

COLING2014 参加報告 (その2)

– 評判分析研究に関して –

小早川 健[†]

1 はじめに

本稿では、COLING 2014 での発表のうち、筆者が重点的に聴講した評判分析に関するものを紹介する。綱川隆司氏の報告にもあるように、この会議には優れた論文を表彰する制度があり、2本が Honorable Mention (選外佳作賞)、2本が Best Papers (最優秀論文賞) に選ばれた。一般の口頭発表は、4つのセッションが並行して行われていたが、Best Papers の発表だけは、1つのセッションしか開催されないように特別な配慮がなされていた。実際に、会議の終了となる Closing セレモニーの直前ということもあって、大勢の聴講者が集まった。

ここで取り上げる発表は、Best Papers のうち1本と、口頭発表から2本、ポスターから1本、それぞれの表題は次の通りである。現在の世相を反映して、ツイート¹ を扱ったものが多い。

Best Paper A context-based model for Sentiment Analysis in Twitter

Oral Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts

Oral Joint Opinion Relation Detection Using One-Class Deep Neural Network

Poster Political Tendency Identification in Twitter using Sentiment Analysis Techniques

2 A context-based model for Sentiment Analysis in Twitter

この発表 (Vanzo, Croce, and Basili 2014) は、すべての発表の最後を締めくくるにはふさわしいストーリーであった。

ツイートは、テキストの長さが短いために、1ツイート毎の解析が難しい。それ故に、context (文脈) を積極的に利用すべきである。

[†]日本放送協会 放送技術研究所 ヒューマンインターフェース研究部, Japan Broadcasting Corp., Science & Technology Research Labs., Human Interface Div.

¹Twitter Inc. が提供する短文メッセージ投稿サービスに投稿されるメッセージのことをツイートと呼ぶ。リプライとは、@に User ID を付加して、特定の人宛に公開メッセージを送ることができる機能である。また、ツイート中に、#と自由文字列を組み合わせてトピックを表す慣習があり、この表記をハッシュタグと呼ぶ。2014年現在、多くの人には既知のことと思われるが念のため。

文脈の取り扱い方や文脈長の比較に主眼をおいた研究となっており，主要な結論は

- 会話，トピック単位での文脈は効果がある．それぞれに適切な文脈長があり，会話においては6メッセージ，トピックにおいては16メッセージあたりに，提案手法の性能の最良値を持つ．
- ユーザーの過去の発言の文脈は，あまり効果がない．

である．

この研究では，ツイートの一般的な使われ方を鑑みて，次の3種類の文脈が考慮されている．

- User Sentiment Profile Context (ユーザーの文脈)
- Conversational Context (会話の文脈)
- Topic Context (トピックの文脈)

User Sentiment Profile Context とは，着目するユーザーの過去の発言を文脈とするものであり，後述するカーネルに取り込むことによって扱われる．

Conversational Context とは，リプライを文脈とするものである．着目するユーザーの発言のみならず，発言をやりとりした相手の発言が含まれる．

Topic Context とは，同一ハッシュタグの発言を含めるものである．ハッシュタグは，表記のゆらぎなどの不統一や，使われる時期や粒度がまちまちであるために，統一的な扱いが難しいと思われる．

ツイートを3範疇 (ポジ・ネガ・中立) に分類するタスクに対して，サポートベクターマシンによる分類を行っている．ツイートで用いられる記法 (リプライ，ハッシュタグ)，ハイパーリンク，感情絵文字 (emoticons) が抽象化された後に，解析器 (parser) にかけてられ言語特徴量が計算される．

文脈という一連のデータを扱うための工夫として，サポートベクターマシンのマルコフ過程的な定式化である SVM^{hmm} を採用している． SVM^{hmm} とは，系列ラベリング問題において，隠れマルコフモデル (hmm) をマージン最大化で学習する方法 (Altun, Tsochantaridis, Hofmann, et al. 2003) である．

サポートベクターマシンに用いられるカーネルとして，次のものが比較されている．

- Bag of Word Kernel (BoWK)
- Lexical Semantic Kernel (LSK)
- User Stream Profile Kernel (USPK)

Lexical Semantic Kernel とは，Bag of Words 特徴量の示す疎性 (sparseness) を改良した分布特徴量を用いるものである．具体的には，Bag of Words 特徴量に特異値分解 (Singular Value Decomposition) を施して得られる上位 250 次元に圧縮されたものである．

User Stream Profile Kernel とは，ユーザーの文脈を扱うためのカーネルであり，一連の文脈に含まれる評価極性語 (positive, negative, neutral) の数を素性として用いるものである．例え

ば、過去にポジティブな発言が多いユーザーは、発言がポジティブに分類される確率が高くなる。正解ラベルが振られていないような過去の発言に対しては、半教師ありによって作成された学習データが用いられている。

文脈を全く用いない分類と比較して、適切な長さのトピックを用いた分類は、 F_1 値で見ても最大 20% 近く改善している。最大の改善を示すのは、BoWK と LSK カーネルを線形補間で組み合わせたカーネルを用いた場合にトピックコンテキスト長として 6 メッセージを用いた場合である。面白いことに、ユーザーの過去の発言を扱う USPK をカーネルに加えても改善効果は薄い。この研究で扱う、会話やトピックの文脈の長さを変化させた場合の適切値が分かるような興味深い図や、他の条件による網羅的な実験結果が原論文に掲載されているので、ぜひとも参照していただきたい。

質疑では、実験設定に関して、「実験において会話の文脈長が無限大とあるが、評価セットにおいて会話は何回くらいのやり取りが続くのか？」という質問が出され、「中には何回もやり取りが続くものもあるが、実験における最良文脈長である 6 を超える会話のやり取りは、そう多くはない。」といった趣旨の回答があった。

3 Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts

この発表 (dos Santos and Gatti 2014) は、単語単位の特徴では情報量が限定的であることに着目し、文字単位の特徴も利用しようという試みで、文字単位から文単位までをたたみ込んだ深層ニューラルネットワークを提案している。

映画レビューとツイートに対して人手で丁寧に正解が付けられた評価セット (SSTb) を用いた実験では、既存の Recurrent Neural Network (Socher, Perelygin, Wu, Chuang, Manning, Ng, and Potts 2013) を上回る性能が得られている。この評価実験によると、句単位の特徴が大きな改善効果をもたらしていて、それに加えて文字単位の特徴がわずかに改善効果をもたらしている。また、10 万文もの多量のツイートに自動的に正解が付けられた評価セット (STS) を用いた実験では、ポジ・ネガの 2 分類で 86.4% の性能が得られ、現行の最高水準を謳っている。この評価実験によると、文字単位の特徴による改善効果が比較的ある。

文字単位という提案手法独特の改善効果は大きくないものの、ツイート単体で分類する場合の精度向上が追求されているという点で興味深い。Deep Neural Network を学習するための、事前学習 (pre-training) の効果なども原論文で比較されている。

質疑では、

- Q 文字素性の平均出現と畳み込みカーネルとの比較は行ったか？
- Q 文字素性は構成的 (compositional) か？

といった質問が出され、今後の研究のために参考にする旨の回答があった。また、「英語以外の他の言語ではどうか?」という質問もあり、「ポルトガル語に対しても提案手法が有効である旨の予備実験がある」という回答であった。

4 Joint Opinion Relation Detection Using One-Class Deep Neural Network

この発表 (Xu, Liu, and Zhao 2014) は、(意見範疇, 意見対象, 意見間の関係) の3種類の抽出を同時に行うという趣旨である。意見範疇については、正解付けにおいて正例と負例を明確に定めやすいが、意見対象や意見間の関係については、正例は明確にしやすいが、明確な負例を見つけることは意外と難しい。これは、評価対象になりえない語や意見間の関係を持ち得ないものを指示することが難しいからである。このような事情によって負例を上手に集めることが難しいため、正例だけからなる、いわゆる 1-class 問題として扱うことにして、1-class 問題が扱える Deep Neural Network の手法を開発したことが、この発表の新規性である。

1-class 問題は、異常値検出と似たような問題になっており、主に3種類のアプローチ — 1. 正常値付近の分布を用いるもの, 2. 正常値と異常値の境界を定めるもの, 3. 構成要素に分解するもの、が考えられる。この研究で用いられている autoencoder は、3. の構成要素に分解するものであり、正常値を基に構成された autoencoder に対して、どれだけ短いステップで入力を処理できるかが、正常値らしさを表す指標として使えることを利用している。

5 Political Tendency Identification in Twitter using Sentiment Analysis Techniques

この発表 (Pla and Hurtado 2014) は、この参加報告で取り上げる発表の中では唯一のポスター発表である。選挙の際のツイートの政治的傾向がどのようなものを当てるタスクに取り組んでいる。具体的には、2013年のスペインの選挙におけるツイートが、政治的な3範疇(左, 右, 中道)のいずれに分類されるかというタスクからなる評価セット (TASS 2013) があり、Weka という機械学習ツールキットを用いて70%の分類率を達成したというものである。言語特徴量として、表層や品詞のほかに、

- 864 の Political Entities (政党名, 党首名など)
- Polarity Lexicon (評価極性辞書)

を用いている。

この発表を取り上げたのは、評判分析手法の新しさよりも、社会情勢分析などのニーズから昨今着目されるようなタスクに対して、きちんと取り組んでいることが興味を引いたためである。

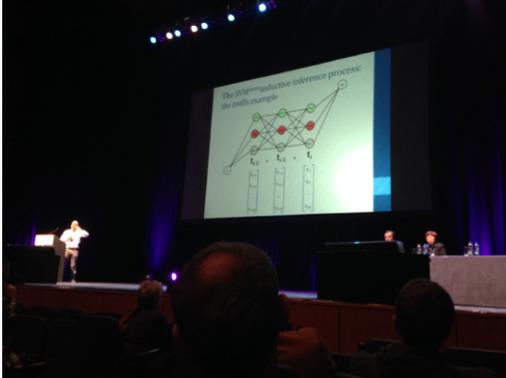


図 1: Best Paper 発表の様子



図 2: ポスター会場の様子

6 おわりに

この会議でも、言語処理における評判分析は、主要なアプリケーションとして相変わらず大きな位置を占めていた。ここで紹介したように、評判分析の分野に対して、Deep Neural Networkを適用する発表も多かった。また、評判分析アプリケーションの発展形として議論されることの多い、より広範囲なタスクとして、社会情勢分析を目指したタスクなど、時代の趨勢にあったアプリケーションが言語処理の観点から議論されていて、この会議の幅広さを実感した。

参考文献

- Altun, Y., Tsochantaridis, I., Hofmann, T., et. al (2003). “Hidden markov support vector machines.” In *ICML*, Vol. 3, pp. 3–10.
- dos Santos, C. and Gatti, M. (2014). “Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts.” In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pp. 69–78 Dublin, Ireland. Dublin City University and Association for Computational Linguistics.
- Pla, F. and Hurtado, L.-F. (2014). “Political Tendency Identification in Twitter using Sentiment Analysis Techniques.” In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pp. 183–192 Dublin, Ireland. Dublin City University and Association for Computational Linguistics.
- Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., and Potts, C. (2013). “Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank.” In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1631–1642 Seattle, Washington, USA. Association for Computational Linguistics.
- Vanzo, A., Croce, D., and Basili, R. (2014). “A context-based model for Sentiment Analysis in Twitter.” In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pp. 2345–2354 Dublin, Ireland. Dublin City University and Association for Computational Linguistics.
- Xu, L., Liu, K., and Zhao, J. (2014). “Joint Opinion Relation Detection Using One-Class Deep Neural Network.” In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pp. 677–687 Dublin, Ireland. Dublin City University and Association for Computational Linguistics.

略歴

小早川 健（正会員）：1993年東北大学理学部物理学科卒，1995年東京大学理学系研究科物理学修了。1995年から，日本アイ・ビー・エム（株）で勤務，1999年から現在まで日本放送協会放送技術研究所で勤務。機械翻訳，音声認識，音声対話の研究を経て，現在は評判分析の研究に従事。

（20xx年x月xx日依頼）

（20xx年x月xx日受付）