

機械学習を利用した看護文献レビューマトリックスシステム

Nursing Literature Review Matrix System based on Machine Learning

山下 貴範^{*1}
Takanori Yamashita伊豆倉 理江子^{*1}
Rieko Izukura中島 直樹^{*1}
Naoki Nakashima廣川 佐千男^{*2}
Sachio Hirokawa^{*1}九州大学病院メディカル・インフォメーションセンター
Medical Information Center, Kyushu University Hospital^{*2}九州大学情報基盤研究開発センター
Research Institute for Information Technology, Kyushu University

The literature review matrix is a method commonly used in writing a summary of literature surveys in the nursing field. In the literature review matrix, the literatures and literary groups obtained by the search are arranged on the vertical axis and the analysis aspects such as purpose, data, method, result and discussion are arranged on the horizontal axis, and features of each object are described in the cell. We applied machine learning to grasp the point of view being performed manually and to extract features of each view. We constructed a system that dynamically generates a review matrix according to the search. The present paper explains the idea, the method and case studies on 215 literatures on “text mining in nursing”.

1. はじめに

多くの論文から研究分野の特徴や傾向の推移を把握するための研究は多くされてきている。IBM 社の Watson は、膨大な医療論文の非構造化データを読み込み、患者のデータと比較して CDSS（医療意思決定支援システム）を目的とした診断や診療行為を推論することは有名である。他にも医療のリアルワールドデータである患者の検査結果数値、画像、テキストなど、あらゆる構造化・非構造化データに対して機械学習を用いた解析が活発になっている。しかし、これらは医師視点の解析が多いのが現状である。一方、看護においては、健常者から患者、病気の早期から末期、新生児からお年寄り、一人の看護師から組織全体と、医師よりも対象の範囲が広く複雑であるため看護を立案・実施するにあたり、医学的な視点だけではなく、心理・社会的・宗教的・スピリチュアルといった様々な視点から対象を捉え、把握することが必要とされる。診療行為はデータ記録されるが、患者等と直接的に関わる時間や看護内容の多様性を考慮すると、患者に対するテキストデータ記録量は医師に比べ多いのが現状である。

近年、看護研究の領域において、テキストマイニングを用いた研究が増えてきており、今後の展開が注目されている[大野 13, 西川 13]。[大野 13]らは、禁煙外来の会話から質的、量的に分析し、探索的研究を行なっている。[西川 13]らは、国際看護を対象にしてレポートなどの自由記載のデータからテキストマイニングを実施し、国際看護の教育を考察している。

そこで本研究では、「看護+テキストマイニング」の研究動向を調査するべく、それに関連する論文概要（215 件 1,166 文）を対象にして、文の分類（背景、目的、技術、結果、議論）を元に機械学習を行い、その結果のレビューシステムを構築した。

2. 動機

看護分野では以前よりグランデット・セオリーや現象学、エスノグラフィー等の手法を用いた質的研究が行われており、複雑な人間・社会的環境、生活の文脈や信念、価値観などを全体的にとらえることに対して有用である[グレッグ 16]。これらの研究では、研究者自身が自分の手で自身の主観性・感受性に基づき対象者の内面に入り深く解析し、客観的なモデルを作成するが、結果の客観性の点で未だ論争の一つとなっている[松原 10]。客観性を高めるといふ点からすれば、テキストマイニングは有用であるし、最近ではテキストマイニングを組み合わせた質的研究もみられる[稲葉 16]。

看護研究におけるテキストマイニングのモチベーションとしては、前述した通り複雑な背景をもつ『人』の抽象的な事象・現象に対して具体性を持たせることに貢献できると考える。また、研究者自身のデータのコード化やカテゴリー化に対して、コンピュータの解析結果を可視化し、照合や検討することで、客観性が高まり、対象がテキストで語る現象に近づく可能性がある。テキストマイニングを用いた看護研究はまだ発展途上ではあるが、複雑な看護分野だからこそ対象を明確化することに意味があると考えられる。

3. 機械学習による文種類の識別

学術論文は文学作品と違い独自の構造に従って書かれる場合が多い。それは、IMRAD あるいは MOVE と呼ばれる[Teufel 10, Hashimoto 16]。IMRAD は Introduction, Method, Result and Discussion の略で、論文はそれらの部分的な文書から構成されることを表している。文献調査のまとめ方として、対象とする論文を縦軸にとり、このような観点を横軸としたレビューマトリックスの形式がある[Judith 12]。文献データベースで文献リストを求め詳細に読みながら、マトリックスとしてまとめることで、文献調査を整理する方法として利用されている。しかし、その作業は人手で行われている。一方、機械学習を利用することで、論文中のそれぞれの文の内容が、IMRAD のどれに合致するかを自動的に推定する研究がある。本論文では、SVM と属性選択による[Sakai 12]の手法を論文概要に適

連絡先: 山下貴範, 九州大学病院メディカル・インフォメーションセンター, 〒812-8582 福岡市東区馬出 3-1-1, 092-642-5971, t-yama@med.kyushu-u.ac.jp

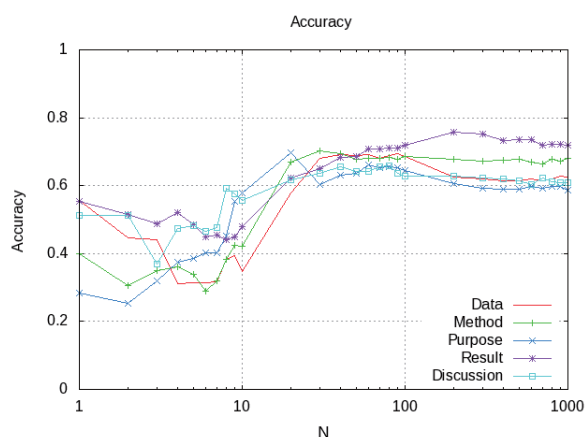


図 1: 属性選択による識別性能向上

用し、それぞれの文に目的、データ、手法、結果、議論という分類ラベルを割り当てる。これにより、自由記述の概要を意味的な半構造化文書として分析することができる。

具体的には、各ラベルに応じて SVM で得られる識別モデルを使い文のラベルを推定する。SVM で得られる識別モデルは、各単語がそのラベル推定にどれだけ有効かを表すスコアのベクトルとして表現される（表 1）。このスコアの上位 N 個の単語（スコアは正）と、下位 N 個の単語（スコアは負）の合計 $2*N$ 個だけで文をベクトル化して、推定する。図 1 は、 N を変化させたときの識別性能 (accuracy) をプロットしたもので、表 2 は最適な N とそのときの accuracy を示している。全ての単語でのベクトル化よりもごく少数の単語で識別性能が向上していることが分かる。

4. 文献レビューマトリックスシステム

本稿では、「看護」と「テキストマイニング」を含む 215 件の論文概要を対象とした検索エンジンを構築した。図 2 は、全論文を求める予約語「z」を検索語として与えたとき検索結果の画面である。検索窓に検索語を入力し、マトリックスの横軸の項目を「h」の行で複数してすることができる。図 2 では、Purpose, Data, Method, Result, Discussion, Author, Keyword の 7 項目を選んでいく。「h」の行の 5 項目は機械学習で識別する文種類であり、残りは、論文概要に本来含まれるメタデータである。分析対象とする項目を「v」の列の中から一つ選ぶことができる。図 2 では、研究者 (author) を選んでいる。その下の部分が検索結果の各対象について、各項目の特徴語の上位 5 個を表示するマトリックスである。一列目には、論文数の降順で研究者名が表示され、各行にその研究者の特徴を IMRAD の文種類とメタデータで示している。例えば、3 列目の Data を見ると 3 行目の「城丸瑞恵」は、他の 4 人とは違い文献情報が分析対象ということが分かる。これは、8 列目の Keyword に計量書史学があることと符号する。

図 3 は、分析対象を組織 (Organization) とした検索結果であり、研究者が 5 列目に表示され、各大学の研究グループや研究概要を把握することができる。

5. まとめと今後の課題

文献レビューマトリックス方式は、看護分野で一般的に利用されている文献調査のまとめ方である。文献レビューマトリッ

クスでは、検索で得られる文献や文献グループを対象として縦軸に並べ、目的、データ、手法、結果、考察などの分析観点を横軸として並べ、各対象の特徴をセルに記載する。本研究は、人手で実施している観点の把握と特徴抽出に機械学習を応用することで、検索に応じて文献レビューマトリックスを動的に生成するシステムを構築した。看護におけるテキストマイニングについての文献 215 件を対象とし、分析事例を紹介した。

観点ごとに生成した機械学習のモデルの識別性能が低いので改良が必要である。また、今回は実験対象が 215 件だけなので規模を拡大し、より一般的なモデルが必要と思われる。一般の文献レビューマトリックスでは、特徴語だけでなく短い文章での記述をする。特徴語を含む文の抽出で、どれだけできるかも今後の課題である。

参考文献

- [Judith 12] ジュディス ガラード: L^AT_EX: 看護研究のための文献レビュー マトリックス方式, 安部 陽子 訳, 医学書院, (2012)
- [Hashimoto 16] Hashimoto, K., Soonklang, S., Hirokawa, S: Feature Words of Moves in Scientific Abstracts, *Proceedings of 5th IIAI - AAI*, pp. 144-149, (2016)
- [Sakai 12] Sakai, T., Hirokawa, S: Feature Words that Classify Problem Sentence in Scientific Article, *Proc. iiWAS2012*, pp.360-367, (2012)
- [Teufel 10] Teufel, S.: The Structure of Scientific Articles: Application to Citation Indexing and Summarization, *Stanford: CSLI Publications*, (2010)
- [西川 13] 西川まり子, 村田直己, 小櫻愛美: 国際看護学への挑戦をテキストマイニングする, 特集看護研究におけるテキストマイニング (II) 看護研究, Vol. 46, No. 6, pp. 577-586, (2013)
- [大野 13] 大野佳子, 磯村毅: テキストマイニングによる成功率の高い禁煙プログラムの会話構造の特徴, 特集看護研究におけるテキストマイニング (II) 看護研究, Vol. 46, No. 6, pp. 552-567, (2013)
- [グレッグ 16] グレッグ 美鈴, 麻原きよみ, 横山美江: よくわかる質的研究の進め方・まとめ方第 2 版看護研究のエキスパートをめざして, 医歯薬出版, 第 2 版, (2016)
- [稲葉 16] 稲葉光行, 抱井尚子: 混合研究法としてのグラウンデッドなテキストマイニング・アプローチ, 看護研究, Vol. 49, No. 1, , pp. 25-35, (2016)
- [松原 10] 松原弘子: 質的研究における客観性に関する論考ーGTA 法と写真表現との比較を素材にアジア太平洋研究センター年報, (2010-2011)

表 1: 最適属性選択による識別性能

文種類	正の特徴語	負の特徴語
データ	名, 対象, もらう, 平成, カンファレンス	目的, なる, 法, いる, 離職
手法	面接, 用いる, 行う, 解析, アンケート	目的, 結果, 出産, 多い, 月
目的	目的, 明らか, 検討, ため, 個別	れる, 結果, 自分, なる, 手術
結果	結果, 多い, 抽出, 上位, 共起	示唆, 目的, 月, 日常, 考える
考察	示唆, 考える, れる, よう, 多く	目的, 分類, 名, 用いる, オノマトペ

表 2: 文種類ごとの特徴語

文種類	全単語性能	最適性能	最適 N
データ	0.6091	0.6938	90
手法	0.6716	0.7019	30
目的	0.5985	0.6962	20
結果	0.7299	0.7573	200
考察	0.6146	0.6582	80

z

Search

☒ detail Menu ☐ detail List

db nurseM

sort weight

☐ stop ☐ debug ☐ count hint 5

lf 1

h(☒ purpose ☒ data ☒ method ☒ result ☒ discussion ☒ author ☐ lastAuthor ☐ firstAuthor ☐ year ☐ source ☐ organization ☒ keyword ☐ title)

v(☐ word ☐ result ☒ author ☐ data ☐ discussion ☐ method ☐ source ☐ title ☐ lastAuthor ☐ year ☐ purpose ☐ keyword ☐ firstAuthor ☐ organization)

"z" returns 215 documents.

Author	Purpose	Data	Method	Result	Discussion	Author	Keyword
上野栄一 14	解析/ 字び/ 動詞/ グループ/ 名	解析/ 形態素/ 単語/ 動詞/ 名詞	動詞/ 基づく/ 出現/ 形態素/ 特徴	対人/ 作用/ 相互/ 関係/ ため	ペア/ 分析/ 作用/ 沿う/ 視点	上野栄一/ 一ノ山隆司/ 舟崎起代子/ 岡山真理/ 山名香奈美	質的研究/ チームナースング/ 看護学校/ 看護短大教育/ 臨床能力
いとうたけひこ 11	語り/ 当事者/ 進行/ 失調/ 乳がん	検索/ 医/ 誌/ 質的/ 量的	検索/ 医/ 誌/ インタビュー/ 質的	単語/ 研究/ 一般/ 全般/ 保健	効果/ 当事者/ 研究/ 姿/ 統合	いとうたけひこ/ 大高庸平/ 城丸瑞恵/ 小平朋江/ 鈴木ゆか	語り/ 計量書誌学/ 統合失調症(看護)/ インターネット/ 看護研究
城丸瑞恵 10	動向/ 医学/ 雑誌/ 中央/ 乳がん	検索/ 文献/ 医学/ 雑誌/ 中央	検索/ 文献/ 医学/ 雑誌/ 中央	文献/ 以降/ 単語/ 件/ 増加	単語/ 対象/ 研究/ 手術/ 頻度	城丸瑞恵/ いとうたけひこ/ 水谷郷美/ 大高庸平/ 加藤千佳	文献研究/ 医学中央雑誌/ 看護研究/ 看護教育/ 計量書誌学
今井多樹子 9	新人/ 紙/ 質問/ 求める/ 文	新人/ 回答/ られる/ 生/ 救命	文/ 回答/ 新人/ 中国/ 理由	主要/ 器/ 呼吸/ 人工/ 概念	職業/ 職務/ 先輩/ 同期/ 器	今井多樹子/ 岡田麻里/ 高瀬美由紀/ 池田敏子/ 船橋眞子	主成分分析/ 新人看護職/ ICU看護/ クラスタ分析/ 救急看護
入江拓 7	精神/ 大学生/ 要点/ 介入/ 内	室/ 保護/ 精神/ デジタル/ 字	精神/ 室/ 保護/ 大学生/ 変化	志向/ 課題/ 傾向/ ユーモア/ 学者	防衛/ 保護/ 精神/ 葛藤/ 室	入江拓/ 横井麗子/ 清水隆裕/ 比嘉勇人/ 松本浩幸	精神看護/ 簡易精神医学の評価尺度/ 患者隔離/ 看護教育/ 病室

図 2: 文献レビューマトリックスシステム

Organization	Purpose	Data	Method	Author	Keyword
聖隷クリストファー大学 10	精神/ 失調/ 統合/ 変化/ 症	保護/ 受講/ 体験/ 精神/ 企画	精神/ 保護/ イメージ/ 大学生/ 変化	入江拓/ 小平朋江/ 横井麗子/ 清水隆裕/ 比嘉勇人	精神看護/ イメージ(知覚)/ 教育手法/ 統合失調症/ 簡易精神医学の評価尺度
関西福祉大学 6	発言/ もつ/ 母親/ 育児/ 内容	ベネッセコーポレーション/ 株式会社/ 発言/ 母親/ ママ	発言/ 母親/ ベネッセコーポレーション/ ママ/ 株式会社	井田歩美/ 猪下光/ 谷川和昭/ 趙敏延/ 橋本裕	ソーシャルメディア/ 母/ 育児/ 不安(予防)/ 母子保健
昭和大学 6	動向/ 点眼/ 動作/ 容器/ わが国	中央/ 医学/ 検索/ 雑誌/ 質的	中央/ 医学/ 雑誌/ 検索/ 質的	城丸瑞恵/ いとうたけひこ/ 加藤千佳/ 鈴木ゆか/ 小川譲一	看護研究/ 医学中央雑誌/ 事故/ 人工呼吸器/ 計量書誌学
県立広島大学 5	必要/ 新人/ コミュニケーション/ スキル/ 表	新人/ 救命/ 当該/ センター/ 知識	知識/ 新人/ 回答/ 基礎/ レベル	今井多樹子/ 池田敏子/ 宮腰由紀子/ 高瀬美由紀/ 伊藤良子	ICU看護/ 新人看護職/ 主成分分析/ 病院救急医療サービス/ CCU
兵庫県立大学 4	感染/ 清掃/ 病室/ システム/ テーマ	感染/ 清掃/ 演題/ 会員/ 学会	感染/ 清掃/ 度数/ 病室/ 物品	高島真美/ 石垣恭子/ 周藤俊治/ 神崎初美/ 東ますみ	ドレナージ/ 経口投与/ 挿管法/ 療養病床/ 臨床検査技師

図 3: 分析対象を組織 (Organization) とした検索結果