

キャラクターエージェント感情表出のための機械学習による テキスト文からの印象抽出

木田信雄[†], 倉田岳人[‡], Santi Saeyor[‡], 石塚満[‡]

[†] 東京大学 工学部電子工学科

[‡] 東京大学大学院 情報理工学系研究科 電子情報学専攻

1 はじめに

従来の対話処理研究の多くは、旅行の申込みのような手続きの業務の対話や、知的教授システムにおける教示的な対話のようにタスク思考的な要素の強い対話を研究の対象とし、目的を効率よく実現する対話の研究が目標であった。しかし、ユーザとの親和性が求められる対話システムにおいては、ユーザの感情を推定し、システムの感情を表現するといった感情情報処理が重要な役割を担うことがわかっている [1], [2]。現在、感情情報処理研究においては感情タグ付きコーパスを作成し、感情を推定する手法 [3][4] が提案されているが、感情情報処理研究自体はまだ萌芽の段階にあり [5] 感情の推定手法もいまだ明らかにされていない。また、感情に関する辞書やコーパスといった研究に必要な資源も、試験的なものがいくつか見られるに過ぎず、感情タグ付きコーパス作成は計算機による自動化が難しい。本研究では、テキスト文から聞き手が受けるであろう印象を機械学習を用いて推定する手法を提案する。携帯電話上のキャラクターエージェントとユーザとの親和性を高めるために、エージェントの感情表出の指標としてこの手法を用い、エージェントに聞き手の印象に合わせた、感情を連想させる動作を含んだ情報伝達システムを構築する研究に貢献する事が本研究の目的である。本研究をさらに発展させた印象推定の際に個人の嗜好を考慮した研究等に貢献できればよいと考えている。本手法を提案する上で、感情タグ付きコーパスについては [3], [4] を、機械学習をテキストに利用する手法については [6] を参考とした。

Extraction of impressions from text based sentences using machine learning to determine character agent's feelings

[†]Nobuo KIDA

Dept. of Electronics Engineering,
School of Engineering, University of Tokyo

[‡]Gakuto KURATA, Santi SAEYOR, Mitsuru ISHIZUKA
Dept. of Information and Communication Engineering,
Graduate School of Information Science and Technology,
University of Tokyo

2 手法

本手法ではテキスト文に含まれる形態素（形態素解析には chasen¹ を使用）の文ベクトルパターンを学習器を用いて学習する。印象を判別する判別器を作成することが本研究のメインテーマである。素性としては人間がテキスト文の印象を判断する際に注目していると考えられる名詞、動詞、形容詞、副詞などの意味があると考えられる形態素に着目した。学習、評価に用いる全テキスト内の該当する形態素を学習器の素性とする。本手法では素性の数が非常に多くなり、各テキストの文ベクトルパターンが非常にスパースなものになることが予想されるため、文ベクトルパターンがスパースであっても意味のある学習が可能であるということ、また逆にスパースであっても高い識別能力が期待できる SVM (support vector machine)² を用いる。

2.1 実験

評価実験に用いるテキスト文として、印象判別をする際に適していると考えられるニュース記事（スポーツ、社会）を選び用いることとした。その為本実験ではまず印象判別の前にドメイン分類を同様の手法で機械学習を用いて行い、記事を社会、スポーツの2ドメインに分けた後に各ドメイン別の素性を抽出し、学習を行った後に印象推定を行うこととした。学習時の特徴量（feature vector）としては、テキスト文中にその形態素が現れるか現れないかの2値（0, 1）を採用し、結果は全ニュース記事を2つに分け、クロスバリデーションにより評価を行った。手動でニュース記事を分類する基準として様々なものが考えられるが、本実験では客観的に平均的な日本人が受けるであ

¹Chasen Homepage: <http://chasen.aist-nara.ac.jp>

²パターン認識の分野において最も優秀な学習モデルの1つであることが知られている。1960年代に Vapnik 等が考案した Optimal Separating Hyperplane を起源とする。SVM は与えられた学習データをすべて正しく識別できるようにするノンパラメトリックなパターン分類アルゴリズムであり、識別境界の位置を決定する明確な基準を持つ。パターン認識性能の優秀な学習モデルの一つである学習機械として知られ、分類学習として SVM を用いた手法は様々な分野で応用されている。その数学的な理論などは参考文献 [7],[8] に譲る。

ろう印象のモデルとして筆者自身の印象基準を採用した。³印象の手動分類基準を変更することにより、様々な嗜好に対応した印象推定器を作成することが可能である。またSVMは2値分類の学習器であるが、本手法の印象推定では3値分類問題となるので2つの判別モデル(良い印象判別モデル、悪い印象判別モデル)を作成し、どちらとも負例と判別された場合に特に印象を与えないものと判別されたとみなした。ドメイン分類は2値分類問題である。

2.2 手法の流れの概要

- 1 Web上のニュース記事⁴を集め、良い印象を与えるもの、悪い印象を与えるもの、特に印象を与えないものの3値に手動で分ける
- 2 全テキストから、該当する素性を抽出する。(ドメイン分類用、各ドメインでの印象推定用)
- 3 SVMを用い、学習データから文ベクトルパターンを学習する。
- 4 SVMを用い学習パターンからスポーツ、社会の記事をわけるドメイン分類を行い、テキストの印象値を自動推定する。

3 実験結果、考察および今後の課題

ドメイン識別の結果(用いた記事数 316+317=633)および社会ニュース記事での悪い印象判定モデルの結果(用いた記事 2週間分 217+203=420)を示す。

表 1: ドメイン識別結果

クロスバリデーション	Accuracy	Precision	Recall
1	92%	96%	82%
2	98%	95%	99%
average	95%	95.5%	90.5%

機械学習を用いたドメイン分類では高い精度の結果を得ることができた。社会ニュース記事での悪い印象判定モデルの結果も学習による精度の上昇が見られた。

c.f クロスバリデーションについて⁵

³サッカー日本代表が勝った うれしい
殺人事件がおきた かなしい 等

⁴ニュース記事は asahi.com(<http://www.asahi.com/>) から集めた。

⁵1は全記事の前半分を学習データ、後ろ半分を評価用データに用いる。2はその逆

表 2: 社会のニュース記事での悪い印象判定モデルの結果

クロスバリデーション	Accuracy	Precision	Recall
1	71%	95%	65%
2	66%	60%	98%
average	68.5%	77.5%	81.5%

4 考察、今後の展望

ドメイン分類、社会ニュース記事での印象判定モデルではよい結果が得られたが、その他の印象判定モデルでは低い数値しかえられず、学習による効果はあまり見られなかった。これはニュース記事の時期による話題の偏り、及び社会、スポーツ各ドメインにおける記事の偏り(スポーツではよい印象の記事が多く(が優勝 など)社会では悪い印象の記事が多い(殺人事件がおきた など))があるためだと考えられる。記事の偏りを防ぐために学習データとしてできるだけ広い領域をカバーすることが精度上昇の為に必要であると考えられる。また印象判別学習の精度向上に関する今後の課題としては、手動でニュース記事を判別する際に参考にした動詞、目的語の係り受け関係、形容詞とその係り受け関係などに注目していく必要がある。

参考文献

- [1] E. Elliott, J. Rickel, and J. Lester. Lifelike pedagogical agents and affective computing. *An Exploratory Synthesis Artificial Intelligence Today*.
- [2] 感情表現を用いた対話システム EDS の開発 (1) ~システムの概要と感情モデル~, 情報処理学会研究報告, No. 89, 2000.
- [3] 徳久良子, 乾健太郎, 徳久雅人, 岡田直之. 言語コーパスにおける感情生起要因と感情クラスの注釈づけ.
- [4] 徳久良子, 乾健太郎, 徳久雅人, 岡田直之. 規模とコストを考慮した感情タグつき言語コーパスの作成方法.
- [5] 乾健太郎, 徳久雅人, 徳久良子, 岡田直之. 感情生起とその反応. 日本ファジィ学会誌, Vol. 12, No. 6, 2000.
- [6] 藤村滋, 松村真宏, 石塚満. 機械学習による電子掲示板からの評判情報抽出. 情報処理学会第 65 回全国大会 5-BZ-6.
- [7] 赤穂昭太郎, 津田宏治. サポートベクターマシン基本的仕組みと最近の発展. 数理科学, 2000.
- [8] 小野田崇. Large margin classifiers. 人工知能学会誌 Vol.17 No.1, 2002.