

テキストの面白さの評価によるユーモアの認識

Humor Recognition by Text Funniness Evaluation

○天谷 祐介¹, 荒木 健治¹, ジェプカ ラファウ¹

○Yusuke Amaya¹, Kenji Araki¹, Rafal Rzepka¹

¹北海道大学大学院情報科学研究科

¹Graduate School of Information Science and Technology, Hokkaido University

Abstract: In recent years, dialogue systems are spreading in society. These systems can talk about any topic. However, they cannot recognize humor, so we lose the motivation to talk to them. Therefore, in order to recognize funniness of a text, we use statistical methods. As a result, our method classifies correctly if a text is humorous or not. We conclude that a computer is able to recognize humorous text or not. In future, we plan to design a system that measures degree of funniness.

1. はじめに

ユーモアは相互コミュニケーションを良好にするために必要不可欠な言語現象の1つである。人は目的を達成するために行う対話に加え、特定の目的を持たずに単に楽しむことを意図した対話、いわゆる雑談を行って生活している。いずれの場合も話の中にユーモアを織り込むことで敵対心や緊張感を緩和することができる。それによって人間関係の距離を縮めることでコミュニケーションを円滑にする効果がある。

近年、コンピュータでも人間との会話ができる対話システムが社会に広まりつつある。具体的な例として、通信回線を用いたサービスの1つであるチャットにおける会話ボットや、発話解析・認識インターフェース Siri^{*1}がある。これらのシステムはユーザ側の入力文に応じて返答を行う。しかし、これらのシステムは未知のユーモアが入力された時に、入力文がユーモアであるかどうかを判断することができない。そのため、ユーザが面白いと思って発したユーモアに対しても、システムがユーザの期待に沿わない反応を返し、ユーザがシステムと対話を続けようとする意欲が損なわれるという問題がある。解決策として、まず入力テキストがユーモアかどうかをシステムが判断できるようにし、判断した結果に応じた反応を返せるようにする必要がある。つまり、システムにユーモアを理解できるようにする必要がある。

コンピュータのユーモアの生成に関する研究は従

来より行われてきた。日本語ユーモアの代表である「駄洒落」では、なぞなぞを利用した生成システム[1]や、謎かけというフレームワークを利用した生成システム[2]がある。それ以外に、最近ではことわざに対してすかしというフレームワークを用いて生成するユーモアもある[3]。一方、ユーモアの理解の分野に関して、日本では一部の例外[4]を除いて研究が進んでないのが現状である。

そこで本稿では、統計的手法を組み合わせることで、ユーザが入力した文を自動的に判断してユーモアかノンユーモアかの分類を行うシステムを提案し、実験およびシステムの評価を行った結果について述べる。

2. ユーモア認識システム

2.1 対象とするユーモア

対象とするユーモアは、話者自身や関係者の体験談や近況、あるいは知識に基づくストーリー型のユーモアであり、これを「物語ユーモア」として定義する。物語ユーモアの例を例1に示す[5]。

例1: 息子がまだ3歳のころ、「お友達の家で、おかしをもらったらすぐにお母さんに言わないとだめよ」とやかましく教えていた。ある日、友達の家で遊んでいた息子を迎えに行ったら、玄関先まで見送りに出てくれたお友達のお母さんの前で、「お母さん、今日はおかしもらわなかったよ」

*1 Siri: <http://www.apple.com/jp/iphone/features/siri.html>

例1では、話者である母親が、自分の息子に対してお菓子をもらったら自分に言うようにと教えている。これは話者が息子の友人の親に対して礼をするために聞いているのだと推測できる。話者の話を受けた息子は、迎えに行った玄関先で話者に報告をするが、その日はお菓子をもらわなかったため、「もらわなかった」と報告をする。その場には友人の母親もいたため、その発言は常識的には失礼に当たるが、息子にとってみれば母親である話者の言ったことを守ったに過ぎない。受け手側は、この話の状況のなかで、話者と息子の友人の母親との間に流れる気まぐれ雰囲気や、話者の焦りや恥ずかしさを想像して、そこにおかしみを感じると考えられる。このように、物語ユーモアにはストーリーの背景や「オチ」があり、これらの特徴は日常的に我々が話すユーモアの形式に近いといえる。

2.2 提案手法

物語ユーモアを認識する手法として解析的なアプローチを用いることが考えられるが、物語ユーモアは一定の形式を持たないユーモアであり、話の背景の推測や、常識の理解などを含めた膨大な非言語情報の理解を必要とする。そのため解析的手法単体での処理は現段階では困難である。したがって、本稿では解析的なアプローチでの処理を行う前段階として複数の統計的手法を組み合わせる手法を提案する。提案手法では教師あり学習を行うため、システムの処理の流れに、分類部の他に独立して学習部が存在する。それぞれ学習部の処理の流れを図1に、分類部の処理の流れを図2に示す。

提案手法では統計的手法として Naïve Bayes[6]と SVM[7]を用いる。これらの詳細な説明は 2.4, 2.5 で述べる。

2.3 学習部

学習部では、物語ユーモア（ユーモア）と一般の物語（ノンユーモア）テキストを訓練データとして分類器の学習を行う。テキストを分類器が学習できる形にするため、形態素解析を行い、各形態素のト

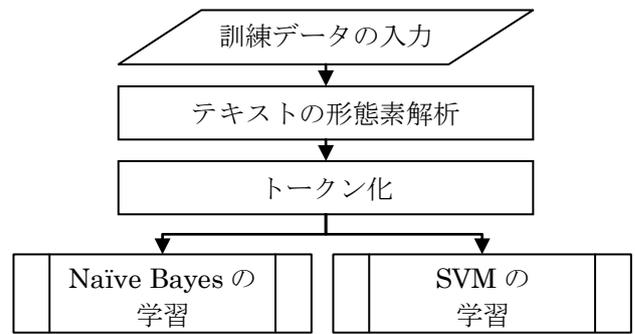


図1 システム学習部の流れ

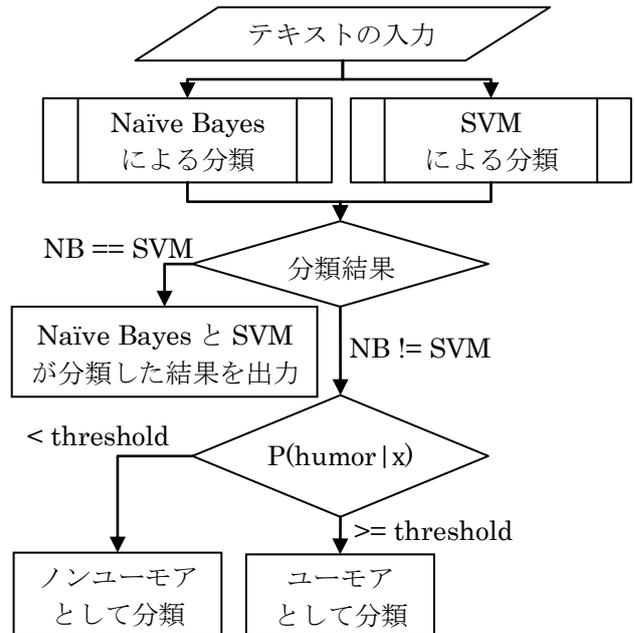


図2 システム分類部の流れ

ークン化を行う。本システムでは形態素の解析に MeCab[8]を用いた。

2.4 Naïve Bayes

Naïve Bayes は、確率的分類器である。訓練データの個数が少なくとも分類に必要なパラメータを見積もることが可能であり、処理が高速でありながら高精度である。

Naïve Bayes は対象となるテキストの、複数のクラスにおける事後確率を比較し、値が最大のクラスへとテキストを分類する。式で表すと、 c_k をクラス、 x をテキスト、 $P(C|D)$ をクラス C におけるテキスト D の事後確率として、以下ようになる。

$$\text{classify}_{NB} = \text{argmax}_{c_k} P(c_k|x) \quad (1)$$

なお、事後確率は以下の式(2)で示される。

$$P(c_k|x) = \frac{P(x|c_k)P(c_k)}{P(x)} \quad (2)$$

$P(c_k)$ は、 c_k における事前確率であり、訓練データに含まれるクラス c_k のテキストの割合で表される。

$P(x|c_k)$ は尤度であり、クラス c_k の時にテキスト x が生成される確率を表している。Naïve Bayes の特徴として単語の出現確率は、テキスト中の他の単語との共起に影響しないと仮定する。そのため、 x_i がテキスト x を構成する形態素とすると、

$$\begin{aligned} P(x|c_k) &= P(x_1 \wedge \dots \wedge x_K | c_k) \\ &= \prod_{i=1}^K P(x_i | c_k) \end{aligned} \quad (3)$$

となり、式(4)に示すように正規化係数である $P(x)$ で正規化を行う。

$$P(x) = \sum_{j=1}^M \left\{ P(c_j) \prod_{i=1}^K P(x_i | c_j) \right\} \quad (4)$$

2.5 SVM

Support Vector Machine(SVM)は、1995 年に Vapnik らにより提案された、教師あり学習を用いた分類手法である。現在知られている手法の中で最も精度が高い学習モデルの1つである。他にも汎化性能が高いといった特徴がある。

SVM は、線形しきい素子を用いることで 2 クラスのパターン識別器を構成する。入力特徴ベクトルに対し、識別関数を用いることで 2 値の出力値を計算し、2 クラスのいずれかに分類する。識別関数の定義式を式(5)に示す。

$$\text{classify}_{SVM} = \text{sign}(\mathbf{w}^T \mathbf{x} - h) \quad (5)$$

\mathbf{w} はシナプス荷重に対応するパラメータ、 h はしきい値である。

さらに、複雑な識別課題に対して性能を向上させ

るため、特徴ベクトルを非線形変換し、識別を行うカーネルトリックと呼ばれる手法を用いる。この手法を用いると、式(5)は、

$$\text{classify}_{SVM} = \text{sign} \left(\sum_{i=1}^N \lambda_i y_i K(\mathbf{x}_i, \mathbf{x}_j) - h \right) \quad (6)$$

と表せ、テキストの特徴ベクトル \mathbf{x}_i に対して、正解ラベルとして y_i が与えられている。さらに制約条件として、

$$\sum_{i=1}^N \lambda_i y_i = 0 \quad (7)$$

$$0 \leq \lambda_i \leq C, \quad i = 1, \dots, N \quad (8)$$

式(7)および式(8)を満たす必要がある。カーネル $K(\mathbf{x}_1, \mathbf{x}_2)$ には多項式カーネルを代入する。多項式カーネルを式(9)に示す。

$$K(\mathbf{x}_1, \mathbf{x}_2) = (1 + \mathbf{x}_1^T \mathbf{x}_2)^p \quad (9)$$

2.6 分類部

分類部における処理の流れを述べる。クラスが未知のテキストを入力すると、Naïve Bayes と SVM がそれぞれ学習内容に基づいてテキスト分類を行う。次に Naïve Bayes と SVM の分類結果を比較し、結果が等しければ分類結果が正しいと判断して出力する。

ここで、それぞれの分類結果が異なっていた場合、Naïve Bayes の結果かあるいは SVM の結果のいずれかを採用することが考えられる。本稿では分類の指標となる値が含まれている Naïve Bayes の結果を採用する。両分類器で分類結果が異なっていたテキストに対して、Naïve Bayes の分類の際に求めたユーモアの場合の事後確率を参照し、しきい値以上であればテキストをユーモアとして分類し、しきい値より小さければノンユーモアとして分類する。

3. 性能評価実験

提案手法がユーモアとノンユーモアを正しく分類

表 1 アンケート調査を行ったテキストの例

クラス(「そう感じた」の回答者数)	テキストの内容
物語ユーモア (11 / 11 人)	こないだガソリンスタンドのに一ちゃんが「不凍液お入れになりませんか」と言うので「いくら？」と聞いたら「メンバー様価格で2千円です」とのこと「じゃ、定価はいくら？」とのマニュアルにない質問に慌てたそのに一ちゃん「定価は・・・2千円です」と正直に答えた。
物語ユーモア (6 / 11 人)	「なにが食べたいの？言っでごらんさいよ。何が食べたいの？」聞くだけ聞いていつものオカズ。
一般の物語 (5 / 11 人)	息子が高校生のとき、参観日に「学校でおれに会っても声かけるなよ」と言われた。現在、茶髪で長髪、フリーターの彼には「街で会っても声かけないでよ」と言い渡してある。
一般の物語 (0 / 11 人)	会場の前に居ても誰も私に気付かない。成人式に出ないと言うと母に泣かれた。あっさり折れた。普段着で行くつもりが振り袖を用意され美容室まで予約されていた。なのにこのざまだ。泣きたかった。会場の中には入らずマックで時間を潰した。「ただいま。みんなに会えて楽しかったよ」

できることを確認するために性能評価実験を行った。

3.1 実験データ

実験に用いるテキストは、Web 上および Twitter[9]から収集した。Web 上で「笑い話」を掲載しているサイトから笑い話を 987 文、Twitter からはハッシュタグとして「#twnovel」が付与されたツイート 1211 文を収集し、各 400 文ずつ無作為に抽出してそれらに対してアンケート調査を行った。

アンケートは、北海道の大学に在籍している 20 代前半の大学生 11 名に対して行なった。内訳として男性 6 名、女性 5 名、文系 5 名、理系 6 名であった。アンケートは各文を読んでもらい、「文章の作者が、書いた文章によって読者を笑わそうとしているかどうか」を尋ね、回答は「そう感じた」か「そう感じなかった」のいずれかで答えてもらった。

集計したアンケートの結果から過半数「そう感じた」と回答があったテキストを物語ユーモアとして分類し、それ以外のテキストを一般の物語として分類すると、物語ユーモアが 438 文、一般の物語が 362 文となった。集計結果により分類されたテキストの例を表 1 に示す。

実験では、集計結果によって分類されたクラス毎にテキストを 362 文ずつ無作為に抽出してデータセットとした。

3.2 実験方法

実験は 3 つに分けて行う。実験 1 では、Naïve Bayes と SVM の両分類器で分類結果が一致したテキスト集合の中での分類性能を評価する。

実験 2 では、実験 1 で両分類器での分類結果が一致しなかったテキストを用いて、ユーモアの場合の事後確率のしきい値を決定する。事後確率は最大値が 1 になるように正規化されており、実験ではしきい値を 0.001, 0.1, 0.3, 0.7, 0.9, 1 と変化させてそれぞれの場合での分類性能を評価する。

実験 3 は、実験 2 で分類精度が最も良かったしきい値を用いて、提案手法全体での分類性能を求め、ベースライン手法と比較した。

いずれの実験でも、検定には k-分割交差検定を用いた。なお、k=10 とした。

3.3 実験結果

それぞれ表 2, 表 3, 表 4 に実験 1, 実験 2, 実験 3 の結果を示す。

実験 1 では、両分類器で結果が一致したテキストは全 724 文中 562 文であったため、562 文の中での分類性能を示している。ベースラインとして、Naïve Bayes と SVM をそれぞれ単体で用いて全 724 文を分類した場合の分類性能と比較を行った。その結果、両分類器で一致した分類結果を用いることによって、ベースライン手法である Naïve Bayes と比較して分類精度に 7.5 ポイント、再現率に 12.0 ポイントの向上が見られた。同様にベースライン手法である SVM

表 2 2つの分類器を用いた手法とベースライン手法の性能比較

分類手法 (対象となる文数)	分類精度(%)	再現率(%)	F 値(%)
Naïve Bayes (724 文)	79.0	78.3	78.6
SVM (724 文)	78.1	83.5	80.7
両分類器の結果が一致 (562 文)	86.5	90.3	88.3

表 3 ユーモアの事後確率のしきい値ごとの性能

しきい値	分類精度(%)	再現率(%)	F 値(%)
0.001	54.7	64.1	59.0
0.1	50.9	45.4	48.0
0.3	51.2	41.9	46.1
0.7	49.9	36.4	42.1
0.9	54.2	36.4	43.6
1	49.8	23.3	31.7

表 4 提案手法とベースライン手法の性能比較

分類手法	分類精度(%)	再現率(%)	F 値(%)
Naïve Bayes	79.0	78.3	78.6
SVM	78.1	83.5	80.7
提案手法	79.4	84.4	81.8

と比較すると、分類精度に 8.4 ポイント、再現率に 6.8 ポイントの向上が見られた。さらに性能を上げる方法として新たに他の分類器を追加するという方法が考えられる。しかし、実際に対話システムに組み込むことを考えると、分類器が増えると処理に時間がかかるという問題が生じる。

実験 2 では、実験 1 の時に両分類器で結果が一致しなかったテキスト 162 文を対象に実験を行った。結果、しきい値を 0.001 に設定した場合、分類精度が 54.7%、再現率が 64.1%と最も高かった。

実験 3 では、実験 1 の結果および実験 2 で最も性能が良かったしきい値 0.001 を採用して提案手法全体の性能を求めた。その結果、ベースライン手法である Naïve Bayes と比較して分類精度が 0.4 ポイント、再現率が 6.1 ポイント向上した。同様にベースライン手法である SVM と比較して分類精度が 1.3 ポイント、再現率が 0.9 ポイント向上した。

しかし、実験 2 の結果を見ると、最も性能が良かったしきい値 0.001 の時の分類精度が 54.7%、再現率が 64.1%と、いずれの場合も性能が良いとは言い

がたい。両分類器の結果が不一致だった場合の分岐手法の性能によって、実験 3 で評価した提案手法全体での性能は、実験 1 の結果と比較すると分類精度が 7.1 ポイント、再現率が 5.9 ポイント低下している。性能が良くない理由として、ユーモアの場合の事後確率は多くの場合において 0 か 1 であり、しきい値を設けても結果が変わらないためと考えられる。そのため、今後は両分類器間で結果が不一致だったテキストの分類を異なる方法で行う必要があると考える。

4. まとめと今後の課題

本稿では、物語ユーモアと一般の物語を Naïve Bayes と SVM のハイブリッド手法を用いて分類を行い、Naïve Bayes 単体と比較して分類精度に 0.4 ポイント、再現率に 6.1 ポイントの向上が見られ、SVM 単体と比較して分類精度に 1.3 ポイント、再現率に 0.9 ポイントの向上が見られた。

本稿では、両分類器の結果が一致しなかったテキストに対してさらにユーモアの場合の事後確率を用

いて分類を行った。しかし、実際の対話において会話の内容がユーモアである場合の方が少ないと考えられるので、対話システムに組み込む際は、両分類器間で結果が不一致だった場合にノンユーモアとして分類することでさらなる精度の向上が見られるとも考えられる。あるいは、分岐の方法に意味的な解析手法を取り入れることも考えられる。内容を考慮した分類を行えることで、性能の向上に加えて面白さの度合いを評価することも可能であると考えられる。

今後は分類性能の向上とともに、面白さの度合いを評価するシステムを構築する予定である。そのためにも、今後はアンケートの段階でより詳細に意見を抽出する必要があると考える。今回のアンケートでは、回答者には話者にユーモアを話す意図があると感じられたかどうかを尋ねたが、今後はアンケートの段階でテキストを書いた時の作者の意図と、テキストの面白さを尋ねる枠をそれぞれ別に設け、「話者はユーモアである」と意図したが、面白くない」、あるいは「話者は意図して話してないが、受け手側は面白いと感じた」といった実際の対話で起こりうる状態を考慮したデータセットを構築してそれを分類することが、人間らしさを持ち、ユーザが対話を続けたいと思うコンピュータを作るために必須であると考えられる。

参考文献

- [1] キム ビンステッド, 滝澤 修: 日本語駄洒落なぞなぞ生成システム”BOKE”, 人工知能学会誌, Vol. 13, No. 6, pp. 920-927, 1998
- [2] Jonas Sjöbergh, Kenji Araki: Automatically Creating Word-Play Jokes in Japanese, 情報処理学会研究報告. 自然言語処理研究会報告 2007(35), pp. 91-95, 2007
- [3] 山根宏彰, 萩原将文: 笑いを生むことわざすかしの自動生成システム, 知能と情報: 日本知能情報ファジィ学会誌, Vol. 24, No. 2, pp. 671-679, 2012
- [4] 田辺公一朗: 駄洒落のコンピュータによる処理——駄洒落生成システムの基本設計——, 産能

大学紀要 26(1), pp. 65-74, 2005

- [5] 笑い話 いろ色,
<http://laugh100.com/index.html>
- [6] M. E. Maron: Automatic Indexing: An Experimental Inquiry, Journal of the ACM (JACM) 8(3), pp. 404-417, 1961
- [7] Corinna Cortes, Vladimir Vapnik: Support-Vector Networks, Machine Learning, 20(3), pp. 273-297, 1995
- [8] MeCab: Yet Another Part-of-Speech and Morphological Analyzer,
<http://mecab.sourceforge.net/>
- [9] Twitter, <https://twitter.com>

連絡先

〒060-0814

札幌市北区北 14 条西 9 丁目

北海道大学大学院情報科学研究科

言語メディア学研究室

天谷 祐介

TEL: 011-706-7389

E-mail: hv.amaya@media.eng.hokudai.ac.jp