

AWSuM開発過程の自動化に関する検討

近藤悠介

はじめに

英語論文執筆支援ツールAWSuMは、英語で論文を執筆する上で有用なツールであるが、その開発コストは膨大である。水本・浜谷・今尾 (2016) によれば、8名の大学院生により1,000本の論文に人手でムーブ・タグの付与に半年以上の時間を要している。タグ付与に関する作業時間が軽減すればAWSuMの発展の可能性はさらに高まる。

筆者は、人手で付与されたタグを正解とし、論文で使用される言語の特徴を用いて、タグ付与の過程を自動化できると考えた。水本・浜谷・今尾 (2016) では高頻度の4語連鎖上位250個を特徴量とし、セクション、ムーブのクラスター分析を行ない、高頻度の4語連鎖により、セクション、ムーブはかなりの程度分類可能であることが示されている。このことは、高頻度の4語連鎖を用いたムーブの予測が可能であること示している。

本稿で報告する自動ムーブ・タグ付与システムautoAWSuMは、水本・浜谷・今尾 (2016) で報告された、1,000本の論文のアブストラクトに自動でムーブ・タグを付与するシステムである。データの前処理、自動タグ付けはそれぞれPython 3.4.5のパッケージnlTK 3.2.1 (Bird, Keim, and Loper, 2009), scikit-learn 0.18 (Pedregosa, Varoquaux, Gramfort, Michel, Thirion, Grisel et al., 2011) を用いた。

1. autoAWSuM開発に使用したデータの概要、仕組み、準備

1.1 タグ付けされたアブストラクト

以下に人手によってセクション、ムーブ・タグが付与された論文の一例としてLarsen-Freeman (2006) のアブストラクトを示す。

<abstract>

<01_introduction>Seeing language as a complex, dynamic system and language use/acquisition as dynamic adaptedness ('a make-do' solution) to a specific context proves a useful way of understanding change in progress, such as that which occurs with a developing L2 system.</01_introduction>

<02_presenting_research>This emergentist shift of perspective provides another way of understanding previously observed characteristics of learner language, that is that its development is not discrete and stage-like but more like the waxing and waning of patterns; that, from a target-language perspective, certain aspects of the behavior are progressive, others, regressive; that change can be gradual and it can be sudden; and that the latter notably heralds the emergence of a new order qualitatively different and novel from earlier organizations.</02_presenting_research>

<03_describing_methodology></03_describing_methodology>

<04_summarizing_findings> In addition, when group data are disaggregated, it is clear that there are many paths to development. By closely examining the oral and written production of five Chinese learners of English, the emergence of complexity, fluency, and accuracy can be seen, not as the unfolding of some prearranged plan, but rather as the system adapting to a changing context, in which the language resources of each individual are uniquely transformed through use.</04_summarizing_findings>

<05_discussing_research></05_discussing_research>

</abstract>

AWSuMではPho (2013)の分類を援用し、アブストラクトでは、Introduction, Presenting the research, Describing the research, Describing the methodology, Summarizing the findings, Discussing the researchの5つのムーブ・タグが採用されている。当該のムーブが出現しない場合は、開始・終了タグのみが記述されている。このタグ付き論文はムーブ・タグに付与された番号順に整列されているので、実際のアブストラクトの文の出現順序とは異なる可能性がある。筆者が30個のタグ付け前のアブストラクトを確認したが文の出現順序が異なるものは確認されなかったので、自動タグ付けシステムの予測精度を評価する上では大きな影響はないと考えた。

1.2 自動タグ付与

前述したように、対象とする論文のムーブは4語連鎖を用いて、まとめられるということがクラスター分析で示されている。このことは言語的特徴を用いてムーブが予測できるということである。また、ムーブの出現順序には大きな制約がある。今回用いる1,000本のアブストラクトでは、最初の文のムーブは49%がIntroduction, 49%がPresenting the researchで残りの2%がDescribing the methodologyであった。また、前の文のムーブがIntroductionである場合、次の文のムーブは42%がIntroduction, 56%がPresenting researchで、残りの2%がDescribing methodology, Summarizing findingsで、Discussing researchは出現しなかった。このようなデータの性質から、ムーブ・タグの予測は文単位とし、ひとつ前の文のタグを所与として、タグ付けする文の言語的特徴を予測変数、ムーブ・タグを目的変数とする仕組みでムーブ・タグを自動付与するシステムの開発を試みた。

1.3 準備

前述したタグ付きアブストラクトでは文ごとにムーブ・タグは付与されていないが、これを文ごとに付与する。それぞれのアブストラクトをムーブ・タグで表現した正解データを用意する。例えば、5文で構成され、最初の文のムーブがIntroduction, 2番目がPresenting the research, 3番目と4番目がDescribing the methodology, 5番目がSummarizing the findingsの場合、[0,1,2,2,3]のように表現する（Introductionが0, Presenting the researchが1, Describing the methodologyが2, Summarizing the findingsが3, Discussing the researchが4のように名義尺度に変換した）。

すべての文をその文のムーブとひとつ前の文のムーブを用いて分類する。組み合わせとしては存在するが実際には出現しないカテゴリもあるので、すべての文は16のカテゴリに分類された。例えば、対象とした1,000本のアブストラクトでは、最初の文のムーブはIntroduction, Presenting the research, Describing the methodologyの3種類しか出現せず、前の文のムーブがDiscussing the researchであった場合、次の文のムーブは必ずDiscussing the researchである。

すべての文を品詞列に変換する。以下はLarsen-Freeman (2006)のアブストラクトの1文を品詞列に変換したものである。

In addition, when group data are disaggregated, it is clear that there are many paths to development.

IN NN, WRB NN NNS VBP VBN, PRP VBZ JJ IN EX VBP JJ NNS TO NN.

すべてのカテゴリでそれぞれの文を品詞の3-gramと4-gramの高頻度上位500種類の頻度でそれぞれのカテゴリ内のtf-idfを計算し、この値を標準化したもので文を表現する。

1,000本のアブストラクトを学習用データ900本とテスト用データ100本に分けた。900の学習用データにおいて、隠れ層が150のニューラルネットワークを上述の500の値を入力として文のカテゴリごとに学習させた。活性化関数は正規化線形関数（Rectified Linear Unit）を用いた。前に文がない場合、前の文がIntroductionの場合、前の文がPresenting the researchの場合、前の文がDescribing the methodologyの場合、前の文がSummarizing the findingsの場合の5個のニューラルネットワークをそれぞれのカテゴリのデータで学習させた。

1.4 自動タグ付与の手順

自動タグ付与は以下の手順で行われる。

1. 入力されたアブストラクトを文に分ける（以下の過程では1文ずつ処理される）。
2. 対象となる文を品詞3-gramと4-gramのtf-idfを計算し、この値を標準化したもので表現する。
3. 前の文のMoveによって異なるニューラルネットワークを用いて当該の文のMoveを決定する。
4. 文がなくなるまで2と3を繰り返し、文がなくなれば処理を終了する。

2. autoAWSuMの精度

テスト用データ100本を用いてautoAWSuMの予測精度を検証した。学習データでの精度は95%、テストデータでの精度は64%であった。それぞれの混同行列を図1および図2に示す。テストデータはランダムに100本のアブストラクトを選ん

でいるが、テストデータにより精度はある程度の幅があった。以下に示すテストデータの混同行列は平均的な精度のものである。

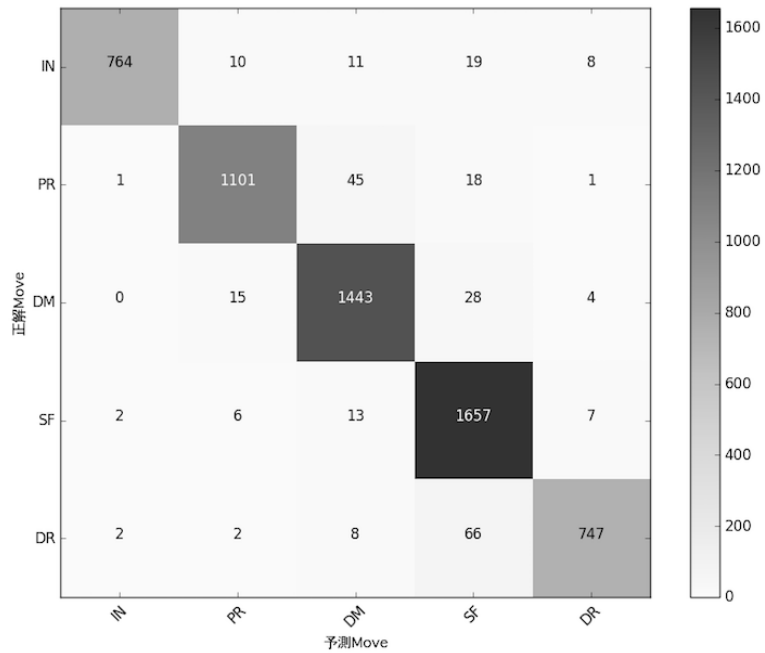


図1 学習データの混同行列

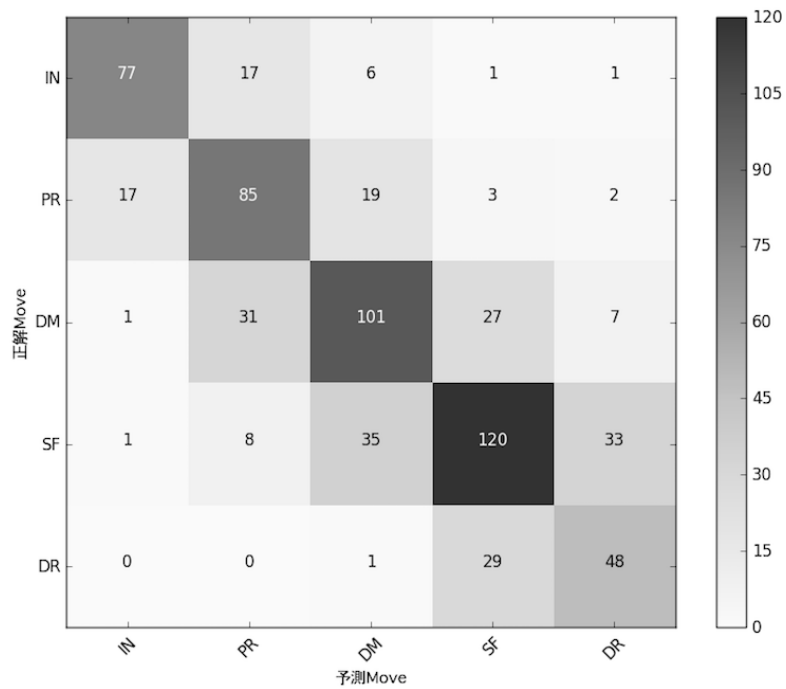


図2 テストデータの混同行列

図中のIN, PR, DM, SF, DRはそれぞれIntroduction, Presenting the research, Describing the methodology, Summarizing the findings, Discussing the researchを示す。

学習データ, テストデータ共に, Introduction, Presenting the researchに比べ, Describing the methodology, Summarizing the findings, Discussing the researchのMoveの予測精度が落ちていることが分かる。これらの精度の改善は今後の課題と言えよう。

おわりに

本稿では開発段階であるautoAWSuMの概要を述べた。近年, 英語学習を支援するツールの開発は盛んに行われており, 開発に関するコストがかかるツールもある。AWSuMは有用なツールであるが, その開発コストは膨大である。この開発コストを削減することにより, さらなる展開に人的, 時間的コストを費やせる。autoAWSuMはAWSuMの開発コスト軽減を目的に開発されたが, 現段階の予測精度ではその目的は達成されていない。予測精度を向上させるためには, 変数選択, データの制約, 予測方法に関して今後精査する必要がある。今回は, 変数として品詞n-gramを利用し, 頻度情報を元に変数を選択したが, 文の構造や語数なども考慮する必要がある可能性が高い。また, 今回の予測では文を単位として, Moveは前の文にある程度制約されていると仮定した。この仮定はある程度妥当なものと考えられるが, テストデータの予測精度に鑑みて, 十分に予測精度の向上に寄与する制約とは言えない。さらに, 今回はニューラルネットワークを予測方法として採用したが, 他の方法も検討すべきである。

引用文献

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*, CA: O'Reilly Media.

Larsen-Freeman, D. (2006). The Emergence of Complexity, Fluency, and Accuracy in the Oral and Written Production of Five Chinese Learners of English, *Applied Linguistics*, 27(4), 590–619.

- 水本篤, 浜谷佐和子, 今尾 康裕. (2016). ムーブと語連鎖を融合させたアプローチによる応用言語学論文の分析—英語学術論文執筆支援ツール開発に向けて—
English Corpus Studies, 23, 21–32.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., & Grisel, O. et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.