

実務実習事前教育に向けたテキストマイニング手法の活用

濃沼政美,* 小池勝也, 中村 均

Application of Text Mining Approach to Pre-education prior to Clinical Practice

Masayoshi KOINUMA,* Katsuya KOIKE, and Hitoshi NAKAMURA

College of Pharmacy Nihon University, 7-7-1 Narashinodai, Funabashi City 274-8555, Japan

(Received September 14, 2007; Accepted February 22, 2008)

We developed a new survey analysis technique to understand students' actual aims for effective pretraining prior to clinical practice. We asked third-year undergraduate students to write fixed-style complete and free sentences on "preparation of drug dispensing." Then, we converted their sentence data into text style and performed Japanese-language morphologic analysis on the data using language analysis software. We classified key words, which were created on the basis of the word class information of the Japanese language morphologic analysis, into categories based on causes and characteristics. In addition to this, we classified the characteristics into six categories consisting of those concepts including "knowledge," "skill and attitude," "image," *etc.* with the KJ method technique. The results showed that the awareness of students of "preparation of drug dispensing" tended to be approximately three-fold more frequent in "skill and attitude," "risk," *etc.* than in "knowledge." Regarding the characteristics in the category of the "image," words like "hard," "challenging," "responsibility," "life," *etc.* frequently occurred. The results of corresponding analysis showed that the characteristics of the words "knowledge" and "skills and attitude" were independent. As the result of developing a cause-and-effect diagram, it was demonstrated that the phase "hanging tough" described most of the various factors. We thus could understand students' actual feelings by applying text-mining as a new survey analysis technique.

Key words—pharmacy education; text mining; free answer; morphological analysis; multivariate analysis; quality control

緒 言

薬剤師実務実習の強化・充実を図り、医療現場で活躍できる高い資質を有する薬剤師の養成を目的に、薬学教育が6年間に延長された。中でも大学で行う実務実習事前学習は、病院・薬局での実務実習に先立ち、薬剤師業務の本務である調剤などの基本的知識、技能、態度の修得を目的とし、実効ある実務実習にするためにも極めて重要な役割を担っている。

医療の現場では患者の背景や病態など様々な状況を考慮した医療が行われているように、教育の現場においても学生の理解度や認識度を十分に踏まえて教育していくことが、教育効果の向上につながるものと考えられる。そこで本研究では、効果的な実務

実習事前教育を行うため、学生の真意を把握するための新たな調査・解析手法の開発を行った。一般にアンケート調査等に用いられる質問形式は、選択肢の中から該当する数値や記号を選ぶ「プリコード形式」と、回答者が回答を自由に記述する「フリーアンサー形式」に大別される。¹⁾プリコード形式は回答者が答え易く、質問者が意図する分析に適した数値データが得られる反面、結果を誘導するような質問構成になる場合がある。一方で、フリーアンサーは回答者に負担がかかるが、回答者の真意を引き出すことが可能とされている。そこで筆者らは学生の真意を測定するためには、定型化されていない文章の集まりを自然言語解析の手法を使って単語やフレーズに分割し、それらの出現頻度や相関関係を分析して有用な情報を抽出するテキストマイニングの手法が適切であると考えた。なお、テキストマイニングとはデータマイニングの一手法であり、文章の中に埋もれている共通性や相関関係を発掘し、有用

日本大学薬学部(〒274-8555 千葉県船橋市習志野台7-7-1)

*e-mail: koinuma.masayoshi@nihon-u.ac.jp

な知見を獲得する分析方法である。²⁾近年では、医療分野において退院サマリーから自動的に疾患分類を行う試み³⁾や、教育分野において大学院改革のワークショップ⁴⁾等に用いられ、その有用性が示されている。さらに本研究では、テキストマイニングによる解析精度の向上を図るために、感性という知性を活用する人間活動の深層心理を探る研究分野である感性工学^{5,6)}の技術を応用した。そして、フリーアンサーから得られた言語から累積度数を定量し言語の出現頻度や割合から、回答者母集団の概略的な構成に関する解析を行った。また、フリーアンサーに含まれる、原因とその結果の内容から、対応分析の布置図を作成し、原因に起因する結果間の関連性について解釈を行った。また回答者の真意を得るため、頻出する言語に対し Fish-Bone 型特性要因図（以下、特性要因図）を作成し、結果に至る諸原因についての説明を行った。

そして実際に開発した手法を用い、学生の心理構造を把握した上で事前教育の方向性について検討を行った。

対象と方法

1. 対象 日本大学薬学部3年次に在籍する学生245名のうち病院・薬局実習の事前教育科目である「病院薬学概論」の受講者224名(91.4%)を対象とした。

性別は男性107名(47.8%)、女性117名(52.2

%)であった。

2. 調査実施 「病院薬学概論」の初回講義開始直後に、対象に本調査内容に関する事前情報を与えない状態で実施した。なお調査時期は2006年9月とした。

3. 方法 テキストマイニングの工程は、定型自由文による調査、形態素解析によるキーワードを用いて再構成した自由文(以下、キーワード自由文)への変換・分解、キーワードのうち特性に分離された言語(以下、特性キーワード)をKJ法により分類、自由文中のキーワード延べ出現度数の定量、対応分析による特性分類とキーワードのうち要因に分離された言語(以下、要因キーワード)の関係図示、特性キーワードと要因キーワードの特性要因図作成という一連の手法とした。

3-1. 定型自由文による調査 対象より、「調剤に対する認識」という題材についてフリーアンサー形式で調査した。記述形式は、林らの開発した文章完成形式の定型自由文^{7,8)}(以下、フリーアンサー自由文)とし、「～ので」の「～」部分をPart.A、「～から」の「～」部分をPart.B、「～です」の「～」部分をPart.Cとした(Fig. 1)。回答は1名当たり3文章以上とし、20分間の回答時間のうち、調査票を回収した。

3-2. 形態素解析によるキーワード自由文への変換 フリーアンサー形式で書かれた文章を解析する場合、形態素解析⁹⁾という技術を用いて文字列か

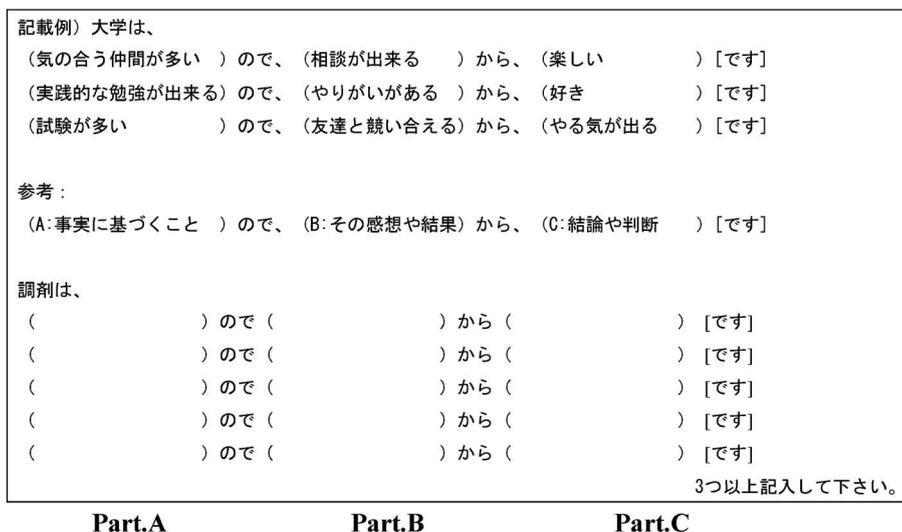


Fig. 1. Schedule (Fixed style complete and free sentences)

ら名詞や形容詞、助詞といった要素（キーワード）に分解する必要がある。現在いくつかの形態素解析システムが存在するが、本研究では、奈良先端科学技術大学院大学松本研究室で開発されたテキストマイニングソフトウェア「Win Cha 2000」¹⁰⁾を使用した。「Win Cha 2000」は、日本語文を単語に分割し、品詞、読み、活用形などの情報を単語に付与することが可能であり、文脈長可変の隠れマルコフモデルに基づく言語モデルをコスト最小法の枠組みに置き換えて解析処理が行われる。¹¹⁾ また、テキストマイニングソフトウェアとしてわが国で汎用される標準的なソフトウェアでもある。調査票に記載されたフリーアンサー自由文をテキストデータ化し、本ソフトウェアにより全文解析を行った。得られた形態素の品詞情報を基に、形容詞（自立・接尾・非自立）、動詞（自立・助詞類接続）、名詞（サ変接続・一般・形容動詞語幹・固有名詞）を抽出し、自由文の主要な構成言語として「キーワード」を生成した。なお、キーワードが2種以上で構成される場合は、“|”記号でキーワードを区切った。ここで得られたキーワードを用い、フリーアンサー自由文をキーワード自由文へ変換した。

3-3. キーワード自由文の分解 自由文の要因及び特性の関係について解析を行うため、キーワード自由文を要因（原因）及び特性（結果）に分解し、要因キーワード及び特性キーワードを組み合わせた言語の対を作った（Fig. 2）。以下、この組み合わ

せを「要因-特性」Pairと表記する。

3-4. 特性キーワードの分類とキーワード累積度数の定量 「調剤」の認識に対する全般的な傾向を把握するため、本研究3名がKJ法に基づいて、¹²⁾ 特性キーワードの分類を行った。なお、これら研究者の背景は、薬剤師実務家大学教員2名（病院・薬局勤務薬剤師歴15年以上）、基礎科目担当大学教員1名（薬剤師）とした。さらに分類した特性分類毎に出現したキーワードの累積度数を定量し累積率を算出した。

3-5. 特性分類と要因キーワードの関係 対応分析は、ともに質的データである2変量の間を視覚的、数量的に評価し、カテゴリ間の反応パターンの類似性を布置図に表すことができる。¹³⁾ そして得られた布置図の軸は成分スコアと呼ばれ、軸の単位はこれら成分の相対的な比率を示し、布置された変数は、変数間の距離よりも中心点からの放射状に引いた線の角度が近いものにより関連が深いものと解釈される。そこで要因キーワードに対応する特性分類の関係を把握するために、クロス集計に基づいた対応分析を行った。

3-6. 特性キーワードと要因キーワードの関係 教育面において、調剤の認識に関する特性がいかなる要因によるものかを把握することは、学生の心理を探る上でも極めて重要なことである。特性要因図を用いた分析は、現状の把握や解析がある程度進んだ段階での特性と要因の関係を整理するために用

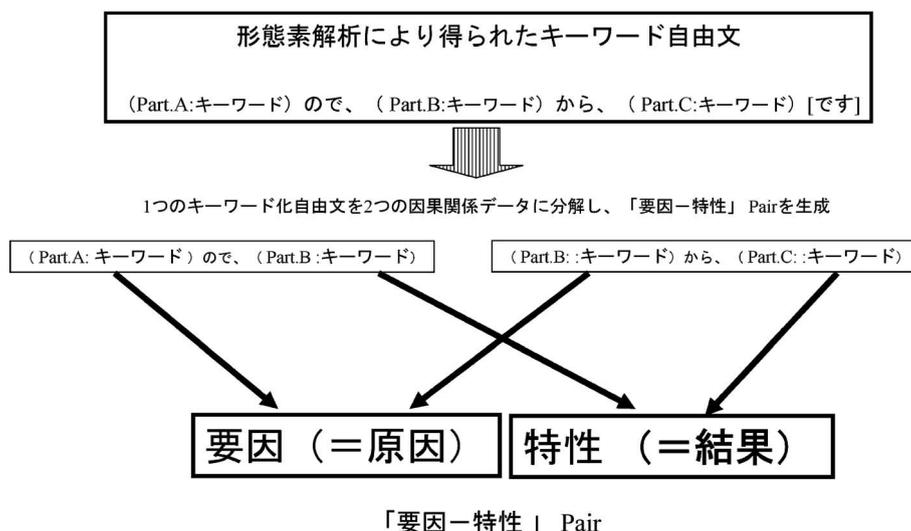


Fig. 2. Resolution of key word free sentence

いられる品質管理分野の判断手法であり、具体的な要因を視覚的に表現することで、次段階の「原因の究明」や「対策の立案」策定に活用される。そこで個々の特性に対する学生の心理構造を表すために「要因-特性」Pairに基づき、それぞれの枝に存在するキーワードが、孫、子、親のような連鎖関係となるようなロジックを組み、ソフトウェアに特性要因図を描画させた。本図では特性に対する要因が複雑であるほど樹形図は深層まで達し、キーワードの種類が多くなるほど枝は長くなる性質を持つ。

3-7. 解析ソフトウェア 対応分析及び特性要因図の描画には、JMP®6.03 (SAS Institute) を使用した。

結 果

1. 集計と形態素解析 調査票の回収率は100% (224/224) で、回答者1名について 3.46 ± 0.83 件 (mean±S.D.)、合計776件のフリーアンサー自由文を得た。これらのフリーアンサー自由文を形態素解析により、キーワード化自由文へ変換したところ、1521対の「要因-特性」Pairが得られた。

2. 特性キーワードの分類とキーワード累積度数の定量 KJ法の作業プロセスに準じて特性キーワードを分類した結果、知識、技能・態度、イメージ、リスク、体調・体力、その他の6項目に分類された。これらの分類は、個々の分類の独立性を高める上で可能な限り独立した概念として取り扱うこととしたが、直感的な作業を必須とするKJ法の作業プロセスにおいて、イメージ及びその他項目に関しては、重複する部分がある概念も存在した。なお分類の内訳は、技能態度510件 (33.5%)、イメージ502件 (33.0%)、リスク208件 (13.7%)、知識

184件 (12.1%)、体調・体力31件 (2.0%) であった (Fig. 3)。なお、特性分類の知識、技能・態度、イメージ、リスクにおいて、累積度数が上位5項目を占めるキーワード (累積率) を Table 1 に示す。

3. 特性分類と要因キーワードの関係 対応分析より得られた布置図を Fig. 4. に示す。本分析の結果、特性分類である知識と技能態度が中心より放射状に引いた線が相反するベクトルを示した。

4. キーワードと要因キーワードの関係 調査対象が薬学生であったことから「調剤」に対する認識において、知識-特性や技能態度-特性が得られたことはある意味必然的でもある。そこでこれらの特性以外の認識について把握するため、イメージ-特性に対する特性要因図について解析を行った。

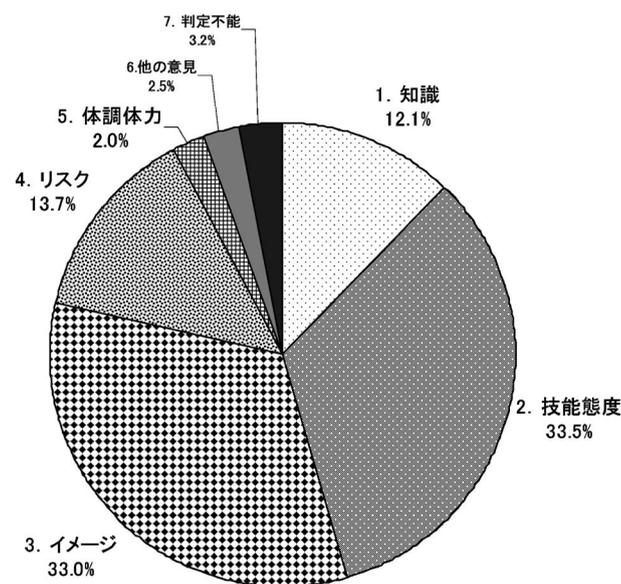


Fig. 3. Classified characteristic: Proportion of characteristic data to "Dispensing of drugs"

Table 1. Cumulative percentage of characteristic key word (Classified characteristic)

特性分類	上位5項目のキーワード (累積率: %)				
	1st	2nd	3rd	4th	5th
知識 (n=184)	知識 必要 20.7	勉強 する 13.0	勉強 必要 6.0	覚える 多い 5.4	考える 4.9
技能・態度 (n=510)	慎重 行う 12.9	注意 必要 8.4	行う 7.5	注意 する 3.9	気 つける 3.7
イメージ (n=502)	大変 14.7	やりがい ある 12.4	責任 重大 9.4	責任 ある 5.6	命 関わる 4.6
リスク (n=208)	間違う 32.2	ミス 起きる 10.1	ミス 許す 9.1	ミス する 5.3	事故 起きる 5.3



Fig. 6. Characteristic key word “Image”: Cause characteristic diagram of 〈ABUNAI〉

んばる」という特性であり、最も複数の要因により構成されていた。

考 察

キーワードの特性分類を行った結果、実習前の学生の「調剤」に対する認識は、知識よりも技能態度、リスク等に関心を持つ傾向が認められた。また、対応分析の結果、特性分類における知識と技能態度が対極的な付置関係となり、知識と技能態度の特性が、相対的に内容を異にする要因によりなり立っていることを示す。具体例を提示し解説する。例えば、「調剤は、《要因 Key Word：難しい》」ので、《特性（技能態度）：慎重に行う》という文章と、「調剤は、《要因 Key Word：難しい》」ので、《特性（知識）：知識が必要である》という文章は、特性分類が（技能態度）と（知識）で異なっているが要因 Key Word が「難しい」というキーワードで一致している。このように特性分類は異なっても要因 Key Word が一致する形式の文章が多く存在するに依り、対応分析の布置関係として技能態度と知識の中心点からの放射状に引いた線が接近する（布置パターン 1）。一方で、「調剤は、《要因 Key Word：難しい》」ので、《特性（技能態度）：慎重に行う》と、「調剤は、《要因 Key Word：リスクがある》」ので、《特性（知識）：知識が必要である》という文章は、特性分類が（技能態度）と（知識）で異なっている上、要因のキーワードに関しても「難しい」と「リスクがある」ということで一致していない。このような形式の文章が多く存在するに依り、逆に布置関係として特性分類の技能態度と知識のそれぞ

れ中心からの遠心線が異なった方向に向かう（布置パターン 2）。本対応分析の布置図は、布置パターン 2 であると解析される。この結果を考察すると、対象となった実習前の学生の意識は、学部 1、2 年次に実施された病院見学や薬局見学等の経験から、調剤はあくまでも知識と技能態度（特性）が必要であると認識していたが、この知識と技能態度はなぜ必要であるかという理由（要因）に関しては、相似した要因キーワードはあまり与えられないことが分かった。筆者らは研究前に、知識と技能態度が特性を示すことは予想していたが、これらの要因はいずれも「患者」等の類似したキーワードが多く該当するものと考えていた。しかし本分析結果から、この予想は誤っていたことが示された。実務としての調剤業務では、知識と技能態度の一方が欠けてもなり立つものではなく、その要因に関しても異なるものではないと考えられる。しかしながらこのような結果となった背景には、早期学年である 1、2 年次に、カリキュラム的にも調剤等の薬剤師業務に関連した科目も少なく、学生の認識がまだ樹立していないことが、個々に独立した要因を生じた原因であったと想定された。このため今後は、できるだけ早期年次より、薬剤師業務に関連した科目設定を行うなどの工夫が必要であるものと考えられ、そして「調剤はなぜ行うのか」という具体的な要因を教育者側が一定の見解を示した上で事前教育することが必要であると思われた。

また、調剤に対し「誇り | 思える」理由に関しては、「患者の命に関わる仕事」であるということが学生の心理に存在していることが分かった。また、ネガティブなイメージとして、「危ない」理由に関しては、「単純作業による慣れ」との認識が学生の心理に存在していることが分かった。このような結果から、大学での事前教育においては、実際の処方例を使用した症例検討を行うなど、常に患者の利益やリスクを第一に考えさせる履修方法を構築すべきと考えた。なお、本解析で最大の樹形図を示したのは「がんばる」という特性であったが、これはすべての学生が真の意味で「がんばる」と考えた学生が記述したものであれば、実習を踏まえた学生の立場としては非常に好ましいことであるといえる。しかしながら、文章完成式定型自由文の調査票であったことから、最後に括弧があまったため、とりあえ

ず都合がよい「がんばる」を入れたものであるのか、本結果からだけでその真偽を検証することはできない。恐らくこの点が、文章完成式定型自由文アンケートの限界であるものと考えられ、解析結果の解釈には、実務実習後の調査結果なども含め総合的に行うべきであろう。

本研究では、効果的な実務実習事前教育に向け、学生の真意を把握するために、テキストマイニング手法を活用した新たな調査解析手法の開発を行った。そして実際に開発した手法を用い、学生の心理構造を把握した上で事前教育の方向性について検討を行った。今後は、事前教育以外の場面においても本手法を活用していくことにより、効果的な薬学教育の実践を行いたいと考える。そして結果的に、国民が求める高い資質を持った薬剤師が1人でも多く世の中に誕生することを望む。

REFERENCES

- 1) Nikkei research report 2004-1, Nikkei research, Tokyo, 2004, pp. 26-28.
- 2) IT terminological dictionary “e-Words”, Incept. Co.,Ltd. (<http://e-words.jp/w/E38386E382ADE382B9E38388E3839EE382A4E3838BE383B3E382B0.html>) 2007.2.20 (Access).
- 3) Suzuki T., Ono H., Yokoi H., Imiya A., Takabayashi K., *Jpn. J. Med. Info.*, **25**, 173-180 (2005).
- 4) Kamoi K., Satoh I., Iwahisa M., Sato T., Nakahara I., *JJADE*, **20**, 25-30 (2004).
- 5) Shimizu Y., Kasuga M., “Role of sensibility engineering in contemporary society,” The Science Council of Japan, 2006, pp. 3-4.
- 6) Hayashi T., *FRAGRANCE J.*, **4**, 46-51 (2004).
- 7) Hayashi T., Takyu H., Dohkan K., Hirano H., *Ann. Conf. Jpn. Soc. Kansei Eng. (JSKE) abstract*, **2**, 58 (2000).
- 8) Hayashi T., Doukan K., Takyu H., Hirano H., *Ann. Conf. Jpn. Soc. Kansei Eng. (JSKE) abstract*, **3**, 149 (2001).
- 9) Ono H., Takahashi K., Suzuki T., Yokoi H., Imiya A., Satomura Y., *Jpn. J. Med. Info.*, **24**, 35-44 (2004).
- 10) Matsumoto Y., *IPSJ Digital Courier*, **41**, 1208-1214 (2000).
- 11) Matsumoto Y., Takaoka K., Asahara M., Kudo T., *Journal of Japanese Society for Artificial Intelligence*, **19**, 334-339 (2004).
- 12) Ohiwa H., Takeda N., Kawai K., Shimomi A., KJ editor: a card-handling tool for creative work support, *Knowledge-Based Syst.*, **10**, 43-50 (1997).
- 13) Hirono M., Hayashi T., “Use technique of multivariate data by JMP software,” Kai-bundo, Tokyo, 2004, pp. 129-146.