感情推定システム用の二視点評価方法
Double Standpoint Evaluation Method for Affect Analysis Systems

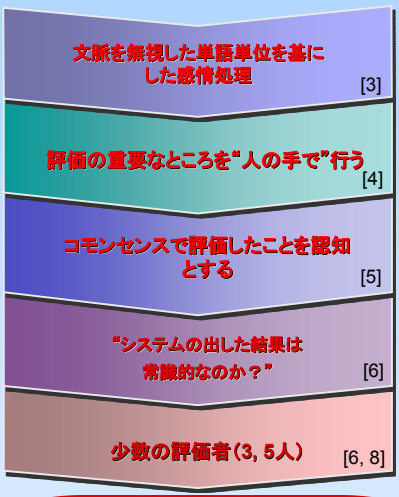
Michal PTASZYNSKI Pawel DYBALA Rafal RZEPKA Kenji ARAKI
Graduate School of Information Science and Technology, Hokkaido University

JSAI 2008 第22回人工知能学会 全国大会

概要

感情推定システムのための二視点に基づいた評価方法(Double Standpoint Evaluation Method)を紹介する．第一視点ではシステムの感情認知が評価され，第二視点ではシステムの感情推定がどの程度一般コモンセンスと一致するかが評価される．以前の研究で紹介したML-Askシステムをこの方法をもって評価する．さらに本評価方法を一般に使われる評価方法と実験的に比較した結果，後者の不十分な点が明らかになり，本方法はより客観的であることが証明された．

評価における問題点



評価…実験？

土屋らの評価手法を基に実験でML-Askシステムを評価
土屋ら感情判断の評価方法:

- システムによる感情判断
- 5人被験者(評価者)
- 可能な評価:常識的、非常識ではない、非常識
- 結果の区分:「常識的」と評価した:
  - ～ ⇒常識的
  - ～ ⇒非常識ではない
  - ⇒非常識

土屋らの結果のまとめ:「正答率は(中略)88.0%と非常に高い結果であり，本システムは有効であると言える.」

土屋の評価方法を用いた我々の結果のまとめ:

土屋式に評価を行った結果、ML-Askシステムは97%の正確さが得られた。それはシステムがほぼ完璧でほとんどの既存システムより性能がすぐれているということになる…？

甘い評価方法では結果を高く見せられる!
このような評価方法は極めて非現実的!!!

ML-Askシステム

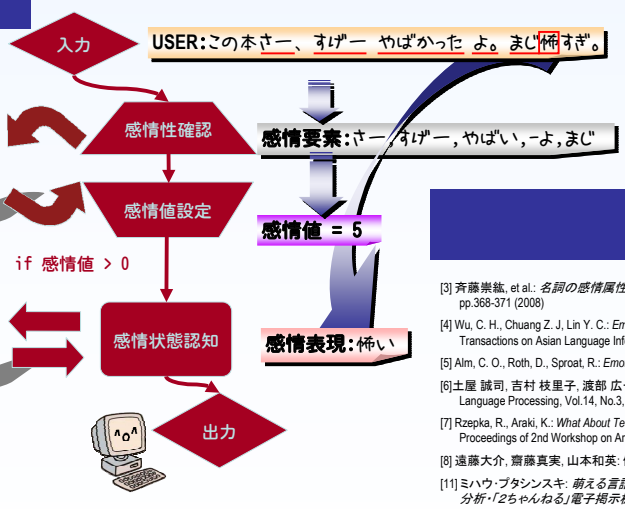
以前の研究[11]から:

- **感情層(日本語)は人間間コミュニケーションを円滑にする**
- **その要素を全般に2種類に分類:**

1.感情要素. 具体的な感情を示さずに発話の感情的コンテクスト設定することで聞き手に感情が表されたことを知らせる. 例, わくわく, すげぇ, !, ???等.

2.感情表現. 常に必ず感情的コンテクストで使われるわけではないが, 感情的コンテクストで使われた場合, 話者の感情状態を表す表現. 例, 喜ぶ, 興奮.

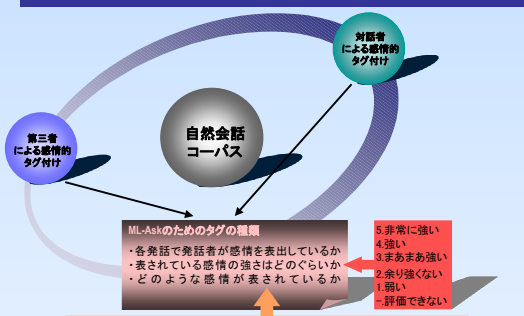
両方に当てはまるものも多少ある
(わくわく[興奮], ～やがる[嫌悪・怒り]等).



二視点評価方法(DSEM)

感情推定システム用の客観的な評価方法の特性

DSEM 用タグ付き対話コーパス



感情の定義と分類については:中村[12]より
定義:気もちや心理などということばでさす、ある臨時的な精神状態のほとんど全て・
分類:喜, 怒, 哀, 怖, 恥, 好, 厭, 昂, 安, 驚.

ML-Ask のDSEM 評価

認知視点評価

感情性推定・感情的コンテクスト決定
感情性有無推定力をそれぞれF値で計算;両方のF値の平均値⇒結果
結果: F値=0.81
人間の認知力は0.4～0.86⇒
システムの感情的コンテクスト推定力は人間レベルにある

感情値設定
システムの判断と発話者の判断との一致率U
感情値設定には曖昧性があるため⇒ APM(ほぼパーフェクトマッチ)条件を許可
APM:発話毎に感情値判断に±1ポイント差異
結果:一致率U=0.5
人間の感情値設定力は~0.84⇒
システムの感情値設定力は人間レベルの60%

感情状態・感情種類推定
感情種類推定力をそれぞれF値で計算;両方のF値の平均値
結果: F値=0.46
人間の認知力は0.71～0.73⇒
システムの感情種類推定力は人間レベルの65%

コモンセンス視点評価

感情性推定・感情的コンテクスト決定
システムの判断と評価者の判断との一致率U;評価者間一致率 ＝ 一般コモンセンス
結果:一致率U=0.58 (条件:APM)
一般コモンセンスの範囲は0.7～0.82⇒
システムの感情的コンテクスト推定の仕方は一般コモンセンスのレベルの83%

感情値設定
システムの判断と発話者の判断との一致率U;
結果:一致率U=0.51 (条件:APM)
一般コモンセンスの範囲は0.61～0.75⇒
システムの感情値設定力の仕方は一般コモンセンスのレベルの84%

感情状態・種類推定
二つの条件で感情状態推定の仕方のコモンセンスのレベルを決定
1.評価者全員の付けた少なくとも一つの感情を抽出
2.評価者大部分の判断と一致
システムの感情種類推定の仕方は一般コモンセンスのレベルの45%

結論

- **感情推定システムの評価方法にいくつかの問題発見**
- **それらを解決する可能な評価方法を提案**
- **本方法→①認知視点及び②コモンセンス視点という2つの視点に基づいた評価方法**
- **ML-Askシステムを本方法で評価**
- **見込みのある結果が得られたが、完璧ではないところも分かり、改善点も決定できた**
- **実験として、ML-Askシステムを一般に使用される評価方法で評価**
- **DSEMに比較して一般の方法の不十分点が明快になった**
- **本方法は高性能でより客観的であることを証明**
- **本評価方法が広く使用されるようになることが期待される**

参考文献

[3] 斉藤崇紘, et al.: *名詞の感情属性の抽出とそれに基づく名詞間類似度の計算*. 自然言語処理学会Proceedings of NLP pp.368-371 (2008)

[4] Wu, C. H., Chuang Z. J, Lin Y. C.: *Emotion Recognition from Text Using Semantic Labels and Separable Mixture Models*, ACM Transactions on Asian Language Information Processing, 2006.

[5] Alm, C. O., Roth, D., Sproat, R.: *Emotions from text: machine learning for text based emotion prediction*, HLT/EMNLP, Vancouver, 2005.

[6]土屋 誠司, 吉村 枝里子, 渡部 広一, 河岡 司: *連想メカニズムを用いた話者の感情判断手法の提案*, Journal of Natural Language Processing, Vol.14, No.3, 2007.

[7] Rzepka, R., Araki, K.: *What About Tests In Smart Environments? On Possible Problems With Common Sense In Ambient Intelligence*, Proceedings of 2nd Workshop on Artificial Intelligence Techniques for Ambient Intelligence, IJCAI'07, 2007.

[8] 遠藤大介, 齋藤真実, 山本和英: 係り受け関係を利用した感情生起表現の抽出. 自然言語処理学会Proceedings of NLP 2006.

[11] ミハウ・プタシンスキ: *萌える言語. インターネット掲示板の上の日本語会話における感情表現の構造と記号論的機能の分析・「2ちゃんねる」電子掲示板を例として*・, アダム・ミツキエヴィッチ大学, 2006

[12] 中村明: *感情表現辞典*. 東京堂出版, 2004