

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 東京医科歯科大学医学部附属病院 医療情報部 (特任講師)  
研究分担者 早川 雅代 国立がん研究センターがん対策研究所がん情報提供部 (室長)

研究要旨

がん相談支援で過去に対応した相談記録情報を利用したテキストマイニング技術による疾患別やカテゴリー別の傾向を分析し可視化する為、2020年度は自由記載で書かれた相談内容や対応内容から単語を集計し相談内容と対応内容に含まれる単語間の繋がりを可視化するプロトタイプを開発した。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。

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。昨年度作成したプロトタイプシステムに改良を加え、実証試験にてシステムの検証を行う。

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](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Word 2Vec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。

(倫理面への配慮)

プロトタイプ開発の段階ではデータは全て実データと異なる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. WordCloudによるビジュアル化と形態素解析による分かち書きで単語を集計  
がん相談内容と対応内容に関するサンプルデー

タを元に頻出語を頻度に比例する大きさと文字を並べたWordCloudによるビジュアル化と単語の出現回数を可視化した。可視化の例を以下に示す。

1-1 がん相談内容のビジュアル化



1-2 がん相談内容の分かれ書きによる単語集計

単語	出現回数
治療	8
相談	7
通院中	6
電話相談	5
受け	5
当院	5
がん患者	4
他院	4
言わ	4

1-3 がん相談内容に対する対応のビジュアル化



1-4 がん相談内容に対する対応の分かれ書きによる単語集計

単語	出現回数
情報提供	15
お伝え	8
確認	6
特徴	5
お話	4
必要	4
受診	4
いただく	4
よう	4
相談	4
適応	4
通院	4
場合	4
治療	4

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。サンプルデータによる抽出結果を以下に示す。

元の文章 (国がんサンプル)	全10データ中最も類似度が高かった文章 (国がんサンプル)	類似度
子宮体部の癌性疾患が再発した。都内で大病院のある病院を教えて欲しい。	他院通院中の大腸がん患者より相談あり。手術と術後の化学療法を行い、現在は経過観察をしている。今後の影響でまた手足の痺れが残っており、なんとか仕事をしている状態。同じような経験をしている人の話が聞きたいが、そういう場所がないかとのご相談。	0.7742452
3年間に乳がんステージIIICで術前化学療法、手術、放射線治療。現在タモキシフェンを内服している。1ヶ月前に頸椎の骨粗鬆症が診断された。今度の診察まで頸椎骨折で通院しているが、固定すると手足が痺れるなど副作用が起きている。病院に通院し続けられるか不安。	乳癌患者より電話相談あり。浸潤性乳がんの診断を受け、今後手術予定。術後の治療が必要と言われている。本日に手術をしようと思っているが、今から通院の準備だけだと判断してよいか迷っている。治療が始まる前に他の意見も聞きたいと思ったが、どうしたらよいかとの相談。	0.5845203
遺伝性潰瘍症があり内視鏡を受けたらポリープがあり、一部生検するといわれて。がんの可能性をおとすと不安で仕方ない。妻がもうすぐ出産なので産にも言えず一人である。ローンは始めるか、がん検査は入れるか。	他院通院中の膵臓がん患者の家族より電話で問い合わせあり。膵臓、膵臓がんと診断されたが、転移は転移があり手術はできないと言われた。今後は化学療法予定。遠方では施設内投与はやっていないと言われたが、当院で施設内投与ができるかとの質問あり。	0.63119334
癌が未発前、昨年8月からキートルと抗がん剤で治療していたが、副作用が続き、入院治療も現在は安静で治療はできず、急いで退院しケアがホスピスを受診する。本日に手術を出ないといけないが、父はこれを拒絶することができない。	他院通院中の膵臓がん患者の家族より電話で問い合わせあり。膵臓、膵臓がんと診断されたが、転移は転移があり手術はできないと言われた。今後は化学療法予定。遠方では施設内投与はやっていないと言われたが、当院で施設内投与ができるかとの質問あり。	0.7210703
2年前に中程度の膵臓がんで放射線を希望してやったがPSAが上がってきたので手術。(説明がわからぬ) 前立腺癌に近い部分の一部に転移があったとのこと。来月のPSAでホルモン療法を行うのがめんどい。得たことがなかったが、進行してしまいが、とても心配。	乳癌患者より電話相談あり。浸潤性乳がんの診断を受け、今後手術予定。術後の治療が必要と言われている。本日に手術をしようと思っているが、今から通院の準備だけだと判断してよいか迷っている。治療が始まる前に他の意見も聞きたいと思ったが、どうしたらよいかとの相談。	0.57097757
元の文章 (国がんサンプル)	【単語一語】最も類似スコアが高かった文章 (国がんサンプル)	【doc2vec】最も類似スコアが高かった文章 (国がんサンプル)
骨質が薄くなった原因が腎臓か、膵臓か、診断と検査が必要で、お話を聞いて、手術と化学療法が必要か、どうしたらよいか。息と息は通っているが、検査の結果は詳しくは分からないので、詳しくお話を聞きたい。当院で施設内投与はできるか、どうしたらよいかとの相談。	乳癌患者より電話相談あり。浸潤性乳がんの診断を受け、今後手術予定。術後の治療が必要と言われている。本日に手術をしようと思っているが、今から通院の準備だけだと判断してよいか迷っている。治療が始まる前に他の意見も聞きたいと思ったが、どうしたらよいかとの相談。	他院通院中の膵臓がん患者の家族より電話で問い合わせあり。膵臓、膵臓がんと診断されたが、転移は転移があり手術はできないと言われた。今後は化学療法予定。遠方では施設内投与はやっていないと言われたが、当院で施設内投与ができるかとの質問あり。
友人がE型肝炎と診断された。慢性肝炎の検査を受けて、今後手術予定。術後の治療が必要と言われている。本日に手術をしようと思っているが、今から通院の準備だけだと判断してよいか迷っている。治療が始まる前に他の意見も聞きたいと思ったが、どうしたらよいかとの相談。	乳癌患者より電話相談あり。浸潤性乳がんの診断を受け、今後手術予定。術後の治療が必要と言われている。本日に手術をしようと思っているが、今から通院の準備だけだと判断してよいか迷っている。治療が始まる前に他の意見も聞きたいと思ったが、どうしたらよいかとの相談。	再発治療中の膵臓がん患者より電話相談あり。手術は手術は必要と言われているが、転移の心配がなく、経過もよいため別の病院に転院して手術を受けたいと言っている。転院先の病院ががん専門の病院かどうかが気になっている。どうしたらよいかとの相談。
父親が膵臓がんの診断がある。現在15日入院中、検査は完了した。膵臓がんの診断と検査の結果は詳しくは分からないので、詳しくお話を聞きたい。当院で施設内投与はできるか、どうしたらよいかとの相談。	他院通院中の膵臓がん患者の家族より電話で問い合わせあり。膵臓、膵臓がんと診断されたが、転移は転移があり手術はできないと言われた。今後は化学療法予定。遠方では施設内投与はやっていないと言われたが、当院で施設内投与ができるかとの質問あり。	他院通院中の膵臓がん患者より電話相談あり。手術は手術は必要と言われているが、転移の心配がなく、経過もよいため別の病院に転院して手術を受けたいと言っている。転院先の病院ががん専門の病院かどうかが気になっている。どうしたらよいかとの相談。
母が乳がんステージIIICで術前化学療法、手術、放射線治療を受けた。その後、骨粗鬆症が診断された。今度の診察まで頸椎骨折で通院しているが、固定すると手足が痺れるなど副作用が起きている。病院に通院し続けられるか不安。	乳癌患者より電話相談あり。浸潤性乳がんの診断を受け、今後手術予定。術後の治療が必要と言われている。本日に手術をしようと思っているが、今から通院の準備だけだと判断してよいか迷っている。治療が始まる前に他の意見も聞きたいと思ったが、どうしたらよいかとの相談。	他院通院中の膵臓がん患者の家族より電話で問い合わせあり。膵臓、膵臓がんと診断されたが、転移は転移があり手術はできないと言われた。今後は化学療法予定。遠方では施設内投与はやっていないと言われたが、当院で施設内投与ができるかとの質問あり。
父がE型肝炎と診断された。慢性肝炎の検査を受けて、今後手術予定。術後の治療が必要と言われている。本日に手術をしようと思っているが、今から通院の準備だけだと判断してよいか迷っている。治療が始まる前に他の意見も聞きたいと思ったが、どうしたらよいかとの相談。	乳癌患者より電話相談あり。浸潤性乳がんの診断を受け、今後手術予定。術後の治療が必要と言われている。本日に手術をしようと思っているが、今から通院の準備だけだと判断してよいか迷っている。治療が始まる前に他の意見も聞きたいと思ったが、どうしたらよいかとの相談。	他院通院中の膵臓がん患者の家族より電話で問い合わせあり。膵臓、膵臓がんと診断されたが、転移は転移があり手術はできないと言われた。今後は化学療法予定。遠方では施設内投与はやっていないと言われたが、当院で施設内投与ができるかとの質問あり。

### 3. プロトタイプWebアプリケーション化と実証試験の開始

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。手順及び画面遷移を以下に示す。

#### 1. ブラウザの起動と分析情報の入力

#### 「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

- 1. ファイルの選択**  
「ファイルを選択」ボタンをクリックし、読み込むファイルを選択して下さい。  
 シンプル、分析 csv
- 2. 頻出ワードの取得個数**  
頻出ワードの取得個数を入力して下さい。  
 個
- 3. 除外ワードの設定**  
前年から除外するワードをカンマ区切りで入力して下さい。
- 4. 類似相談内容の抽出**  
抽出するキーワードをカンマ区切りで入力して下さい。  
  
類似相談を抽出する対象の文章を入力して下さい。(なるべく長い文章を入力)  
  
類似相談に関する情報は多くありませんが、ゲノム検査をしてほしいと考えている、または、治療は手術になると思われるがゲノム検査は必要なのかと質問する、どちらにも意見を載せてもらってほしいとお願いしたい、
- 5. 解析実行**  
「解析」ボタンをクリックし結果が表示されるまでそのままお待ち下さい。(数分かかる場合があります)

#### 2. 解析の実行

#### 解析結果

頻出ワードの取得個数: 100個

除外ワード: いる,あり,され,の,しい,られ,よい,なる,あり,ある,いる,する,ない,れる,ため,こと,もの

- 1. 記録内容の頻出ワード分析**  
【相談内容 上位100ワード】  
[「治療」, 19], [「ゲノム」, 11], [「医療」, 10], [「中」, 9], [「受け」, 9], [「相談」, 8], [「言わ」, 8], [「患者」, 7], [「院」, 6], [「通院」, 6], [「化学療法」, 5], [「後」, 5], [「手術」, 5], [「たけ」, 5], [「状況」, 5], [「話」, 5], [「聞い」, 5], [「もろえる」, 5], [「先生」, 5], [「これ」, 5], [「確し」, 5], [「効果」, 5], [「卵巣がん」, 5], [「転移」, 5], [「転」, 5], [「良体」, 5], [「的」, 5], [「治療」, 5], [「パネル」, 5], [「後」, 5], [「コース」, 5], [「取り上げ」, 5], [「がん」, 5], [「薬」, 5], [「受け」, 5], [「費用」, 5], [「流れ」, 5], [「実態」, 5], [「代」, 5], [「相」, 5], [「遺伝」, 5], [「腫瘍」, 5], [「可能性」, 5], [「説明」, 5], [「結果」, 5], [「受け止め」, 5], [「いけ」, 5], [「経過」, 5], [「最近」, 5], [「体調不良」, 5]  
【対応内容 上位100ワード】  
[「治療」, 16], [「検査」, 11], [「情報提供」, 10], [「確認」, 10], [「相談」, 9], [「ゲノム」, 8], [「説明」, 8], [「医療」, 7], [「本人」, 7], [「実態」, 6], [「容」, 6], [「標準」, 6], [「後」, 6], [「パネル」, 6], [「提案」, 6], [「選択肢」, 6], [「場合」, 6], [「期」, 6], [「用い」, 6], [「医師」, 6], [「冊子」, 6], [「形」, 6], [「終了」, 6], [「患者」, 6], [「用」, 6], [「ガイドライン」, 6], [「適応」, 6], [「状況」, 6], [「子宮体がん」, 6], [「手術」, 6], [「特徴」, 6], [「一緒に」, 6], [「臓器」, 6], [「いく」, 6], [「遺伝子」, 6], [「部」, 6], [「院」, 6], [「でき」, 6], [「伝え」, 6], [「見込み」, 6], [「条件」, 6], [「代」, 6], [「別」, 6], [「薬法」, 6], [「位置づけ」, 6], [「ページ」, 6], [「術後」, 6], [「経過」, 6], [「画」, 6], [「心配」, 6], [「来週」, 6], [「予定」, 6]
- 2. WordCloudによるビジュアル化**  
【相談内容】

順位	類似スコア	キーワード	類似内容	対応内容
1	0.9629361	15	相談内容として治療中、検査結果待ちで悩んでいる、ゲノム検査をしてほしいと聞いている、手術内容は、治療は手術になると思われるが、ゲノム検査は必要なのかと聞かれる、中から治療内容を選択し、	ゲノムの検査結果は検査結果が出るまで、検査内容、手術内容、検査結果を待ちます。検査結果がでたら医師と相談してください。検査結果がでたらゲノム検査の結果を医師と相談してください。検査結果がでたらゲノム検査の結果を医師と相談してください。検査結果がでたらゲノム検査の結果を医師と相談してください。
2	0.7446679	7	相談内容として検査中、検査結果待ちで悩んでいる、ゲノム検査をしてほしいと聞いている、手術内容は、治療は手術になると思われるが、ゲノム検査は必要なのかと聞かれる、中から治療内容を選択し、	ゲノムの検査結果は検査結果が出るまで、検査内容、手術内容、検査結果を待ちます。検査結果がでたら医師と相談してください。検査結果がでたらゲノム検査の結果を医師と相談してください。検査結果がでたらゲノム検査の結果を医師と相談してください。検査結果がでたらゲノム検査の結果を医師と相談してください。
3	0.83326	11	相談内容として検査中、検査結果待ちで悩んでいる、ゲノム検査をしてほしいと聞いている、手術内容は、治療は手術になると思われるが、ゲノム検査は必要なのかと聞かれる、中から治療内容を選択し、	ゲノムの検査結果は検査結果が出るまで、検査内容、手術内容、検査結果を待ちます。検査結果がでたら医師と相談してください。検査結果がでたらゲノム検査の結果を医師と相談してください。検査結果がでたらゲノム検査の結果を医師と相談してください。検査結果がでたらゲノム検査の結果を医師と相談してください。
4	0.684713	8	相談内容として検査中、検査結果待ちで悩んでいる、ゲノム検査をしてほしいと聞いている、手術内容は、治療は手術になると思われるが、ゲノム検査は必要なのかと聞かれる、中から治療内容を選択し、	ゲノムの検査結果は検査結果が出るまで、検査内容、手術内容、検査結果を待ちます。検査結果がでたら医師と相談してください。検査結果がでたらゲノム検査の結果を医師と相談してください。検査結果がでたらゲノム検査の結果を医師と相談してください。検査結果がでたらゲノム検査の結果を医師と相談してください。
5	0.6592204	4	相談内容として検査中、検査結果待ちで悩んでいる、ゲノム検査をしてほしいと聞いている、手術内容は、治療は手術になると思われるが、ゲノム検査は必要なのかと聞かれる、中から治療内容を選択し、	ゲノムの検査結果は検査結果が出るまで、検査内容、手術内容、検査結果を待ちます。検査結果がでたら医師と相談してください。検査結果がでたらゲノム検査の結果を医師と相談してください。検査結果がでたらゲノム検査の結果を医師と相談してください。検査結果がでたらゲノム検査の結果を医師と相談してください。

#### D. 考察

これまで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、Word2vec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覧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。次年度は実証試験の結果を分析し、他施設を含めた情報共有の仕組みへの活用を検討していく予定である。

#### 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。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ていく予定である。

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