
ポスター | 知識工学／医療データ解析

ポスター1

知識工学／医療データ解析

2019年11月22日(金) 14:50 ～ 16:20 ポスター会場2 (国際展示場 展示ホール8)

[2-P2-1-07] がん電話相談の質的分析に基づいたテキストマイニングによる可視化

○石川 大介¹、片山 佳代子²（1. 地方独立行政法人神奈川県立病院機構 神奈川県立こども医療センター, 2. 地方独立行政法人神奈川県立病院機構 神奈川県立がんセンター）

キーワード：Cancer Telephone Consultation, Qualitative Analysis, Text Mining

神奈川県立がんセンターでは、2006年から2014年までがん電話相談事業を行っており、がん電話相談窓口には約14000件の電話相談が寄せられた。しかし、この電話相談の内容を記録したデータは、会話に基づく相談内容を書き留めたテキストデータであり、機械処理が不向きなテキストである。また件数も多く全てを目視するのは容易ではない。

そこで本研究では、相談データからサンプリング抽出した対象に質的分析を実施し、その結果に基づいて全てのデータを対象にテキストマイニングを行い、相談内容の可視化を試みた。

まず、電話相談を次の4グループに分類した。話者の性別と相談の対象から、

1.男性話者による自身の相談、2.男性話者による家族の相談、3.女性話者による自身の相談、4.女性話者による家族の相談、の4グループに分類した。この各グループから100件の相談をランダム抽出し、合計400件の相談内容に対して質的分析を実施した。

次に、この質的分析の結果に基づいて、主要な相談内容に横断して出現する表現に着目し、テキストマイニングによって4グループ毎の相談内容の可視化を行った。

本研究の可視化結果により、がん患者の相談内容の概要を把握し、また4グループにおける相談内容の差異を明らかにした。

がん電話相談の質的分析に基づいたテキストマイニングによる可視化

石川大介^{*1}、片山佳代子^{*2}、

^{*1} 地方独立行政法人神奈川県立病院機構 神奈川県立こども医療センター、

^{*2} 地方独立行政法人神奈川県立病院機構 神奈川県立がんセンター臨床研究所

Visualization of Cancer Telephone Consultations through Text Mining based on Qualitative Analysis

Daisuke Ishikawa^{*1}, Kayoko Katayama^{*2}

^{*1} Kanagawa Prefectural Hospital Organization, Kanagawa Children's Medical Center,

^{*2} Kanagawa Prefectural Hospital Organization, Kanagawa Cancer Center Research Institute

The Kanagawa Cancer Center offered telephone consultations for cancer between 2006 and 2014, giving about 14,000 telephone consultations in this time. However, the records of these telephone consultations are text data of notes made by the consultant about the conversation contents and are not suitable for natural language processing such as text mining. Furthermore, with such a large number of cases, reading each conversation note and conducting qualitative analysis would not be easy. Therefore, in this study, we tried visualizing the consultation contents by conducting a qualitative analysis of the results extracted through sampling the consultation data, before using those results to carry out text mining on the data as a whole.

First, we classified the telephone consultations into four groups: a) consultation by a male patient, b) consultation by a male family member, c) consultation by a female patient, d) consultation by a female family member. We conducted a qualitative analysis of the content of a total of 400 consultations by randomly extracting 100 consultations from each of these groups. Next, we focused on "worry/anxiety," which the qualitative analysis revealed had appeared in many of the consultations, and used text mining to extract the keywords related to this. We then represented the relationships between the extracted keywords in the form of a graph.

From the results of this study, we successfully visualized the "worry/anxiety" of cancer patients and clarified the differences in "worry/anxiety" among the above four groups.

Keywords: Cancer Telephone Consultation, Qualitative Analysis, Text Mining

1. はじめに

神奈川県立がんセンターでは、2006年から2014年までが
ん電話相談事業を行っており、このがん電話相談窓口には
約14000件の電話相談が寄せられた。しかし、この電話相談
の記録は、相談者との会話内容を書き留めたテキストデータ
であり、自然言語処理などのテキストマイニングに不向きなデ
ータである。また件数も多く全ての内容を解釈して質的分析
を行うことも容易ではない。

そこで本研究では、まず相談データからサンプリング抽出
した結果に対して質的分析を行い、その分析結果に基づい
て全てのデータを対象にテキストマイニングを行い、相談内
容の可視化を試みた。

以下、本研究の背景、方法、結果及び考察を述べ、最後
に今後の展開を述べる。

2. 背景

2.1 質的研究と量的研究

質的研究(質的分析)と量的研究(データマイニング、テキ
ストマイニングなど)は、これまで使われてきた研究分野およ
び研究対象が異なるため別々の手法として扱われていたが、
1990年代後半から両方の手法を取り入れた混合研究法

(mixed method research)によって両手法の統合が試みられて
きた¹⁾。2010年頃から混合研究法は専門ジャーナルが創刊さ
れ、両手法の統合方法の一般化²⁾などが研究されている。例
えば、がん告知の可否について、質的分析とテキストマイニ
ングを相互に利用する研究³⁾などが行われている。

本研究で対象とするがん電話相談のデータは、会話を書
き留めたテキストであり、そのまま自然言語処理などのテキ
ストマイニングを実施しても有用な情報を得ることは難しい。一
方、質的分析を行うためには一つ一つの相談内容を解釈し
て重要な概念を特定し、それを横断的にまとめる作業が必要
となり、1万件以上の事例を対象に行うのは困難な手法であ
る。そのため、本研究ではサンプリング抽出結果に対して質
的分析を行い、そこで得られた知見や重要な観点に着目し、
それをターゲットにテキストマイニングを行うことを試みた。

2.2 がん患者の相談支援の研究

日本において、がんの罹患率および死亡率は男性の方が
高いにもかかわらず、患者会への参加や相談支援センター
の利用は女性が圧倒的に多いことが指摘されている⁴⁾⁵⁾。また、
男性のがん経験者は、75%が誰にも相談できなかったと述べ
ていることも報告されている⁶⁾。

このように、がん相談に関する行動の男女差については従

来研究で言及されているものの、その差異の詳細まで含めた研究はまだ見当たらない。そのため、本研究では、相談内容の男女差の違いに着目し、これを可視化することを試みた。

3. 方法

以下、我々が行った研究⁷⁾の概要を述べる。

3.1 がん電話相談の概要

神奈川県立がんセンター内に設置されたがん電話相談窓口は、2006 年 10 月から 2014 年 3 月まで、匿名による電話相談を無料(通話料は別)で実施した。電話相談は、平日 10 時～12 時と 13 時～15 時の間に行い、経験豊富な看護師(3 名)による交代制で対応を行った。この電話相談の内容を Microsoft Excel で記録した 13962 件の相談事例が本研究のデータである。

3.2 電話相談データの概要

電話相談データの概要を以下に述べる。電話相談者の話者の性別は、男性からの相談が 3 割、女性からの相談が 7 割であり、女性からの相談が多い。

話者の年代は、男性話者は 60 代、70 代が多く、合計で半数以上を占める。女性話者は 40 代、50 代、60 代が多く、合計 6 割以上を占める。

話者と患者の関係では、自分自身のがんの相談が男性話者、女性話者ともに最も多く、それぞれ半数以上を占めている。次いで、男性話者は配偶者の相談が多く、女性話者は両親の相談が多い。男性話者による家族に関する相談は約 3 割、女性話者による家族に関する相談は約 4 割を占める。なお、家族以外の相談事例としては、親戚や会社関係者からの相談や、他の医療機関のソーシャルワーカーからの問い合わせなどがあつた。

以上の結果から、相談データの内訳として話者は男性と女性があり、相談の対象は本人と家族があるという大まかな傾向が伺えた。そこで、本研究では相談データをこの 4 つのグループに分けて分析を行った。

3.3 質的分析の概要

以下に本研究で実施した質的分析の概要を述べる。本研究の電話相談は、初回で相談内容の概要が語られ、2回目以降はその後の経過などの話題となる。そのため、質的分析には初回の電話相談の内容を対象とした。初回の電話相談は 10891 件あり、そのうち本人と家族からの相談事例 10534 件を質的分析の対象とした。

この 10534 件の相談事例に対して、質的分析⁸⁾を用いて相談内容の特定を行った。

まず、電話相談を話者の性別と相談の対象から、次の 4 グループに分類した。

- a. 男性話者による本人の相談
- b. 男性話者による家族の相談
- c. 女性話者による本人の相談
- d. 女性話者による家族の相談

上記 4 グループから、それぞれランダムに 100 件をサンプリング抽出した。この 4 グループから抽出した 100 件、合計 400 件のデータを対象に質的分析のコーディング作業を行った。コーディング作業の具体的な手順は下記の通り。

- 1. 電話相談のテキストを読み、相談内容に該当する部分をマーキングする。
- 2. そのマーキングした部分が表す相談内容の概要に名称(コード)を付与する。

- 3. 1～2 の作業を各グループの全 100 件に対して繰り返し行い、新たなコードが出現しなくなるまで繰り返す。
- 4. グループを横断してコードとマーキングを概観し、類似する複数のコードをより大きな概念の名称(カテゴリ)でまとめる。
- 5. カテゴリとグループによるクロス集計を行う。

上記手順によって特定された 19 の相談内容(カテゴリ)を表 1 に示す。

全てのグループで、[1]治療について、[2]施設の問い合わせが最も多かった。次いで男性本人および男性家族から[3]転院相談が多かった。次いで女性本人からは[5]進行に対する不安が多く、女性家族からは[4]医療者への不満が多かった。

質的分析の結果から、ほぼ全てのカテゴリに共通して相談内容の「不安」と「心配」が現れていた。つまり、がん電話相談は、基本的に患者が「不安」や「心配」に思う事柄を相談していると理解できる。従って、「不安」と「心配」の対象が相談内容の重要な要素であると考えた。そこで、この「不安」と「心配」の対象を特定するため、テキストマイニングを用いて「不安」と「心配」の係り受け関係となるキーワードを抽出し、可視化することを試みた。

表 1 質的分析によって得られた相談内容(カテゴリ)

[1] 治療について(方針、薬剤、治療法)
[2] 医療施設の問い合わせ
[3] 転院相談(当がんセンターの受診希望)
[4] 医療者への不信、コミュニケーション不足
[5] 進行に対する不安・心配
[6] 検査・検診(施設)の問い合わせ
[7] 症状について
[8] 副作用について
[9] 手術について
[10] 治療費、お金に関する心配、医療保険
[11] がんの疑いに関する相談
[12] 日常生活について(食事、運動、生活全般)、就労問題
[13] セカンドオピニオンについて
[14] 告知、余命宣告について
[15] 一般的な質問、情報収集
[16] 転移・再発について
[17] 精神的苦痛、心のケア
[18] 家族の心配(子供が小さい、一人暮らし、高齢の両親、など)
[19] 身体的、外見に関する相談

3.4 テキストマイニングによる抽出と可視化

相談者の相談内容の具体的な対象を把握するため、テキストマイニング⁹⁾を用いてキーワードの抽出と可視化を試みた。今回は、前述の質的分析の結果から、多くのカテゴリで出現した「不安」と「心配」に着目し、その対象のキーワードの抽出と可視化を試みた。

まず、全相談データ 13,962 件に対して、相談者本人による相談と、家族(配偶者、家族、子、親、兄弟姉妹)による相談を抽出した。抽出された 12544 件のテキストに対して日本語形態素解析ツール Mecab¹⁰⁾を利用して形態素に分解した。

次に、キーワード(主に名詞の集合)とそれ以外(動詞、助動詞、助詞など)を特定するために、名詞と接頭辞で構成さ

れた語の集合をキーワードと見なしてタグ付けを行った。ここで接頭辞をキーワード構成要素に含めた理由は、例えば「抗がん剤」などの語は「抗:接頭辞、がん:名詞、剤:名詞」というように、「抗」が接頭辞として処理されてしまうためである。

そして、このタグ付けされたキーワードとキーワードの関係を抽出した。今回は「不安」と「心配」の係り受け関係となるキーワードを抽出するため、2つのキーワード関係のうち、後方のキーワードが「不安」または「心配」である組を抽出した。抽出されたキーワード関係から重複を取り除き、目視によって前方のキーワードが「不安」及び「心配」の対象であると読み取れる関係(係り受け関係)を適合結果として抽出した。例えば、キーワード関係が係り受け関係にならないもの(否定、並列など)や、キーワードの抽出エラーなどを取り除いた。抽出したキーワード関係の件数を表2に示す。

この適合結果に対して Graphviz ツール¹¹⁾による線画を行い、可視化を試みた。今回は、ばねモデルに基づく線画方法¹²⁾ fdp を使い、エッジの線画がノードに重ならないよう曲線で描く splines オプションを用いた。また、今回は4グループにおける「不安」あるいは「心配」の対象となるキーワード関係を全て統合した上でグラフを生成した。生成したキーワードグラフを図1、2に示す。

表2 キーワード関係の抽出結果の件数

	男性 本人	男性 家族	女性 本人	女性 家族
テキストマイニング対象のテキスト件数	2202	1482	4855	4020
「不安」の対象となるキーワード関係:				
抽出件数	80	26	341	136
重複を除いた件数	59	25	198	102
適合件数	23	12	107	39
「心配」の対象となるキーワード関係:				
抽出件数	124	74	338	228
重複を除いた件数	85	55	227	149
適合件数	38	26	137	73

4 結果と考察

まずキーワード関係の抽出件数について述べる。表2で示したように、抽出件数は女性本人 > 女性家族 > 男性本人 > 男性家族の順番で多かった。しかしながら、最終的な適合件数の対象テキスト数による割合で考えると、女性本人だけが他グループよりも倍近く多く、他グループの割合は同程度であった。「不安」は他グループが約 1%に対して女性本人は 2.2%、「心配」は他グループが約 1.8%に対して女性本人は 2.8%であった。以上のことから、女性本人による「不安」や「心配」の対象が他のグループよりも多いことが分かった。

次に、生成したキーワードグラフについて述べる。図1,2に描かれたグラフでは、前述のように女性本人によるノードが多数を占めていることが分かる。例えば「インターネット」のノードは「インターネットで調べると不安」という記述であり、他にも「ストレス」「個人情報」「誤診」など、女性本人の「不安」や「心配」の対象は病気以外にも及んでいることが分かった。その他、男性家族からは「妻」「弟」のノードがあるのに対して、女性家族からは「義兄」「義妹」などのノードが見られ、家族における男女の役割の差が表れていた。男性本人からは、例えば

「体力低下」のノードが見られた。なお、「不安」と「心配」のグラフを比較すると、「不安」の方が漠然としたキーワードが多く、「心配」の方が具体的なキーワードが多い傾向にあった。

5 おわりに

本稿では、神奈川県立がんセンターで実施されたがん電話相談の内容について、質的分析を行った結果に基づいてがん患者の不安・心配の可視化を行った。今後の展開として、質的分析においては対象範囲を更に広げて複数人によるコーディング作業を行う方向や、テキストマイニングにおいては「不安・心配」以外のキーワード(例えば「不信・不満」など)に基づく係り受け関係の抽出、可視化においてはばねモデル以外の線画方法の検討など、更なる研究の余地がある。

なお本研究の成果は、日本がん登録協議会と全国がん患者団体連合会の共同プロジェクト J-CIP (Japan Cancer Information Partnership)¹³⁾の情報提供の準備に活用されている。

参考文献

1) 抱井尚子. 混合研究法入門: 質と量による統合のアート. 医学書院, 2015.

2) Creswell, John W., and Vicki L. Plano Clark. Designing and Conducting Mixed Methods Research. SAGE, 2011.

3) 稲葉光行, 抱井尚子. 質的データ分析におけるグラウンデッドなテキストマイニング・アプローチの提案—がん告知の可否をめぐるフォーカスグループでの議論の分析から—. 政策科学, 2011: 18.3: 255-276.

4) 町田朱美. がん相談支援センターにおけるがん相談の現状と課題. 日本看護学会論文集. 看護総合, 2012, 42: 238-240.

5) 上田育子, 他. 当院がん患者サロン「ひだまり」の評価と課題—サロンの現状と参加者へのアンケート調査から—. 癌と化学療法, 2013, 40.9: 1195-1200.

6) HOPE プロジェクト. 粒子家族時代のがん暮らし・生活ニーズ調査. 2013. http://kibou.jp/images/20130710_SingleLife_can.pdf. (cited 2019-Aug-23)

7) Kayoko Katayama, Daisuke Ishikawa, Yohei Miyagi, Shoji Takemiya, Naoyuki Okamoto and Asao Ogawa. Qualitative analysis of cancer telephone consultations: Differences in the counseling needs of Japanese men and women (submitted to Patient Education and Counseling)

8) Glaser, Barney G., Anselm L. Strauss, and Elizabeth Strutzel. The discovery of grounded theory; strategies for qualitative research. Nursing research 17.4 (1968): 364.

9) Hearst, Marti A. Untangling text data mining. Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics. Association for Computational Linguistics, 1999.

10) MeCab: Yet Another Part-of-Speech and Morphological Analyzer. <http://taku910.github.io/mecab/>

11) Graphviz: Graph Visualization Software. <https://www.graphviz.org/>

12) Fruchterman, Thomas MJ, and Edward M. Reingold. Graph drawing by force-directed placement. Software: Practice and experience 21.11 (1991): 1129-1164.

13) J-CIP: Japan Cancer Information Partnership. <http://jacr.info/j-cip/>

