Mining Effective Features Using Quantum Entropy for Humor Recognition

Yang Liu and **Yuexian Hou**
College of Intelligence and Computing, Tianjin University, Tianjin, China
{lauyon,yxhou}@tju.edu.cn

Abstract

Humor recognition has been studied with several different methods in the past years. However, existing studies on humor recognition do not understand the mechanisms that generate humor. In this paper, inspired by the incongruity theory, any joke can be divided into two components (the setup and the punchline). Both components have multiple possible semantics, and there is an incongruous relationship between them. We use density matrices to represent the semantic uncertainty of the setup and the punchline, respectively, and design Quantum Entropy Uncertainty (QE-Uncertainty) and Quantum Entropy Incongruity (QE-Incongruity) with the help of quantum entropy as features for humor recognition. The experimental results on the SemEval2021 Task 7 dataset show that the proposed features are more effective than the baselines for recognizing humorous and non-humorous texts.

1 Introduction

Humor is one of the most distinctive features of human behavior and a sign of mental maturity (Pasquali, 1990). The study of humor has received extensive attention in the fields of linguistics, philosophy, psychology, and sociology (Mihalcea et al., 2010). Computational humor is of particular interest, with the potential to transform computers into creative and motivational tools (Nijholt et al., 2003).

This paper restricts research to humor recognition in computational humor, which aims to recognize whether a piece of text is humorous. As shown in Figure 1, a joke usually includes two components: the setup and the punchline. The reader generates an expectation of the following text (the punchline) based on the content of the setup, and if the following text violates the reader's expectation, humor is generated, and vice versa.

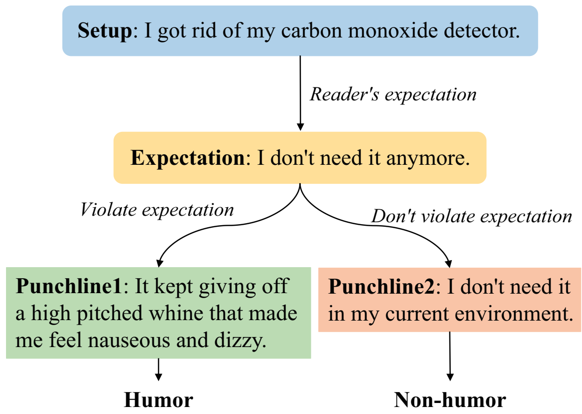

Figure 1: A humor and non-humor example containing the setup and the punchline.

In fact, the incongruity theory of humor can explain the above process of producing humor. The incongruity theory states that humor is generated because a thing (the setup) has multiple underlying concepts, and there is an incongruity between the concept involved in the situation and the real object it represents (the punchline).

Features based on semantic similarity (Mihalcea et al., 2010; Yang et al., 2015) and word association (Liu et al., 2018; Cattle and Ma, 2018) have achieved certain results, but they lack consideration of humorous mechanisms. Xie et al. (2021) calculated the uncertainty and the surprisal values of the joke with the help of the GPT-2. But they did not model the semantic incongruity between the setup and the punchline. While the above approaches are somewhat effective, the incongruity theory requires us to model semantic uncertainty and the incongruity between the setup and the punchline. We take inspiration from quantum theory and use density matrices to represent the uncertainty of text semantics. Specifically, the setup and the punchline are represented as density matrices, respectively. Then, take the quantum entropy of the setup as **Q**uantum **E**ntropy **Uncertainty** (**QE-Uncertainty**) and the conditional quantum entropy

Findings of the Association for Computational Linguistics: EACL 2023, pages 2048–2053
May 2-6, 2023 ©2023 Association for Computational Linguistics

between the setup and the punchline as **Q**uantum **E**ntropy **Incongruity** (**QE-Incongruity**). Experiments conducted on a manually-labeled dataset demonstrate that these two features are better than existing baselines in distinguishing between humorous and non-humorous texts, confirming the necessity of correlating semantic uncertainty with quantum theory.

2 Background

2.1 The Incongruity Theory

The most widely accepted theory for explaining humor is the incongruity theory. The theory suggests that laughter is caused by an incongruity between the understanding of the text and its actual meaning (Mulder and Nijholt, 2002). Immanuel Kant describes humor as "*the sudden transformation of a strained expectation into nothing* (Hickey-Moody and Laurie, 2017)." Schopenhauer (1966) also believed that perceived incongruity exists between a concept and the real object it represents. The incongruity theory has also been developed in the field of linguistics. The Semantic Script-based Theory of Humor (SSTH) proposed by Raskin (1979) is a scripted expression of the incongruity theory. SSTH is our bridge to mathematically model the incongruity theory. SSTH requires humorous texts to meet the following conditions: (1) The text is compatible, fully or in part, with two different (semantic) scripts. (2) The two scripts with which the text is compatible are opposite.

2.2 Density Matrix

The mathematical form of quantum mechanics represents the probability space as a vector space (i.e., the Hilbert space $\mathbb{H}^n$) (Von Neumann, 2018). Researchers often use Dirac's notation to represent unit vectors in this space. For example, a unit vector $\vec{u}$ and its transpose $\vec{u}^T$ are represented as $|u\rangle$ and $\langle u|$, respectively. The inner product of two unit vectors $|u\rangle$ and $|v\rangle$ is written as $\langle u|v\rangle$. The projector onto the direction $|u\rangle$ is its own outer product $|u\rangle\langle u|$. The rank of each projector is one and each projector represents a quantum fundamental event, often called a *dyad*. The density matrix (Nielsen and Chuang, 2010) is a generalization of the classical probability distribution. A density matrix ρ can be defined as a mixture of dyads:

$$\rho = \sum_{i=1}^{n} p_i |\psi_i\rangle\langle\psi_i| \tag{1}$$

where $|\psi_i\rangle$ represents a pure state with probability p_i. The density matrix ρ is symmetric, positive semi-definite, and its trace is one.

2.3 Quantum Entropy

Quantum entropy is a generalization of the quantum case of classical Shannon entropy (Shannon, 1948). If a quantum system is described by a density matrix ρ, its quantum entropy (Von Neumann, 2018) is defined as:

$$\mathrm{S}(\rho) = -\mathrm{tr}(\rho \ln \rho) \tag{2}$$

The conditional quantum entropy (Cerf and Adami, 1999) of the density matrix σ given the known density matrix ρ is defined as:

$$\begin{aligned} \mathrm{S}(\sigma|\rho) &= \mathrm{S}(\sigma\rho) - \mathrm{S}(\rho) \\ &= -\mathrm{tr}(\sigma\rho \ln(\sigma\rho)) + \mathrm{tr}(\rho \ln \rho) \end{aligned} \tag{3}$$

unlike classical conditional entropy, conditional quantum entropy can be negative.

3 Methodology

The incongruity theory holds that the prerequisite for humor is that the text has multiple semantic aspects. The reader does not understand one meaning but expects one while the punchline provides another, leading to incongruity. According to the incongruity theory, we should design features to represent the multiple semantic overlaps of the setup, as well as the incongruity of the semantics of the setup and the punchline.

Normalize each word $w_i \in V$ as follows:

$$|w_i\rangle = \frac{\vec{w_i}}{\|\vec{w_i}\|} \tag{4}$$

where $\|\cdot\|$ represents the L_2-norm. The representation of each word can be viewed as a superposition in Hilbert space.

A sentence of length l is represented by an n-by-n density matrix ρ:

$$\rho = \frac{1}{|l|} \sum_{i=1}^{l} |w_i\rangle\langle w_i| \tag{5}$$

where the diagonal values of ρ reflect the superposition semantics of sentences, and the non-diagonal values encode the correlation between semantics in a quantum way.

3.1 QE-Uncertainty

We take evidence of humor recognition from the setup, model the setup as a density matrix to represent its uncertainty semantics, and use the quantum entropy of the density matrix to represent the value of uncertainty. Formally, the QE-Uncertainty is calculated as follows:

$$\mathrm{U}(\rho) = -\mathrm{tr}(\rho \ln \rho) \tag{6}$$

where ρ represents the density matrix of the setup. The value of QE-Uncertainty reflects the amount of information contained in the text and the uncertainty of semantics. The larger the value, the more information the text contains, and the more likely the text is humorous.

3.2 QE-Incongruity

Another aspect of the incongruity theory is how different the semantics of the punchline is from expectations when the semantics of the setup are known (i.e., how much information we don't know about the punchline). In other words, how much information about the punchline is included in the setup? Specifically, the QE-Incongruity is defined as follows:

$$\begin{aligned} \mathrm{I}(\sigma|\rho) &= \mathrm{U}(\sigma\rho) - \mathrm{U}(\rho) \\ &= -\mathrm{tr}(\sigma\rho \ln(\sigma\rho)) + \mathrm{tr}(\rho \ln \rho) \end{aligned} \tag{7}$$

where ρ and σ represent the density matrices of the setup and the punchline, respectively. The value of QE-Incongruity describes how unknown the semantics of the punchline is when the setup is known. We argue that when the setup contains less semantics in the punchline, there will be incongruity, and there will be humor.

4 Related Work

The existing text humor recognition methods are mainly divided into feature-based methods and deep learning-based methods. Mihalcea and Strapparava (2005) use automatic classification techniques to integrate humor-specific features (alliteration, antonymy, slang) and content-based features into a machine-learning framework for humor classification tasks. Mihalcea et al. (2010) divide the humor text into two components: the setup and the punchline. Humor recognition is performed by calculating the semantic correlation between the setup and the punchline based on the incongruity theory. Morales and Zhai (2017) use a generative language model combined with background text resources to construct multiple features to identify whether a comment is a humorous text. Liu et al. (2018) combine discourse analysis and sentiment analysis to extract sentiment-related features to address humor recognition. Xie et al. (2021) developed uncertainty and superisal with the help of the prediction results of the pre-trained language model GPT-2. In recent years, with the development of deep learning, some deep learning-based methods have been proposed. Chen and Lee (2017) use convolutional neural networks to identify humor in the TED talks corpus. Chen and Soo (2018) used the highway network architecture to implement deep convolutional neural networks to predict humor on datasets of different types and different languages. Weller and Seppi (2019) used pre-trained BERT for the humor classification task. Fan et al. (2020) combine the Bi-GRU network with phonetic structure and ambiguity for humor recognition.

5 Experiments

5.1 Settings

We build a Support Vector Machine (SVM) classifier for humor classification. Experiments are performed on the SemEval 2021 Task 7[1] dataset modified by Xie et al. (2021). The dataset consists of a total of 3,052 labeled samples, half of which are humor and the other half are non-humor. The text of each sample in the dataset is split into two parts (the setup and the punchline). For each sample in the dataset, the lengths of the setup and the punchline are both below 20, and the percentage of alphabetical letters is greater than 75%, all of which start with alphabetical letters. We use Accuracy(Acc), Precision(P), Recall(R) and F1-Score(F1) as the evaluation metrics. P, R and F1 are macro-averaged. The experiments adopt 10-fold cross-validation, and the result is the average value of repeated experiments.

5.2 Baselines

Semantic similarity and semantic distance are the most commonly used text features, and we choose three such features as our baselines:

- **Path similarity** (Rada et al., 1989) is a similarity measure based on the shortest path, defined as follows:

$$\mathrm{Sim}_{path} = \frac{1}{1 + \mathrm{D}(c_1, c_2)} \tag{8}$$

[1] https://semeval.github.io/SemEval2021/

where $\mathrm{D}(c_1, c_2)$ represents the shortest path in WordNet between concepts c_1 and c_2.

- **Disconnection** (Yang et al., 2015) is defined as the maximum distance between word pairs in the text.
- **Repetition** (Yang et al., 2015) is defined as the minimum distance between word pairs in the text.

In addition, we consider two GPT-2 based features proposed by Xie et al. (2021) as baselines. They feed the text into GPT-2 model to predict the next token. While predicting the tokens of y, GPT-2 produces a probability distribution v_i over the vocabulary.

- **Uncertainty** is obtained by calculating the average entropy of the probability distribution v_i on the vocabulary, defined as:

$$\mathrm{U}(x, y) = -\frac{1}{|y|} \sum_{i=1}^{n} \sum_{w \in V} v_i^w \log v_i^w \tag{9}$$

where n represents the length of y and V is the vocabulary.

- **Surprisal** describes the degree of surprise when the language model generates the punchline, which is defined as follows:

$$\begin{aligned} \mathrm{S}(x, y) &= -\frac{1}{|y|} \log \mathrm{p}(y|x) \\ &= -\frac{1}{|y|} \sum_{i=1}^{n} \log v_i^{y_i} \end{aligned} \tag{10}$$

5.3 Predict Using Individual Features

Table 1 shows the results of individual feature prediction. Compared with the baselines, our proposed features QE-Uncertainty and QE-Incongruity achieve higher scores on all four metrics, with QE-Incongruity achieving the best results. In particular, compared with Uncertainty based on classical Shannon entropy, QE-Uncertainty under our quantum framework is greatly improved. This shows the necessity of quantum generalization for semantic uncertainty problems.

5.4 Boost a Content-Based Classifier

To demonstrate the effectiveness of our proposed features combined with content-based classifiers. We use the 50-dimensional GloVe (Pennington et al., 2014) embedding as the baseline. We encode the setup and the punchline as the average of their respective word embeddings, resulting in two vectors with dimensions 50. Concatenate these two vectors with our features to form a vector with dimension 101. Finally, put it into an SVM classifier for humor classification. The results are shown in Table 2, our features achieve higher improvements on content-based classifiers compared to baselines.

Table 1: Experimental results of individual features. The results for features with an asterisk are reported by Xie et al. (2021).

Features	P	R	F1	Acc
Random	0.5000	0.5000	0.5000	0.5000
Sim_{path}	0.5123	0.5070	0.4555	0.5062
Disconnection	0.6475	0.5503	0.4610	0.5501
Repetition	0.5592	0.5577	0.5538	0.5567
Uncertainty*	0.5840	0.5738	0.5593	0.5741
Surprisal*	0.5617	0.5565	0.5455	0.5570
QE-Uncertainty	0.6589	0.6318	0.6146	0.6314
QE-Incongruity	**0.6690**	**0.6450**	**0.6319**	**0.6451**

Table 2: Experimental results of concatenating a content-based classifier. The results for features with an asterisk are reported by Xie et al. (2021).

Features	P	R	F1	Acc
GloVe	0.8233	0.8232	0.8229	0.8234
GloVe+Sim_{path}	0.8246	0.8246	0.8233	0.8237
GloVe+Discon.	0.8262	0.8264	0.8258	0.8263
GloVe+Repeti.	0.8239	0.8241	0.8237	0.8240
GloVe+U*	0.8355	0.8359	0.8353	0.8359
GloVe+S*	0.8331	0.8326	0.8321	0.8326
GloVe+QE-U	0.8361	0.8363	0.8355	0.8359
GloVe+QE-I	**0.8363**	**0.8365**	**0.8356**	**0.8360**

5.5 Feature Visualization

Figure 2 shows the distribution histograms of the values of QE-Uncertainty and QE-Incongruity for the joke and non-joke samples. From the figure, it can be found that jokes have higher QE-Uncertainty and QE-Incongruity values than non-jokes, which is consistent with what we stated in Section 3.

6 Conclusion

In this paper, we model semantic uncertainty with a quantum framework. Inspired by the incongruity theory, we design two features, QE-Uncertainty and QE-Incongruity. We conduct experiments on the humor dataset, and the experimental results

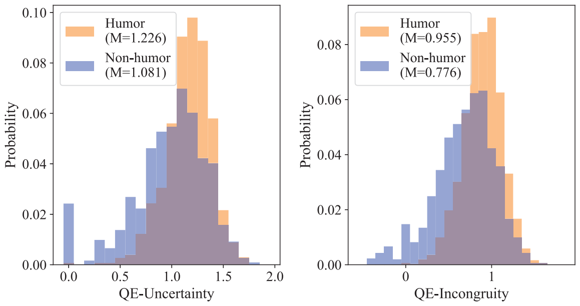

Figure 2: Histograms of our proposed features. The x-axis is the value of the feature, and the y-axis is the proportion of the feature in the total number of samples. **M** is the **M**edian of the current feature.

demonstrate the effectiveness of our proposed features. This suggests that the density matrix is an excellent framework for describing uncertainty and that the quantum entropy of the density matrix is a better feature to distinguish jokes from non-jokes than previously proposed features. We believe that the quantum framework can also be used for semantic uncertainty modeling for other tasks in the future.

Limitations

In this paper, the density matrix representation of text is constructed in an averagely weighted manner, without considering the influence of weights on words. In addition, the density matrix as a text representation does not consider the position information of words. Furthermore, quantum generalization on the problem of multimodal humor recognition is also an interesting topic compared to unimodal humor recognition.

References

Andrew Cattle and Xiaojuan Ma. 2018. Recognizing humour using word associations and humour anchor extraction. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1849–1858.

N. J. Cerf and C. Adami. 1999. Quantum extension of conditional probability. *Phys. Rev. A*, 60:893–897.

Lei Chen and Chong Min Lee. 2017. Convolutional neural network for humor recognition. *arXiv preprint arXiv:1702.02584v1*.

Peng-Yu Chen and Von-Wun Soo. 2018. Humor recognition using deep learning. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 113–117.

Xiaochao Fan, Hongfei Lin, Liang Yang, Yufeng Diao, Chen Shen, Yonghe Chu, and Tongxuan Zhang. 2020. Phonetics and ambiguity comprehension gated attention network for humor recognition. *Complexity*, 2020:1–9.

Anna Hickey-Moody and Timothy Laurie. 2017. Masculinity and ridicule. In *Gender: laughter*, pages 215–228. Macmillan Reference USA.

Lizhen Liu, Donghai Zhang, and Wei Song. 2018. Modeling sentiment association in discourse for humor recognition. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 586–591.

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

Rada Mihalcea, Carlo Strapparava, and Stephen Pulman. 2010. Computational models for incongruity detection in humour. In *Computational Linguistics and Intelligent Text Processing*, pages 364–374.

Alex Morales and Chengxiang Zhai. 2017. Identifying humor in reviews using background text sources. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 492–501.

Mauk Mulder and Antinus Nijholt. 2002. Humour research: State of the art. *CTIT Technical Report Series*, (02-34):1–24.

Michael A Nielsen and Isaac L Chuang. 2010. *Quantum Computation and Quantum Information*. Cambridge University Press.

Anton Nijholt, Oliviero Stock, Alan Dix, and John Morkes. 2003. Humor modeling in the interface. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, page 1050–1051.

EA Pasquali. 1990. Learning to laugh: humor as therapy. *Journal of Psychosocial Nursing and Mental Health Services*, 28(3):31–35.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

Roy Rada, Hafedh Mili, Ellen Bicknell, and Maria Blettner. 1989. Development and application of a metric on semantic nets. *IEEE transactions on systems, man, and cybernetics*, 19(1):17–30.

Victor Raskin. 1979. Semantic mechanisms of humor. In *Annual Meeting of the Berkeley Linguistics Society*, pages 325–335.

A. Schopenhauer. 1966. *The World as Will and Representation*. Dover books on philosophy. Dover Publications.

Claude Elwood Shannon. 1948. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.

John Von Neumann. 2018. *Mathematical foundations of quantum mechanics: New edition*, volume 53. Princeton university press.

Orion Weller and Kevin Seppi. 2019. Humor detection: A transformer gets the last laugh. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3621–3625.

Yubo Xie, Junze Li, and Pearl Pu. 2021. Uncertainty and surprisal jointly deliver the punchline: Exploiting incongruity-based features for humor recognition. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 33–39.

Diyi Yang, Alon Lavie, Chris Dyer, and Eduard Hovy. 2015. Humor recognition and humor anchor extraction. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2367–2376.