instances of SO and LM are the same for both jokes, their corresponding similarity is 1. For TA, the word2vec similarity between *poles* and *graduates* is 0.046 using the methods defined by the Gensim library (version 3.8.1) on the pre-trained model. For SI, we look at the fragment of the SI hierarchy along with the P(c), as depicted in Figure 1. The nodes of interest are highlighted. This results in the following:

$$sim_{Lin}\left(SI_{going\ to\ a\ bar}, SI_{intellectual\ discussion}\right) = \frac{2 * \log(1)}{\log(0.0066) + \log(0.0066)} = 0$$

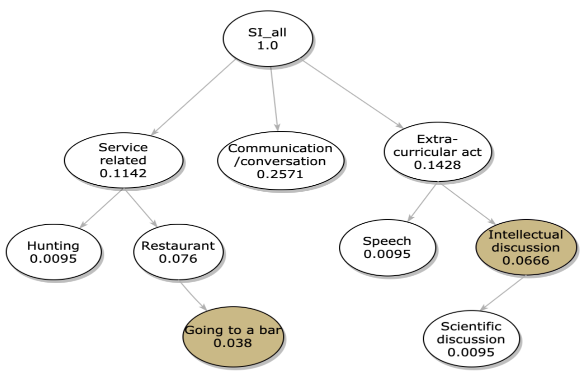


Figure 1: SI hierarchy fragment

To take into consideration the hierarchy of KRs themselves, as proposed by SSTH, we assign a weight, $w_{SO}$, $w_{LM}$, $w_{SI}$, and $w_{TA}$, to each of the KRs such that $w_{SO} < w_{LM} < w_{SI} < w_{TA}$:

$$\mathrm{sim}(joke_i, joke_j) = \frac{[w_{SO}\ \ w_{LM}\ \ w_{SI}\ \ w_{TA}]\begin{bmatrix}\mathrm{sim}(SO_{jokei}, SO_{jokej})\\ \mathrm{sim}(LM_{jokei}, LM_{jokej})\\ \mathrm{sim}(SI_{jokei}, SI_{jokej})\\ \mathrm{sim}(TA_{jokei}, TA_{jokej})\end{bmatrix}}{w_{SO}+w_{LM}+w_{SI}+w_{TA}}$$

For this paper, the following values are assigned: $w_{SO}=5$, $w_{LM}=4$, $w_{SI}=3$ and $w_{TA}=2$.

$$\mathbf{sim}(\mathbf{joke_5}, \mathbf{joke_6}) = \frac{[\mathbf{5\ 4\ 3\ 2}]\begin{bmatrix}\mathbf{1}\\ \mathbf{1}\\ \mathbf{0}\\ \mathbf{0.046}\end{bmatrix}}{\mathbf{5+4+3+2}} = \frac{\mathbf{0.092}}{\mathbf{15}} = \mathbf{0.649}$$

The value calculated after the weighted average is 0.649 which quantifies how similar $Joke_5$ and $Joke_6$ are.

[4] Annotation has been changed from *liberal arts graduate* to *graduates* since the former was not in word2vec-based-model

### Joke Clustering

We aim to cluster the jokes so that the most similar ones are close to each other. To achieve this, we implement the hierarchical clustering algorithm in which we use the joke similarity values for the distance calculation. Figure 2 shows the dendrogram of the jokes after clustering.

### Joke Recommendations

To recommend jokes to a user, we would identify the user's favorite joke with the help of the user ratings of the corpus. The joke which is the immediate sibling of the favorite joke in the dendrogram is recommended first. To recommend more jokes, we move up in the hierarchy and if there are multiple jokes available on the same level of the hierarchy, then a random selection of the jokes is done for that level.

## Results

A qualitative evaluation was performed on the GTVH-based model. A user was randomly selected for comparison of the recommendations made by the baseline and the GTVH-based model. To compute the top recommend jokes from the GTVH-based model, we use the selected user's top-rated joke from the dataset which is known to our system as the favorite joke. The same user's ratings of the recommended jokes from both the models were used to compare them. Table 2 shows the results for five randomly selected users.

We are restricted in selecting the users due to the sparsity of the rating dataset which sheds light on one of the difficulties with working with this dataset. We meticulously select report results on the users who have rated the jokes in both the baseline and the GTVH, to ensure that the comparison of the two models is possible. The results for randomly selected 5 users are shown in Table 2. The highest-rated joke for user 1, as well as recommended jokes by the baseline and the GTVH-based model are presented as well. For user 1 in Table 2, we observe that the top joke recommended by the GTVH-based model (Joke 87) has a better rating than the top joke recommended by the baseline model (Joke 89). We can see by the text of the jokes that the favorite joke of user 1 and Joke 87 are very similar whereas Joke 89 is very different from these jokes. We provide the modified versions of some of jokes from the dataset for the analysis.

User1's favorite joke: *An artist has been displaying his paintings at an art gallery and he asked the owner if there had been any interest in his paintings. "I've got good news and bad news," says the owner. "The good news is that a gentleman inquired about your work and wondered if it would be worth more after your death. When I told him it would, he bought all ten of your paintings." "That's wonderful!" the artist says. "What's the bad news?" With concern, the gallery owner replied: "The man was your doctor."*

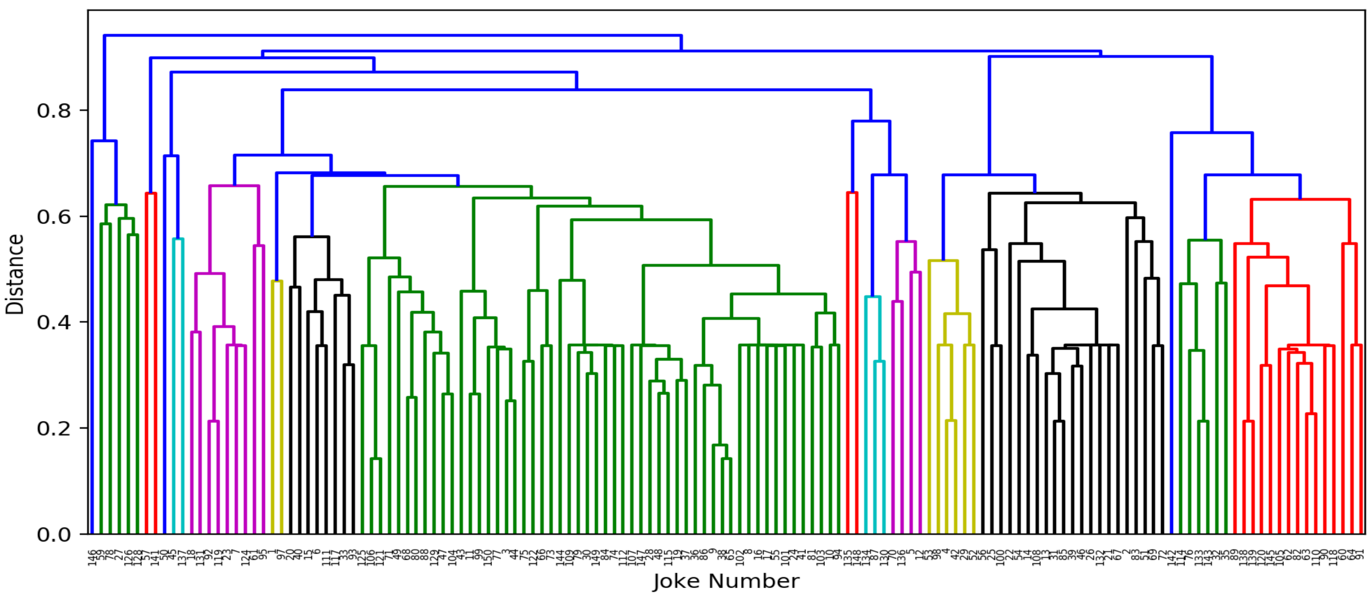

Figure 2: Hierarchical clustering of jokes

Table 2: Comparison of Recommended Jokes

| | **Baseline Model** | | **GTVH-based Model** | |
|---|---|---|---|---|
| | **Top Recommended Joke** | **Rating** | **Top Recommended Joke** | **Rating** |
| User 1 | Joke 89 | 8.18 | Joke 87 | 9.37 |
| User 2 | Joke 73 | -1 | Joke 42 | 5.71 |
| User 3 | Joke 53 | 3.56 | Joke 72 | 3.46 |
| User 4 | Joke 5 | 9.87 | Joke 112 | 0.93 |
| User 5 | Joke 89 | 4.56 | Joke 126 | 5.62 |

Table 3: Cluster Analysis for the five selected users

| | **Cluster 1** | | **Cluster 2** | | **Cluster 3** | | **Cluster 4** | | **Cluster 5** | | **Cluster 6** | | **Cluster 7** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Joke Id** | **92** | **119** | **106** | **121** | **38** | **65** | **31** | **85** | **133** | **143** | **138** | **139** | **110** | **63** |
| **User 6** | 3.02 | 2.88 | 3.39 | 2.91 | 3.10 | 3.58 | 4.29 | 3.22 | 2.66 | 3.46 | 2.86 | 3.05 | 3.13 | 3.64 |
| **User 7** | *2.45* | *4.22* | 3.45 | 4.66 | *4.91* | *3.13* | *1.02* | *4.14* | 0.88 | 0.87 | 4.68 | 4.14 | 0.99 | 0.95 |
| **User 8** | *1.40* | *4.31* | 4.53 | 3.30 | 4.25 | 3.36 | 3.65 | 3.45 | 2.94 | 4.17 | 3.16 | 3.29 | 3.83 | 4.22 |
| **User 9** | 3.63 | 4.12 | 4.80 | 4.96 | *4.79* | *1.35* | 2.68 | 1.75 | 4.28 | 3.13 | 3.13 | 3.19 | 2.82 | 2.77 |
| **User 10** | *0.19* | *4.31* | 3.85 | 4.34 | *3.90* | *1.21* | *4.31* | *0.13* | 4.41 | 3.81 | 1.64 | 1.07 | 1.85 | 2.90 |

Joke 87: *A man after undergoing a routine physical examination receives a phone call from his doctor. The doctor says, "I have some good news and some bad news." The man says, "I want to hear the good news first." The doctor says, "The good news is, you have 24 hours to live." The man replies, "If this is good news then what's the bad news?" The doctor says, "The bad news is, I forgot to call you yesterday."*

Joke 89*: A radio conversation between a US naval ship and Canadian authorities... Americans: Please divert your course 15 degrees to the North to avoid a collision. Canadians: Recommend you divert YOUR course 15 degrees to the South to avoid a collision. Americans: You divert YOUR course. Canadians: No. You divert YOUR course. Americans: This is the second largest ship in the United States; Atlantic Fleet. We are accompanied by three destroyers, three cruisers and numerous support vessels. I demand that you change your course…., or countermeasures will be undertaken to ensure the safety of this ship. Canadians: This is a lighthouse. Your call.*

The proposed model works better than the baseline for users 1, 2 and 5, works moderately well for user 3, and fails to perform better for users 4. It should be noted that it is equally possible to find similar jokes to all highly rated jokes for a particular user. However, based on the results of user 4, we wanted to check whether highly similar jokes are typically rated similarly.

To further investigate how users rate jokes that are considered similar by the proposed model, we selected 5 users who have rated 140 jokes which the maximum number of jokes rated by any user. Also, we normalize the ratings to 0-5 for the experiment. We selected all the joke clusters which are formed near the distance value of 0.2 for the analysis. Table 3 lists 7 such clusters each consisting of 2 jokes for the comparison of user ratings of closely clustered, and thus similar, jokes. We observe that the intra-cluster ratings of the users 6, 8 and 9 are largely similar for all the clusters where they differ greatly for user 7 and user 10. Both these users rate jokes in clusters 1, 3 and 4 were differently which implies that the jokes which are considered similar by the model are not equally appreciated by both the users. There are several explanations for this result, assuming that the ratings in the dataset accurately represent user preferences. The first one is that the similarity metric that we produced does not accurately represent joke similarity, and Target may need to be weighted heavier than the rest for the recommendation system. The second one is that the annotation of one of the jokes in the clusters may be flawed. The third, and perhaps most interesting one, is whether users tend to rate familiar jokes lower. We do not have the data on the ordering of jokes that were presented to the users, and thus, this is impossible to test this hypothesis. However, we can look at rating of almost identical jokes for these users.

As stated earlier, the corpus has ratings of two pairs of identical jokes, ratings of which for the same 5 users are summarized in table 3.

We can observe that users 7, 8 and 9 have given different ratings to identical jokes. Since the users were given a scroll button to rate the jokes, some variation in the ratings is acceptable but this difference is very high for users 7 and 8. It is tempting to conclude that the effect of a previously heard or rated joke must be taken into consideration while recommendations are made. It is also possible that for some users the almost identical jokes were presented very close to each other, while for others they were spread much farther apart among the 140 jokes. Lastly, the dataset also does not consider the effect fatigue effects of the users which may affect the ratings.

Table 3: Ratings given to Identical Jokes

| | **Identical Pair 1** | | **Identical Pair 2** | |
|---|---|---|---|---|
| **User 6** | 5.12 | 0.62 | 3.15 | 1.37 |
| **User 7** | *-5.25* | *4.68* | -7.31 | -5.84 |
| **User 8** | 5.81 | 0.62 | *9.62* | *1.68* |
| **User 9** | 5.12 | 5.59 | *3.53* | *7.28* |
| **User 10** | 4.78 | 0.53 | 1.15 | 5.34 |

## Conclusion

By taking into account the text of the jokes along with the user rating for joke recommendations, we observe that the model can select similar jokes, however, it is not clear that this by itself is the winning mechanism. To attain a more generalized framework for joke recommendations we need to 1) Conduct more research focusing on the manipulation of the weights assigned to the KRs 2) Collect user ratings while keeping track of the order of jokes in which they appear, thus taking into consideration the effect of a previously heard joke. We suspect that understanding user preference will go a long way towards more friendly interaction between various devices that have a functionality of telling a user a joke.

## Acknowledgement

The authors are grateful to anonymous reviewers for their helpful comments and to Ken Goldberg for releasing the dataset to the public.

## References

Attardo, S. (1994). Linguistic theories of humor (Vol. 1). *Berlin: De Gruyter Mouton. Barsoux, JL (1996). Why organisations need humour. European Management Journal, 14*(5).

Attardo, S., Hempelmann, C. F., & Di Maio, S. (2002). Script oppositions and logical mechanisms: Modeling incongruities and their resolutions. *Humor - International Journal of Humor Research, 15*(1). doi:10.1515/humr.2002.004

Attardo, S., & Raskin, V. (1991). Script theory revis (it) ed: Joke similarity and joke representation model. *Humor-International Journal of Humor Research, 4*.

Binsted, K. (1995). *Using humour to make natural language interfaces more friendly.* Paper presented at the Proceedings of the AI, ALife and Entertainment Workshop, Intern. Joint Conf. on Artificial Intelligence.

Binsted, K., Nijholt, A., Stock, O., Strapparava, C., Ritchie, G., Manurung, R., . . . O'Mara, D. (2006). Computational humor. *IEEE Intelligent Systems, 21*(2).

Francis, L., Monahan, K., & Berger, C. (1999). A laughing matter? The uses of humor in medical interactions. *Motivation and emotion, 23*.

Goldberg, K., Roeder, T., Gupta, D., & Perkins, C. (2001). Eigentaste: A constant time collaborative filtering algorithm. *information retrieval, 4*.

Hetzron, R. (1991). On the structure of punchlines. *HUMOR: International Journal of Humor Research*.

Jurafsky, D., & Martin, J. H. (2018). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. *Inc. Upper Saddle River, NJ*.

Lin, D. (1998). *An information-theoretic definition of similarity.* Paper presented at the Icml.

Lurie, A., & Monahan, K. (2015). Humor, aging, and life review: Survival through the use of humor. *Social Work in Mental Health, 13*.

McGhee, P. E., & Pistolesi, E. (1979). *Humor: Its origin and development*: WH Freeman San Francisco.

McKay, J. (2002). Generation of idiom-based witticisms to aid second language learning. *Stock et al*.

Meyer, J. C. (2000). Humor as a double-edged sword: Four functions of humor in communication. *Communication theory, 10*.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv.org*.

Morkes, J., Kernal, H. K., & Nass, C. (1998). *Humor in task-oriented computer-mediated communication and human-computer interaction.* Paper presented at the CHI 98

Conference Summary on Human Factors in Computing Systems.
Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2*. doi:10.1080/14786440109462720
Raskin, V. (1985). *Semantic mechanisms of humor*. Dordrecht ; Boston: D. Reidel Pub. Co.
Raskin, V., Hempelmann, C. F., & Taylor, J. M. (2009). How to understand and assess a theory: The evolution of the SSTH into the GTVH and now into the OSTH. *Journal of Literary Theory, 3*.
Resnik, P. (1995). Using Information Content to Evaluate Semantic Similarity in a Taxonomy. *arXiv.org*.
Ritchie, G. (1999). *Developing the incongruity-resolution theory*.
Ritchie, G. (2001). Current directions in computational humour. *Artificial Intelligence Review, 16*.
Ruch, W., Attardo, S., & Raskin, V. (1993). Toward an empirical verification of the general theory of verbal humor. *Humor, 6*.
Ruch, W., & Heintz, S. (2016). The virtue gap in humor: Exploring benevolent and corrective humor. *Translational Issues in Psychological Science, 2*.
Stock, O. (1996). *Password Swordsfish: Verbal Humor in the Interface*: Citeseer.
Suls, J. M. (1972). A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis. *The psychology of humor: Theoretical perspectives and empirical issues, 1*.
Suls, J. M. (1977). Cognitive and disparagement theories of humor: A theoretical and empirical synthesis. *Chapman and Foot, It's A Funny Thing Humour*.
Taylor, J. M., & Raskin, V. (2015). Comparing Apples and Orange Cottages: Classifications and Properties. *The Seventh International Conference on Advanced Cognitive Technologies and Applications*.