

# ACL2014 参加報告 (その3)

## —言語の分散表現と少し文書要約—

菊池 悠太<sup>†</sup>

### 1 はじめに

本報告では ACL2014 において発表された, 表現学習と文書要約について取り上げる.

### 2 表現学習

まずは表現学習 (Representation Learning) について取り上げる. 表現学習では, 主に教師なし学習によってある言語単位 (主に単語) の分散表現を学習し, 言語の意味的な類似度を捉える必要のある問題で, 近年注目を集めている. 単語の表現学習では, 各単語に低次元で密な実数値ベクトルを割り当て, 各単語ベクトル表現がその単語の (意味的, 構文的) 特徴を捉えるように大量のテキストコーパスからベクトル空間を学習する. 学習されたひとつひとつのベクトルは embedding と呼ばれる. 他に伝統的な単語のベクトル表現の方法に, Bag of words を作る時に用いる One-hot 表現がある. One-hot 表現における単語ベクトルの次元は語彙数と等しく, 各次元が一つの単語に対応づく. すなわち, ある単語を表す単語ベクトルは, その単語に対応した次元のみが 1 で他の次元は全て 0 となる. Bag of words は文脈に現れた単語の one-hot 表現の総和によって作られる. これに対し分散表現は, ある単語を表すのに全ての次元が関連している点で異なる. One-hot 表現では表層的に異なった単語同士は全く別の記号として扱うが, 分散表現では似た単語同士が似たベクトルになるよう学習を行うことができる.

なお, 分散表現という用語そのものはニューラルネットワークの分野などでは古くから知られている概念である. 実際, 単語の分散表現学習ブームの先駆けとなった手法は最も基本的なフィードフォワード型ニューラルネットワークの枠組みの中に単語ベクトルを組み入れることで, 単語の分散表現を獲得した (Bengio et al. 2003).

<sup>†</sup>東京工業大学, Tokyo Institute of Technology

## 2.1 文単位への拡張

単語よりも大きな言語単位を対象とした分散表現を扱う研究ももちろん存在する．文や文書<sup>1</sup>を対象とするタスク（評価極性予測などの分類問題）を解くためである．もっとも，それらの研究も，多くは直接的に文をベクトルとして表すのではなく，文中に含まれる単語の分散表現を足がかりにする事が多い．そのため，処理の過程では文単位の分散表現が出来上がるが，学習されるのはベースとなる単語の分散表現と，それらを文単位に組み合わせるための関数である．また，特定の問題を解くことが目的であるため，単語の分散表現や関数は最終的にはその問題のために用意された教師データを用いて調整 (fine-tuning) される．Fine-tuning の結果，“そのタスクを解くのに特化した単語の分散表現ベクトル” が手に入る．

“A Convolutional Neural Network for Modeling Sentences” も，そのように単語の分散表現を用いて文レベルのタスクを解くことに取り組んでいる (Kalchbrenner et al. 2014)．この論文は，文レベルの処理を行うために畳み込みニューラルネットワーク (Convolutional Neural Network; CNN) の枠組みを利用している．CNN は，特に画像処理の分野でよく知られているニューラルネットワークの一種である<sup>2</sup>．CNN では，（特徴抽出用）フィルタのデータへの畳み込み (convolution) による特徴抽出と，特徴のサブサンプリングを行い微小な幾何学的変化に対する頑健性を得るためのプーリング (pooling) と呼ばれる処理を用いてデータの特徴を捉える（得られた特徴は feature map と呼ばれる）．得られた特徴 (feature map) に更に畳み込み層やプーリング層をつなげ多段に積み重ねる事で，より抽象的な特徴を捉えることができるようになる．本論文の手法も基本的な枠組みは画像処理における CNN と同じであるが，convolution の方式や pooling の方法など，いくつかの点において自然言語処理に即した変更が加えられている．まず畳み込みについては，画像処理においてはフィルタは基本的に正方形であり，一つのフィルタを画像全体に走査させる．本論文では，入力文中の単語の分散表現ベクトルを並べた行列である．そして各フィルタは，単語ベクトルの各次元毎に独立に用意されその次元のみで走査される．これは，フィルタサイズ以下の  $n$ -gram について，次元毎に特徴的なパターンを抽出する処理と捉えることができる．またプーリングの処理においても，画像処理では画像中の全ての正方領域に対して行われるが，本論文では単語ベクトルの次元毎にサブサンプリングが行われる．また，入力文の単語数に依存せず  $k$  個の要素だけ残すようにプーリングを行うことで，文ごとに入力サイズ（単語数）が異なっても，最終的には全て同じサイズのベクトルになるよう調節している．

他に単語の分散表現を用いて文レベルの処理を行う手法に，Recursive Neural Network (RecNN) や Recurrent Neural Network (RNN) などがある．それらの手法は本会議で提案された新しい

<sup>1</sup>本節では文を単位とする拡張を取り扱う．文書単位への拡張について，本会議では (Hermann and Blunsom 2014) などがある．

<sup>2</sup>物体認識に関するコンペティションである Large Scale Visual Recognition Challenge (ILSVRC) では，今年度はほぼ全ての手法が CNN をベースとしているほどである (<http://image-net.org/challenges/LSVRC/2014/results>)

枠組みではないので取り上げないが、次節で取り上げる話題ではそれらの手法が多く使われている。

## 2.2 異なる二つの概念を一つのベクトル空間に埋め込む

ここまで、文レベルの処理を行う新しいモデルとして CNN を用いた手法を紹介したが、本節では、異なる二つの概念を同じベクトル空間に埋め込むという研究に焦点を当てる。ACL2014 では特にこのような設定の研究が多かったように思われる。ここで異なる二つの概念とは、異なる言語(マルチリンガル)であったり、テキストと画像(マルチモーダル)のことを指す。それらに共通するのは、あらかじめ人手で“この二つは意味的に等価である”とアノテートされたデータを利用するという点である。それは例えばマルチリンガルなトピックではパラレルコーパスや対訳辞書、単語アライメントなどである。マルチモーダルなトピックでは、画像とそれに対する説明文を用いたものがあつた。

“Grounded Compositional Semantics for Finding and Describing Images with Sentences” が、画像とその説明文を用いた論文である (Socher et al. 2014)<sup>3</sup>。本論文は、画像とその説明文を同じベクトル空間にマッピングすることを目的としている。CNN により画像の特徴ベクトルを学習し、係り受け木をベースとした新しい RecNN の亜種を学習することで説明文の分散表現ベクトルを構築する。その後、画像の特徴ベクトルを説明文ベクトルのベクトル空間にマッピングするために線形写像を行う。目的関数では画像ベクトルと説明文ベクトルの正しい組み合わせの時に二つの内積が大きくなるようにパラメータの調整を行う。この調整は、説明文ベクトル構築用のパラメータと線形写像を行うための行列のみに対して行われる。これは学習データのサイズが十分に大きくないためで、学習データのサイズが大きくなれば単語ベクトルそのものや、画像ベクトル構築用の CNN も合わせた同時学習も可能となり、より精度も高まるであろうと著者らは述べている。

マルチリンガルなトピックにおいても基本的な原理は同じである。同じ言語内で類似した単語同士のベクトルを近づけたいのはもちろんだが、さらに言語間で対訳関係にある単語同士も近いベクトルになるように学習を行う。そのためにパラレルコーパスなどの資源を用いて、正しいパラレルデータのベクトル同士が、そうでないベクトル同士よりも近くなるように目的関数を設定する。以下、上記のような手法をベースとした、関連するロングペーパーのうちいくつかの論文を取り上げて簡単に説明する。

(Devlin et al. 2014) は本会議の best paper である。softmax 関数の分配関数が 1 になるよう学習することで計算コストを抑え、デコードに用いることに成功した。(Tamura et al. 2014) は RNN を用いた単語アライメントのモデルを提案している。IBM モデルなどの教師なしモデルを元に (Dyer et al. 2011) の手法を改良してアライメントの教師なし学習を実現してい

<sup>3</sup>正確には Transactions of the Association of Computational Linguistics (TACL) に投稿された論文である。

る。(Hermann and Blunsom 2014) は、文内の単語ベクトルを足しあわせたものを文ベクトルとみなし、対訳文のベクトル同士のコサイン類似度を最小化する。さらに論文内で文書単位までの拡張を行なっている。(Zhang et al. 2014) は対訳文から抽出した句をそれぞれの言語用の RecNN に入力し、出来た句ベクトル間の二乗誤差を最小化する。両言語はお互いに他方の言語の句ベクトルを gold なベクトルとみなし、共訓練の枠組みでパラメータを調整する。上記のうち (Devlin et al. 2014; Tamura et al. 2014) は、従来単言語で使われていたモデルを、ソース言語とターゲット言語の両方を入力とするように拡張している。対して、(Hermann and Blunsom 2014; Zhang et al. 2014) は単言語用のモデルを二つ用意し、出力が近いものになるように双方のモデルパラメータを調整する。

### 3 文書要約

ここまで分散表現について扱ってきたが、他に筆者が興味を持っている文書要約についても簡単に説明する。

まず、本会議における Long paper の Oral session において、要約のために設けられていたのは Parallel Session 4 の Track D における “Summarization” である。4 件の発表があるが、そのうち 1 件は生成的要約へ向けた要素技術の研究であり、あるイベントを説明する表現のテンプレートを集めるというテーマであった (Pighin et al. 2014)。

残り 3 件のうち 2 件は新しいタスクの提案、残りの 1 件は、文書の時系列を考慮することで Generic な<sup>4</sup> 要約の精度を向上させるというものであった (Ng et al. 2014)。

2 件の新しいタスクはそれぞれ Hierarchical summarization と query chain focused summarization と名付けられていた。Hierarchical summarization は、要約文書中のある文について、さらにその詳細を知りたい場合は “その文の要約” を更に提示させるという階層的な要約提示システムであった。元文書に階層クラスタリングを施し、クラスタ毎に要約を作成するというシンプルなモデルであった (Christensen et al. 2014)。

Query chain focused summarization は、query focused summarization (与えられたクエリに対する要約を生成する) と、update summarization (すでに知っている情報があるという設定で、新しい情報のみを含めた要約文書を生成する) を組み合わせたようなタスクである (Baumel et al. 2014)。文書を読んでいる途中で知らない用語が出てきた時に、その用語について簡単にまとめた文書が欲しいという設定を想定している。それを呼んだらまた知らない用語が出てくるのでそれについても同様に... という意味で “query chain” という名前になっている。この時、chain の途中ではすでにいくつかの要約を読んでいるため、既に出てきた情報は要約に含めない

---

<sup>4</sup>クエリを伴うといった特殊な条件設定のない、いわゆる一般的な文書の “要約” を行う。



図 1 優雅な昼食



図 2 主役はハンバーガーのはず

という update summarization の要素が含まれている。

#### 4 まとめ

本稿では、ACL2014 で発表された論文のうち、分散表現と文書要約に焦点を当てた。初めての参加だったが、会場ではスターバックスのコーヒーなどの飲み物が飲み放題で、部屋の wifi も無料で使えるなどの粋な計らいが多くてさすが ACL ともなると太っ腹である。会場となったホテルは港のすぐ横にあり、ホテル横には川も流れておりそこで散歩や昼食を取ると気持ちよかったです (図 1)。周辺は飲食店も充実しておりハンバーガーはどこも美味である。ただし、ハンバーガーを頼んだつもりでも、ときどきガーリックフライドポテトのハンバーガー添えが来たりするので注意が必要である (図 2)。来年は ACL が北京、NAACL がデンバーで開催されるが、発表のときに NAACL の方が拍手や歓声が大きかった気がしたのが印象的だった。

#### 参考文献

- Baumel, T., Cohen, R., and Elhadad, M. (2014). “Query-Chain Focused Summarization.” In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 913–922 Baltimore, Maryland. Association for Computational Linguistics.
- Bengio, Y., Ducharme, R., Vincent, P., and Janvin, C. (2003). “A Neural Probabilistic Language Model.” *J. Mach. Learn. Res.*, **3**, pp. 1137–1155.
- Christensen, J., Soderland, S., Bansal, G., and Mausam (2014). “Hierarchical Summarization:

- Scaling Up Multi-Document Summarization.” In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 902–912 Baltimore, Maryland. Association for Computational Linguistics.
- Devlin, J., Zbib, R., Huang, Z., Lamar, T., Schwartz, R., and Makhoul, J. (2014). “Fast and Robust Neural Network Joint Models for Statistical Machine Translation.” In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1370–1380 Baltimore, Maryland. Association for Computational Linguistics.
- Dyer, C., Clark, J. H., Lavie, A., and Smith, N. A. (2011). “Unsupervised Word Alignment with Arbitrary Features.” In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 409–419 Portland, Oregon, USA. Association for Computational Linguistics.
- Hermann, K. M. and Blunsom, P. (2014). “Multilingual Models for Compositional Distributed Semantics.” In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 58–68 Baltimore, Maryland. Association for Computational Linguistics.
- Kalchbrenner, N., Grefenstette, E., and Blunsom, P. (2014). “A Convolutional Neural Network for Modelling Sentences.” In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 655–665 Baltimore, Maryland. Association for Computational Linguistics.
- Ng, J.-P., Chen, Y., Kan, M.-Y., and Li, Z. (2014). “Exploiting Timelines to Enhance Multi-document Summarization.” In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 923–933 Baltimore, Maryland. Association for Computational Linguistics.
- Pighin, D., Cornolti, M., Alfonseca, E., and Filippova, K. (2014). “Modelling Events through Memory-based, Open-IE Patterns for Abstractive Summarization.” In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 892–901 Baltimore, Maryland. Association for Computational Linguistics.
- Socher, R., Karpathy, A., Le, Q. V., Manning, C. D., and Ng, A. Y. (2014). “Grounded Compositional Semantics for Finding and Describing Images with Sentences.” *TACL*, **2**, pp. 207–218.
- Tamura, A., Watanabe, T., and Sumita, E. (2014). “Recurrent Neural Networks for Word Alignment Model.” In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1470–1480 Baltimore, Maryland. Association for Computational Linguistics.
- Zhang, J., Liu, S., Li, M., Zhou, M., and Zong, C. (2014). “Bilingually-constrained Phrase

Embeddings for Machine Translation.” In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 111–121 Baltimore, Maryland. Association for Computational Linguistics.

## 略歴

菊池 悠太 (非会員) :

2013 年，東京工業大学 総合理工学研究科博士前期課程修了．同年，同大学博士後期課程に進学，現在に至る．

(2014 年 6 月 24 日依頼)

(2014 年 8 月 23 日受付)