

薬物の系統的レビューにおける選択基準ベクトルの利用

Utilizing Selection Criterion Vectors in Pharmaceutical Systematic Reviews

佐々木 裕[†] 三輪 誠[†]
Yutaka Sasaki[†] Makoto Miwa[†]

安部 賀央里[‡] 頭金 正博[‡]
Kaori Ambe[‡] Masahiro Tohkin[‡]

[†]豊田工業大学工学部

[‡]名古屋市立大学大学院薬学研究科

[†]Toyota Technological Institute [‡]Graduate School of Pharmaceutical Sciences, Nagoya City University

1 はじめに

系統的レビュー (Systematic Review) とは、臨床的疑問 (Clinical Question) を解決するために、予め設定した基準に合致するあらゆる実証的根拠 (学術論文) を網羅的に収集・統合し、分析することで、臨床的疑問に関する科学的な結論を導く手法である。バイアスを最小限にするために体系的かつ明確な手法を用いて学術論文を選択することで、信頼性の高い知見の提示が可能となる。「エビデンス (科学的根拠) に基づく医療」 (Evidence-Based Medicine) の実践が重要視されるなか、系統的レビューは、エビデンスレベルが最も高いとされている。

系統的レビューは基本的に次のようなプロセスで行われる。

- (1) 研究対象とする臨床的疑問の決定
- (2) 研究論文の網羅的な収集および選択
- (3) 各研究のデータ抽出・妥当性評価
- (4) 統計学的解析
- (5) 結果の解釈

研究論文の網羅的な収集および選択のためには、選択基準を設定する。分析の対象とするための条件である「組み込み基準」 (Inclusion Criteria) と対象外とするための条件である「排除基準」 (Exclusion Criteria) を設定することにより、バイアスを最小限にして、対象となる論文を系統的に選択することができる。

研究論文の収集と選択は、専門家にとって非常に負荷のかかる部分である。現時点では、ある研究論文が基準を満たしているかどうかは、専門家が論文の内容を読み、理解することでしか達成できない。このような人手による厳密な検証が必要なため、分析できる対象論文数も限られる。

このような制約の下で、本研究では、ステップ(2)において、選択基準を文書分類の特徴とすることが、研究論文の網羅的収集・選択にどの程度有用かを検証する。特に、文書分類問題の中で選択基準をどのように特徴として取り込むかが課題となる。

2 系統的レビューデータ構築

Ito ら[4]は、名古屋市立大学大学院薬学研究科医薬品安全性評価学分野において、「2型糖尿病薬 (DPP-4 阻害薬) における有効性の民族差」を臨床的疑問に設定して、系統的レビューを以下のように実施した。

<PubMed アブストレビュー手順>

- ① PubMed にアクセス
- ② Advanced 検索により下記を検索
“Dipeptidyl peptidase 4 inhibitor”
+ “placebo”
- ③ Article types を Randomized clinical trial に限定し、112 件の文献を取得
- ④ 取得した文献のアブストを薬学の専門家が読み、組み込み基準、排除基準で研究の対象となる文献を選別

組み込み・排除基準は表 1 のように設定された。まず、アブストラクトを読んで表の排除条件を確認し、1 つでも排除条件にあてはまるものがあれば、アブストラクト判定表に排除した理由を記入し、その文献は除外した。不明なもの、情報がないものはすべてフルペーパーレビューを行った。その結果、検索により得られたアブストラクト 112 件のレビューを 2 人で独立して行い、48 件が最終的なフルペーパーレビュー対象と判断された。

表1 組み込み基準、排除基準の定義文

| | 組み込み基準 | 排除基準 |
|---------------|---|---|
| Study design | Randomized clinical trial(RCT), double blind | Crossover study, a pooled analysis, population pharmacokinetic/ pharmacodynamics analysis, single blind |
| Duration | Treatment 4 weeks or over | Treatment less than 4 weeks |
| Participants | Patients with T2DM, adult(over 18 years) | Healthy subjects, renal impairment, cardiovascular disease, metabolic syndrome, coronary syndrome, obese subjects, NODAT (new-onset diabetes after transplantation) |
| Interventions | DPP-4 inhibitors (once-daily) as monotherapy or add-on to other oral agents | Non- DPP-4 inhibitor, twice daily or once weekly DPP-4 inhibitors |
| | Placebo | Active comparator, no-comparator |

この膨大な作業を文書分類技術により軽減できれば将来的に系統的レビューにおける研究の大幅な効率化が期待できる。本論文では、最初のステップとして、検索により得られたアブストラクト集合から分析対象論文を選択する問題を、組み込み・排除基準を特徴に組み入れた SVM 文書分類問題として解き、どの程度の性能が得られるかを明らかにする。

3 準備

3.1 パラグラフベクトル

Le and Mikolov [5]は、word2vec を拡張して、文書に対するパラグラフベクトルの生成法を提案している。パラグラフベクトルの生成は、word2vec と同様に文脈内の単語と文脈外の単語のベクトルの距離を更新していく際に、文書に対して割り当てたベクトルの更新をかける。

PV-DM モデルは、文書 ID と特定の文脈内の単語を入力として、その文脈内でターゲットの単語が出現するかどうかを学習することで、文書 ID に対応するベクトル表現を生成する。

PV-DBOW モデルは、文書 ID のみを入力とし、その文書内に出現する単語を予測するモデルを学習することで、文書 ID に対応するベクトル表現を学習する。

パラグラフベクトルと同時に生成される単語ベクトルを用いて、新しい文書に対してもパラグラフベクトルを作成することもできる。本研究では、全体のデータサイズが小さいためテスト文書も含めてパラグラフベクトルを

作成するトランスダクティブなアプローチを採用する。

3.2 正規化

埋め込みベクトル表現を他の特徴と組み合わせて学習に利用する場合、ベクトル内の値が取る範囲を調整することで分類器の性能が向上することが知られている。

たとえば、SVM による学習では、学習結果の重みベクトル w は、訓練事例ごとの重み α_i に基づく事例の加重和として得られるため、各特徴が取り得る値の範囲が異なる場合は、振れ幅の大きい特徴の影響を受けやすくなる。そのため、 $\{0, 1\}$ の値と取る特徴と埋め込みベクトルを組み合わせる場合には、ベクトルの値の正規化 (normalization) が必要となる。

本研究では、埋め込みベクトルの要素の分散 σ と平均 μ を用いて、下式のように全体の分布の平均が 0、分散が 1 であるベクトルになるように調整する。

$$\hat{x}_i = \frac{x_i - \mu}{\sigma}$$

3.3 単語重みの利用

SVM モデルの学習は入力ベクトル間の内積に基づいているため、単語ベクトルを入力として利用する場合、単語の出現の $\{0, 1\}$ 値の代わりに単語ベクトルを配置するだけでは、各単語ベクトルの L2 ノルムの二乗を特徴値とすることと同等であり、ベクトル中に埋め込まれた情報が活用されない。また、単純に文書中に出現する単語の単語ベクトルの和や平

均を特徴としても、分類性能が良くならないことが知られている。

そこで単語の出現に関する特徴と単語ベクトルを組み合わせる場合、文書データから計算された tf-idf 等の単語の重み (term weighting) を利用する手法が提案されている [6]。単語の出現の有無を $x_i \in \{0,1\}$ で表す入力ベクトル $\mathbf{x}=(x_1, \dots, x_d)$ に対し、単語 ID が i の埋め込みベクトル \mathbf{v}_i 、その単語の重み α_i が与えられたとき、下記のような出現単語の加重和を文書の特徴とする。

$$\mathbf{x}' = \sum_{i: x_i=1} \alpha_i \mathbf{v}_i$$

4 提案手法

提案手法の概観を図 1 に示す。文書集合 $D=\{d_1, \dots, d_n\}$ 、組み込み基準 $I=\{i_1, \dots, i_m\}$ と排除基準 $E=\{e_1, \dots, e_k\}$ ($I \cup E$ を IE と略記) が与えられたとき、 $D \cup IE$ に対して gensim の doc2vec によりパラグラフベクトルを生成する。同時に単語ベクトルも生成される。組み込み・排除基準から作成された $\mathbf{v}(i_j)$, $\mathbf{v}(e_j)$ に対して、それぞれ $\mathbf{v}(d_k)$ との要素積を計算したベクトル列に正規化を適用する。正規化された特徴ベクトルと BoW (Bag of Word) 特徴ベクトルの連結を SVM の入力とする。

この方法によって、組み込み・排除基準を自然に SVM による学習に取り入れることができる。このような方法はパラグラフベクトルの構築が 2014 年に提案されて初めて実現可能となった。

5 実験

5.1 実験設定

下記の組み合わせを比較対象とする。

$\mathbf{v}_{\text{bow}}(d)$: 文書 d を BoW とみなしたときの単語の出現 $\{0,1\}$ のベクトル (baseline)

$\mathbf{v}_{\text{tfidf}}(d)$: 文書 d 中の単語の埋め込みベクトルの tf-idf 加重和 (3.3 節)

$\mathbf{v}_{\text{dm}}(d)$: 文書 d の PV-DM ベクトル

$\mathbf{v}_{\text{dbow}}(d)$: 文書 d の PV-DBOW ベクトル

$\mathbf{v}_{\text{dm}}(IE)$: 文集合 IE の PV-DM ベクトル列

$\mathbf{v}_{\text{dbow}}(IE)$: 文集合 IE の PV-DBOW ベクトル列

ベクトル \mathbf{v} に対して正規化 (3.2 節) をかけることを $\bar{\mathbf{v}}$ で表す。またベクトルの連結を $\|$ で表

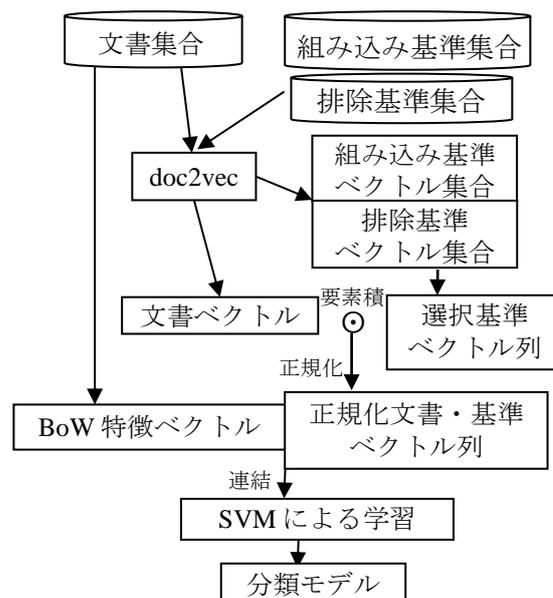


図 1 提案手法の概観図

す。たとえば、文書 d の単語の出現の特徴と文書 d の正規化 PV-DM ベクトルを連結することを $\mathbf{v}_{\text{bow}}(d) \| \overline{\mathbf{v}_{\text{dm}}(d)}$ で表す。要素積は $\mathbf{x} \odot \mathbf{y}$ で表し、ベクトル列 \mathbf{y} の各ベクトルと \mathbf{x} の要素積のベクトルの列と定義する。

薬物論文のレビューの負荷が大きく、深層学習に必要なデータ数を得ることが難しいため学習手法として LIBSVM を用いた。線形カーネルおよび 2 次の多項式カーネルを使用し、パラメータ C はデフォルトの 1.0 とした。 C の値を 0.01~10 の範囲で試行的に振ってみたが、スコアは改善されなかった。

評価は、データ数が少ないため安定した結果を得るため leave-two-out 交差検証による正解率 (Accuracy) で評価した。取りおいた 2 つのうち一方を開発用に用い、もう一方をテスト用に用いた。64/112=57.1% が多数側のラベルを常に選択したときに得られる得られる正解率である。

5.2 実験結果

実験結果を表 2 に示す。単語の出現を特徴とすること ($\mathbf{v}_{\text{bow}}(d)$) で、線形の SVM で 75% の正解率が得られた (baseline)。アブストラクトをパラグラフベクトルに変換したもの ($\mathbf{v}_{\text{dm}}(d) \sim \overline{\mathbf{v}_{\text{dbow}}(d)}$) を特徴としただけでは正解率は低下する。BoW 特徴に tf-idf 特徴ベ

クトルを連結しても正解率の改善は見られなかった ($v_{\text{bow}}(d) \parallel v_{\text{tfidf}}(d)$)。BoW 特徴にパラグラフベクトルの正規化を連結すると正解率が向上する ($v_{\text{bow}}(d) \parallel \overline{v_{\text{dbow}}(d)}$)。アブストラクトと基準の PV-DBOW の要素積の正規化を連結したとき、開発データにおいて 2 次の多項式カーネルによる正解率が 81.3% と最も高く、このときテストデータによる正解率が 80.4% であった ($v_{\text{bow}}(d) \parallel \overline{v_{\text{dbow}}(d) \odot v_{\text{dbow}}(IE)}$)。また、1 回の leave-two-out の学習時間ほどの特徴設定でも 0.03 秒程度であった。

6 関連研究

Aphinyanaphong and Aliferis [1] は、自動文書分類技術を用いて系統的レビューシステムを構築し、SVM が機械学習手法として最も効果的であることを示した。Frunza ら [3] は高齢者の健康管理サービスの普及戦略についての系統的レビューデータを対象に、医学シソーラス UMLS の概念や情報利得を利用して選択基準を適用する方法を述べた。Hashimoto ら [3] は、系統的レビューの能動学習にパラグラフベクトルと LDA を利用する方法を提案しているが、選択基準文のパラグラフベクトルを分類のための特徴としていない。一方、本研究は文書分類における選択基準の効果の確認を対象としている。

7 考察

単語ベクトルの tf-idf 加重和は本データについては正解率向上に寄与しなかった。これはデータ数の不足により、tf-idf 値があまり効果的でないためと考えられる。

参考のために畳み込みニューラルネットを本データに対して適用してみたが、データ数 112 と少ないことから 2 分割交差検証で 65% 程度の正解率しか得られなかった。同じ実験設定で線形 SVM による BoW は 75% であった。

8 おわりに

組み込み・排除基準のパラグラフベクトルの薬物系統的レビューデータ分類に対する効果を明らかにした。今後、フルペーパーに対するパラグラフベクトルの利用法を検討し、正解率の向上を図って行きたい。

表 2 実験結果 (正解率[%])

| 手法 | カーネル | | | |
|---|-------------|-------------|-------------|-------------|
| | 線形 | | 2 次多項式 | |
| | 開発 | テスト | 開発 | テスト |
| $v_{\text{bow}}(d)$ (baseline) | 75.0 | 75.0 | 70.5 | 69.6 |
| $v_{\text{bow}}(d) \parallel v_{\text{tfidf}}(d)$ | 75.0 | 75.0 | 70.5 | 70.5 |
| $v_{\text{bow}}(d) \parallel v_{\text{dm}}(d)$ | 75.0 | 75.0 | 70.5 | 69.6 |
| $v_{\text{bow}}(d) \parallel \overline{v_{\text{dm}}(d)}$ | 77.7 | 76.8 | 72.3 | 72.3 |
| $v_{\text{bow}}(d) \parallel v_{\text{dbow}}(d)$ | 75.0 | 75.0 | 70.5 | 69.6 |
| $v_{\text{bow}}(d) \parallel \overline{v_{\text{dbow}}(d)}$ | 80.4 | 78.6 | 57.1 | 57.1 |
| $v_{\text{dm}}(d)$ | 57.1 | 57.1 | 57.1 | 57.1 |
| $\overline{v_{\text{dm}}(d)}$ | 64.3 | 66.1 | 61.6 | 60.7 |
| $v_{\text{dbow}}(d)$ | 57.1 | 57.1 | 57.1 | 57.1 |
| $\overline{v_{\text{dbow}}(d)}$ | 67.0 | 67.0 | 67.0 | 67.0 |
| $v_{\text{bow}}(d) \parallel v_{\text{dm}}(d) \odot v_{\text{dm}}(IE)$ | 75.0 | 75.0 | 70.5 | 69.6 |
| $v_{\text{bow}}(d) \parallel \overline{v_{\text{dm}}(d) \odot v_{\text{dm}}(IE)}$ | 75.0 | 75.0 | 70.5 | 69.6 |
| $v_{\text{bow}}(d) \parallel \overline{v_{\text{dm}}(d) \odot v_{\text{dm}}(IE)}$ | 76.8 | 75.9 | 76.8 | 75.9 |
| $v_{\text{bow}}(d) \parallel \overline{v_{\text{dm}}(d) \odot v_{\text{dm}}(IE)}$ | 75.9 | 75.9 | 76.8 | 76.8 |
| $v_{\text{bow}}(d) \parallel v_{\text{dbow}}(d) \odot v_{\text{dbow}}(IE)$ | 75.0 | 75.0 | 70.5 | 69.6 |
| $v_{\text{bow}}(d) \parallel \overline{v_{\text{dbow}}(d) \odot v_{\text{dbow}}(IE)}$ | 75.0 | 75.0 | 70.5 | 69.6 |
| $v_{\text{bow}}(d) \parallel \overline{v_{\text{dbow}}(d) \odot v_{\text{dbow}}(IE)}$ | 77.7 | 77.7 | 81.3 | 80.4 |
| $v_{\text{bow}}(d) \parallel \overline{v_{\text{dbow}}(d) \odot v_{\text{dbow}}(IE)}$ | 77.7 | 77.7 | 79.5 | 80.4 |

参考文献

- [1] Aphinyanaphongs and Aliferis, Text Categorization Models for Retrieval of High Quality Articles, *JAMIA* 12:207-216, 2005.
- [2] Frunza *et al.*, Building Systematic Reviews Using Automatic Text Classification Techniques, *COLING-2010*.
- [3] Hashimoto *et al.*, Topic detection using paragraph vectors to support active learning in systematic reviews, *Journal of Biomedical Informatics*, 62:59-65, 2016.
- [4] Ito *et al.*, Ethnic Difference in the Pharmacodynamics-efficacy Relationship of Dipeptidyl Peptidase-4 Inhibitors between Japanese and non-Japanese Patients: A Systematic Review, *Clinical Pharmacology and Therapeutics*, 102(4):701-708, 2017.
- [5] Le and Mikolov, Distributed representations of sentences and documents, *ICML-2014*, 2014.
- [6] Sohrab *et al.*, Centroid-Means-Embedding: An Approach to Infusing Word Embeddings into Features for Text Classification, *PAKDD-2015*, 2015.