

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。

以下の下位ワード表示は2022年度に追加した新機能である。

【相談内容 下位20ワード】	
ワード	件数
【助動詞】で	3
【名詞】どこ	3
【動詞】言わ	4
【名詞】他院	4
【名詞】がん患者	4
【助動詞】たら	4
【形容詞】よい	4
【名詞】当院	5
【動詞】受け	5
【名詞】電話相談	5
【名詞】通院中	6
【動詞】れ	7
【名詞】相談	7
【名詞】の	7
【助動詞】ない	8
【名詞】治療	8
【動詞】し	9
【助動詞】た	9
【動詞】あり	13
【動詞】いる	14

2-2) Word2vecによる類似文章検索

単語を数値ベクトルに変換し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。抽出結果を以下に示す。

3. 類似相談内容の抽出				
キーワード: がん、治療				
検索対象内容:				
検索から出てきた内容中、特に興味をかけたくないため、がん治療をしてほしいと書いている。主治医には、治療は手術になると書かれたので、がん治療は担当しないと言われる。ちゃんと話を聞いてもらえ先生に相談したい、どうしたらその治療が受けられるのか。				
順位	類似スコア	キーワードスコア	類似内容	対応内容
1	0.9528361	15	相談内容: 相談内容: 特に興味をかけたくないため、がん治療をしてほしいと書いている。主治医には、治療は手術になると書かれたので、がん治療は担当しないと言われる。ちゃんと話を聞いてもらえ先生に相談したい、どうしたらその治療が受けられるのか。	ご本人の相談と病状理解を踏まえて、状況を確認。予後について手紙を書き送られている。がん治療サービスの提供が目的の患者と、患者様ご自身の意思で、病状の改善と手術療法との選択について情報提供。相談内容については、医師の経験から7-10歳の子供から、標準治療の選択肢となること、がん治療の重要性を把握して説明した。病状の経過や体力面への配慮、また治療方針の決定手前であり、疑問点も主治医に確認することも提案。医師への質問の仕方や工夫についても説明した。今後も相談を継続し、引き続き相談可能であることを伝えました。
2	0.9469219	7	相談内容: 相談内容: 特に興味をかけたくないため、がん治療をしてほしいと書いている。主治医には、治療は手術になると書かれたので、がん治療は担当しないと言われる。ちゃんと話を聞いてもらえ先生に相談したい、どうしたらその治療が受けられるのか。	がん治療の適応ではないことを説明。説明について医師とよく相談していただく。
3	0.9133366	11	相談内容: 相談内容: 特に興味をかけたくないため、がん治療をしてほしいと書いている。主治医には、治療は手術になると書かれたので、がん治療は担当しないと言われる。ちゃんと話を聞いてもらえ先生に相談したい、どうしたらその治療が受けられるのか。	まだ情報不足であり、乳癌の検査の進め方と治療についてがん治療サービスの提供を主として情報提供。体への負担を軽減したいというご本人の気持ちを受け止めつつ、まずはしっかりと再検と方針を確定いただくようお願いした。がん治療サービスについて、がん治療サービスの内容を一緒に確認しながら、疑問と確認について説明。医師の経験、標準治療の選択の仕方についても伝えました。
4	0.6840123	8	相談内容: 相談内容: 特に興味をかけたくないため、がん治療をしてほしいと書いている。主治医には、治療は手術になると書かれたので、がん治療は担当しないと言われる。ちゃんと話を聞いてもらえ先生に相談したい、どうしたらその治療が受けられるのか。	現在の病状の経過と治療方針の相談内容を踏まえて、がん治療サービスの提供と手術療法について、患者様の情報提供。遺伝子検査の結果ががん治療方針の決定に与える影響について、医師が可能な範囲で説明し、相談が可能であることを伝えました。
5	0.65922004	4	相談内容: 相談内容: 特に興味をかけたくないため、がん治療をしてほしいと書いている。主治医には、治療は手術になると書かれたので、がん治療は担当しないと言われる。ちゃんと話を聞いてもらえ先生に相談したい、どうしたらその治療が受けられるのか。	医師の経験と病状理解を踏まえて情報提供。ご本人の希望を踏まえて、ホームページの情報から医師の経験と費用について確認し情報提供した。標準治療の提供、相談支援センターについて情報提供し、相談が可能であることを伝えました。

D. 考察

これまでががん相談に関しては、過去の相談データを解析するなどのデータ利活用はあまり進んでいなかった。今回作成したWordCloudによるビジュアル化と形態素解析による分かち書きによる単語集計により傾向を分析し可視化することで相談内容の重要なキーワードを直観的に把握することができ、経時的にトレンドを追う事で自施設の相談支援業務に活かす事が可能になると考える。また、W

ord2vecによる類似文章検索では、相談内容の文章を入力する事で、過去の相談から類似度の高い順に内容を一覽で抽出する事が出来、キーワードのハイライトを合わせる事で、類似の相談内容に対してどのような対応を行ったのかを把握する事も出来るので、相談内容と対応内容を分類することで、FAQの作成支援につなげることが出来る事が示唆された。2022年度は研究班や実証施設の意見を反映し、システムの機能を追加したが、これまで相談内容や対応内容の頻出ワードを上位から表示するのみであったが、頻度が少ない=新しいトピックであり、情報が整備されていない可能性があるため頻度の低い単語にも注目する必要があることから頻度の少ない情報から重要なワードをキャッチしやすいように出力結果を変更した。また、分析の際に登録する除外ワードはこれまで個別に入力を行っていたが、品詞別一括で登録できるように修正を行い、今後、多施設共同で解析を行う際に、共通認識の基で作業を進められるような工夫を行なった。

E. 結論

2021年度はプロトタイプを改良し、単語の出現頻度によるビジュアル化を行い、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。今後は実際のがん相談支援記録を用いてテキストマイニングを行う事で、相談内容および対応内容の分類やFAQの作成支援へと発展させる予定である。

F. 健康危険情報

特になし

G. 研究発表

(発表誌名巻号・頁・発行年等も記入)

1. 論文発表 なし

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