ACL2013 参加報告（その２）
– ベストペーパーおよび TACL セッションから論文を紹介 –

高瀬 翔[†]

1　はじめに

本稿では ACL2013 において発表された論文の中から，ベストペーパーとして表彰された論文 “Grounded Language Learning from Video Described with Sentences”(Yu and Siskind 2013) および，TACL（Transactions of the Association for Computational Linguistics）セッションにおいて筆者が特に関心を持った論文 “What Makes Writing Great? First Experiments on Article Quality Prediction in the Science Journalism Domain”(Louis and Nenkova 2013) について簡単な紹介を行う．ベストペーパーに表彰された論文をはじめとして，ACL2013 では，画像，動画と言語処理の融合に取り組んだ研究が散見された．これについて，詳しくは岡崎氏の「ACL2013 参加報告（その３）」報告を参照されたい．

TACL とは 2012 年より募集開始された論文誌であり，これに採択されると ACL の主催する会議[1]のいずれかで発表する資格を得られる．1 年間での投稿数は約 100 本，うち採択数は 28 本[2]であり，採択率が低過ぎるわけでは決してないが，丁寧な議論，調査が要求されるのは間違いない．ACL2013 では TACL 掲載論文の口頭発表は TACL セッションとしてまとめられており，非常に質の高い発表を集中して聞く機会を得られた．

2　Grounded Language Learning from Video Described with Sentences

本論文 (Yu and Siskind 2013) では，3-5 秒程度の動画とそれを説明する文のペアの集合から単語の意味，すなわち動画内での物体や物体の動作と単語との対応を学習する手法を提案する．具体的には，各単語の引数（$person(\alpha)$ や $carried(\alpha, \beta)$ における α や β のような，単語の取る

[†]東北大学, Tohoku University

[1]2013 年 9 月現在は ACL，NAACL，EACL，EMNLP の 4 つ．

[2]余談だがこのうち 9 本は改稿無しでの採択らしく，意外と採択数が多い印象を受ける．

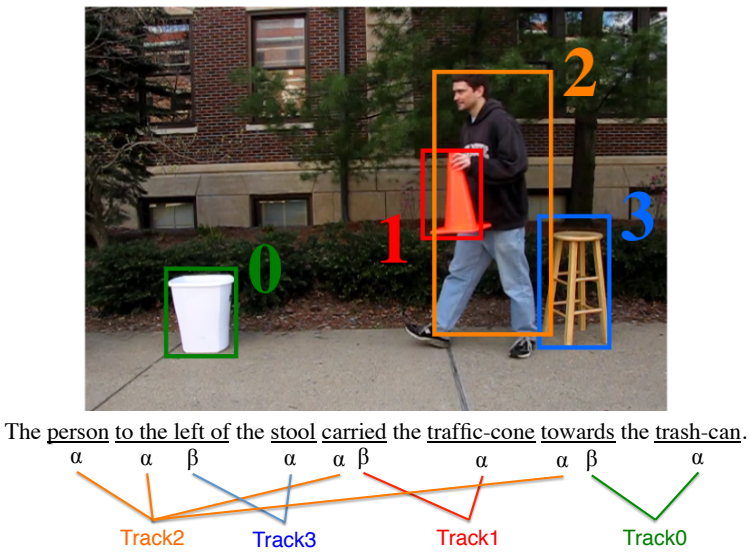

図 1　本論文の目的とする学習結果の例

項）が動画内のどの物体（領域）に対応するかについて，HMM を用いた教師無し学習を行う．図 1 にこの論文で入力としている動画の 1 フレームと，その説明文の例を示す[3]．なお，図 1 では動画内でのどれが物体か既に特定されているが，実際の入力にはこの情報は含まれていない．本論文はこのような動画とその説明文の集合を訓練事例とし，動画内での物体の特定，および図 1 下部に示したような物体と各単語の引数との対応を学習する．

本論文ではまず，動画内の各フレームにおける物体の特定を行う．各フレーム内で物体と推定された領域は，色や形，大きさなどの素性を持ち，さらにフレーム間の比較によって速さや移動の向きなどの素性も得られる．提案手法では，Yamato らの手法 (Yamato, Ohya, and Ishii 1992) にならい，この素性ベクトルを出力する HMM を学習する．各単語はそれぞれの HMM に対応するとしており，この HMM を単語の意味と考える．例えば “jump” という単語は速さと向きの素性を出力する 2 つの状態を持つ HMM で表現され，“quickly” は速さの素性を出力する 1 つの状態を持つ HMM で表現される．

なお，本論文では語彙はいくつかの名詞，動詞，副詞，前置詞に限られており，各単語の引数，すなわち単語毎の取りうる項の数も既知であるとする．これにより，複数の動画に同じ単語が出現するため，似た素性を持つ物体に同じ単語を割り当てることが可能となる．例えば一旦上昇し下降する物体には “jump” を，素早く動く物体には “quickly” を，というように，物体（素性ベクトル）と単語との対応を学習できる．また，本手法では各説明文における単語の引数間の対応

[3] この図は著者の論文に関する説明ページである http://haonanyu.com/research/acl2013/ より引用した．

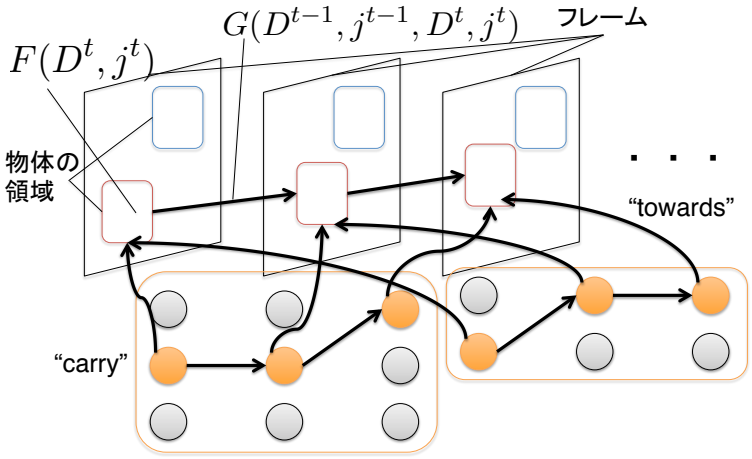

図 2　提案手法の概要

も既知であると仮定している．すなわち，図 1 の文について，$person(p_0)$，$to\text{-}the\text{-}left\text{-}of(p_0, p_1)$，$stool(p_1)$，$carried(p_0, p_2)$ のように，対応は明らかになっているものとしている．さらに，動画の各フレームにおいて物体を特定する detector は既存のものを使用するとし，detector は物体の領域を確信度（スコア）と共に出力するとする．

本論文での提案手法の概要を図 2 に示す[4]．提案手法では図に示したように，detector の出力した物体の領域スコア F とフレーム間での領域の遷移，すなわち領域が滑らかに動いているかを表すスコア G から物体を特定する．さらに図 2 の “carry” や “towards” のように，説明文内の単語に対応した，各フレームにおける物体の素性ベクトルを出力する HMM を学習する．なお，HMM の状態数，出力する素性の種類は単語の品詞に依存している．本論文では図 2 のように，複数の HMM を扱うが，これは Ghahramani ら (Ghahramani, I. Jordan, and Smyth 1997) の提案した Factorial HMM にもとづいている．実際には，動画を説明する文内の各単語（HMM）について積を取り確率を計算している．学習は EM アルゴリズムによって行うが，詳細は割愛する．

実験では与えられた動画について，説明文候補の中から適切な説明文を選択できるかどうかを検証する．結果はランダムに選択するベースラインよりも正解率が大きく向上しており，また，人手でパラメータを調整したモデルとの比較ではほとんど同じ結果が得られた．これにより提案手法では単語の意味，すなわち対応する物体および物体の動きについて学習できていると結論づけている．この論文はテキストだけでは表現することが難しい，物体や物体の動作という視覚によってとらえられる概念と単語との対応の学習を可能としており，非常に興味深い

[4] この図は，統計数理研究所の持橋大地氏の紹介スライド（http://chasen.org/~daiti-m/paper/sNLP2013-video.pdf）を参考にさせていただいた．

内容である．しかしながら，論文そのものは恐ろしく読みづらいため，著者らの説明ページ[5]や本報告筆者の第 8 回 NLP 若手の会における本論文の紹介スライド[6]，および持橋大地氏の紹介スライド[7]などを併読することをオススメする．

3 What Makes Writing Great? First Experiments on Article Quality Prediction in the Science Journalism Domain

本論文 (Louis and Nenkova 2013) では，読む価値の高い文書の検索を容易にするため，文書の質を自動で推定する手法を提案する．語のつづりや文法から質を予測する研究は存在する (Brill and Moore 2000; Tetreault and Chodorow 2008; Rozovskaya and Roth 2010) が，本論文では，対象ドメインを新聞（New York Times）の科学に関係のある記事（物語やインタビュー記事など複数のジャンルを含む）と定め，ドメイン特有の素性を設計し，その有用性を検証している．

概要としては，New York Times の科学記事について，The Best American Writing に掲載された著者の科学記事を「良質な記事」，それ以外の記事を「典型的な記事」として分けたデータを作成し，著者らの設計した科学記事に特有の素性を用い，SVM によりこの 2 種類の記事の分類器を構築する．科学記事に特有の素性は，「青々とした芝生」のように何らかのイメージを想起させるかという点や，独創的な表現が使われているかなど，6 つの観点を示し，それぞれについての素性を設計している．実験では「良質な記事」と「典型的な記事」の分類精度を検証しており，既存研究で提案されている素性よりも精度が向上したことを示している．

この論文は文書の質や面白さに貢献する要素について，心理言語学についても参照するなど広く調査している．また，各素性について，分類すべき 2 つのデータで有意に差があるか，すなわち，分類に貢献しそうかどうかを検証しており，論文として非常に丁寧に書かれている印象を受ける．反面，素性別では bag of words が最も精度の高い結果となっている点についての分析など，実験結果への考察があまいように感じられる．さらに，今回提案した素性には，coherence など，談話的な特徴を扱えるものが含まれていない．著者らは昨年の EMNLP において “A coherence model based on syntactic patterns”(Louis and Nenkova 2012) のタイトルで発表しているので，その辺りに取り組んだ結果が待たれる．

[5] http://haonanyu.com/research/acl2013/
[6] http://yans.anlp.jp/symposium/2013/files/takase.pdf
[7] http://chasen.org/~daiti-m/paper/sNLP2013-video.pdf

4 おわりに

本報告の筆者は数えるほどしか国際会議に参加していないが，ACL では特に，質の高い研究発表を聞く機会にめぐまれ，また，発表の質疑を通して自分では見落としていた観点に気づかされるなど，非常に刺激的な経験ができた．月並みな言葉かもしれないが，研究発表の場に参加する重要性を強く印象づけられる．しかしこのような心持ちになるのは一度体験してこそという気も有り，個人的には，多くの学生が研究発表および交流の場に積極的に参加することを願っている．

謝辞

本稿を執筆する機会を与えていただきました言語処理学会に感謝いたします．TACL の採択率についての報告は，国立情報学研究所の原忠義氏に情報を提供していただきました．ありがとうございます．また，本稿を執筆する上で，第 5 回最先端 NLP 勉強会における持橋大地氏の発表は大変参考になりました．発表資料中の図を使用することについても快く許諾いただき，深く感謝いたします．

参考文献

Brill, E. and Moore, R. C. (2000). “An improved error model for noisy channel spelling correction.” In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, pp. 286–293. Associationg for Conmputational Linguistics.

Ghahramani, Z., I. Jordan, M., and Smyth, P. (1997). “Factorial Hidden Markov Models.” In *Machine Learning*. MIT Press.

Louis, A. and Nenkova, A. (2012). “A coherence model based on syntactic patterns.” In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, EMNLP-CoNLL ’12, pp. 1157–1168 Stroudsburg, PA, USA. Association for Computational Linguistics.

Louis, A. and Nenkova, A. (2013). “What Makes Writing Great? First Experiments on Article Quality Prediction in the Science Journalism Domain.” In *Transactions of Association for Computational Linguistics*. Association for Computational Linguistics.

Rozovskaya, A. and Roth, D. (2010). “Generating confusion sets for context-sensitive error correction.” In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP ’10, pp. 961–970 Stroudsburg, PA, USA. Association for Computational

Linguistics.

Tetreault, J. R. and Chodorow, M. (2008). “The ups and downs of preposition error detection in ESL writing.” In *Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1*, COLING ’08, pp. 865–872 Stroudsburg, PA, USA. Association for Computational Linguistics.

Yamato, J., Ohya, J., and Ishii, K. (1992). “Recognizing Human Action in Time-Sequential Images using Hidden Markov Model.” In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

Yu, H. and Siskind, J. M. (2013). “Grounded Language Learning from Video Described with Sentences.” In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 53–63 Sofia, Bulgaria. Association for Computational Linguistics.

略歴

高瀬 翔：2012 年東北大学工学部情報知能システム総合学科卒業．同年，同大学大学院情報科学研究科博士前期課程に進学，現在に至る．知識獲得，情報抽出の研究に従事．

（2013 年 8 月 1 日依頼）
（2013 年 9 月 20 日受付）