

厚生労働科学研究費補助金 (がん対策推進総合研究事業)
(分担研究報告書)

がん相談支援内容の分析と分類に関する研究

研究分担者 内村祐之 東京医科歯科大学病院医療情報部 (特任講師)

研究要旨

がん相談支援で過去の相談記録情報を利用したテキストマイニング技術により疾患別やカテゴリー別の傾向を分析し可視化する為、2020年度からプロトタイプの開発を進めてきた。2021年度には単語の出現頻度によるビジュアル化や単語を数値ベクトルに変換してその意味を把握する自然言語処理の手法であるWord2Vecを用い、過去の相談内容から類似度の高い順に類似相談内容を抽出する機能の追加とWebアプリケーション化を実現しがん相談支援を行っている施設での実証試験を開始した。2022年度は研究班の意見を反映し、除外ワードに加えて品詞別除外機能を追加し、出力された結果が表形式で出力されるように変更を行い、下位の出現頻度のワードも出力するようにした。また、除外品詞やワードの再調整を行いやすいように品詞も出力するように変更を行ない、引き続き実証を行なった。

A. 研究目的

現在、全国のがん相談支援センターでは、一定の基準を満たした上で治療の内容や設備、がんに関する情報提供などを行っており、がん相談支援における相談内容をデータとして管理している。しかし、蓄積されたデータの利活用や相談員間の知識共有は十分に出来ていないのが現状である。将来の相談内容の更なる多様化に備えるためには、蓄積された相談情報を活用し、FAQの作成や相談員間の情報交換など他施設を含めた相談員のノウハウを共有する仕組みを作り、相談員による経験や知識に起因する対応の差を埋めることにより、一定水準以上の相談対応が行えるような相談対応支援の仕組みの構築が必要であると考えられる。仕組みの構築に向けて、がん相談支援で過去に対応した相談記録情報を利用し、テキストマイニング技術により疾患別やカテゴリー別の傾向を分析し可視化する。さらに、現状の相談内容や個々の相談員の対応の傾向を把握した上で、相談内容と対応内容を分類することで、FAQの作成支援や他施設を含めた情報共有の仕組みを検討し、自施設の相談支援業務に活かすことを目的とする。これまでのWebアプリケーションのプロトタイプに更なる改良を加え、実証試験にてシステムの検証を行う事を目的とする。

B. 研究方法

全国の「がん相談支援センター」の相談支援内容は施設により、がん相談支援センター「相談のため

の基本形式」(https://ganjoho.jp/med_pro/consultation/support/registration_sheet.html)の形式を基として、相談記録の要旨が電子カルテシステムまたはExcelのシート、FileMakerなどの独自開発システムなどに相談員が入力したデータとして記録されている。この記録を収集し、計量テキスト分析とテキストマイニングにより分析および可視化、内容による分類を行う。2020年度は10件の相談記録「形態素解析による分かち書きで単語を集計」「係り受け解析」「共起ネットワークでビジュアル化」の3種のプロトタイプをプログラミング言語Pythonにより作成した。2021年度はプロトタイプに改良を加え、単語の出現頻度によるビジュアル化(Word Cloud)を行い、単語を数値ベクトルに変換してその意味を把握する自然言語処理の手法であるWord2Vecを用い、過去の相談内容から類似度の高い順に類似相談内容を抽出する機能を実現した。更にプロトタイプシステムをWebアプリケーションとしてブラウザのみで動くように改良を行い、がん相談支援を行っている施設での実証試験を開始した。2022年度は研究班の意見を反映し、除外ワードに加えて品詞別除外機能を追加し、出力された結果が表形式で出力されるように変更を行い、下位の出現頻度のワードも出力するようにした。また、除外品詞やワードの再調整を行いやすいように品詞も出力するように変更を行ない、引き続き実証を行なった。

(倫理面への配慮)

データは全て実データを元にしたサンプルデータを用いた。実証試験では実証試験施設である群馬大学医学部附属病院の倫理審査委員会の承認(テキストマイニングによるがん相談支援内容の分析と分類に関する研究)を経て、倫理規定等を遵守しながら進めている。

C. 研究結果

1. 作成したWebアプリケーション概要と利用手順

Webアプリケーションはプログラミング言語「Python」のFlaskフレームワークで作成し、形態素解析エンジンには「Mecab」を使用した。Webアプリケーションは、ブラウザを開き特定のURLを入力する事で利用が可能となるが、利用の手順は以下の通り。

「1. ファイルの選択」で「ファイルを選択」ボタンを押して読み込むファイルを選択する。読み込むファイルには「相談内容」「対応内容」の列を必ず含める。

「2. 頻出ワードの取得個数」で頻出ワードを数える個数を入力する。(デフォルト100)

「3. 除外ワードの設定」で解析から除外するワードをカンマ区切り(または、)で入力する。

「4. 類似相談内容の抽出」で抽出キーワードをカンマ区切り(または、)で入力する。また、類似相談を抽出する対象の文章を入力する。

最後に「解析」ボタンをクリックすると解析処理が開始される。

「がん相談支援内容の分析と分類に関する研究について」評価ツール

Version: 1.1

1. ファイルの選択

2. 頻出ワードの取得個数

3. 除外品詞・ワードの設定

4. 類似相談内容の抽出

5. 解析実行

2. 分析結果の表示

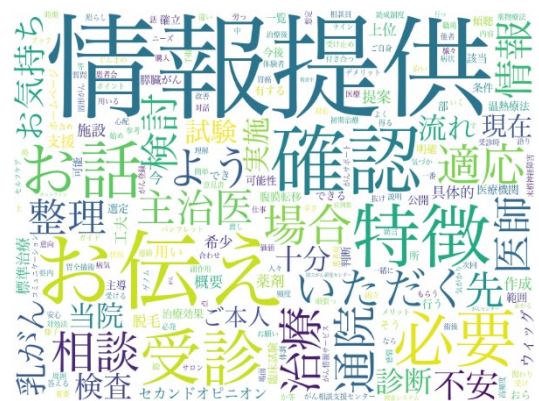
2-1) がん相談内容、対応内容のビジュアル化

がん相談内容と対応内容に関するデータを元に頻出語を頻度に比例する大きさと並べたWord Cloudによるビジュアル化を行なった。

① 相談内容のビジュアル化



② 対応内容のビジュアル化



③ 単語の出現回数の可視化(上位、下位)

解析結果

頻出ワードの取得個数: 20

除外品詞: 助詞・接頭詞・連帯詞・記号・フィラー・その他

除外ワード:

1. 記録内容の頻出ワード分析

【相談内容 上位20ワード】

ワード	件数
[動詞] いる	14
[動詞] あり	13
[助動詞] た	9
[助動詞] し	9
[名詞] 治療	8
[助動詞] ない	8
[名詞] の	7
[名詞] 相談	7
[助動詞] れ	7
[名詞] 通院中	6
[名詞] 電話相談	5
[助動詞] 受け	5
[名詞] 当院	5
[形容詞] よい	4
[助動詞] たら	4
[名詞] がん患者	4
[名詞] 他院	4
[動詞] 言わ	4
[名詞] どこ	3
[助動詞] で	3

2. 学会発表

なし

H.知的財産権の出願・登録状況

(予定を含む)

1.特許取得

なし

2.実用新案登録

なし

3.その他

なし

