

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情報センターがん情報提供部 (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10件の相談記録の要旨のサンプルデータを作成し、「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」「共起ネットワークによる可視化」の3種のプロトタイプ作成を行った。その結果、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、係り受け解析により単語間の関係性が示し、共起ネットワークによるビジュアル化により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ことができた。

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。

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件の相談記録の要旨のサンプルデータ(平均文字数 相談内容:111文字、対応内容:73文字)をCSV形式で作成し、「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。

C. 研究結果

1. 形態素解析による分かち書きで単語を集計

10件のサンプルデータにおける単語数の集計結果は以下の通りとなった。

① 相談内容に多く含まれる単語数

順位	単語	件数
1	がん	10
2	病院	9
3	治療	6
4	療法	6
5	転移	5
6	副作用	5
7	入院	3
8	医師	3

② 対応内容に多く含まれる単語数

順位	単語	件数
1	相談	7
2	がん	5
3	状況	5
4	検索	4
5	医師	4
6	情報	4
7	医療	3
8	提供	3

2. 係り受け解析

10件のサンプルデータにおける係り受け解析の結果の一部を以下に示す。

① 相談内容の係り受け解析（ツリー構造出力）

子宮体部の-D
 絨毛性疾患が-D
 再発した。-----D
 都内で-----D
 実績の-D |
 ある-D |
 病院を-D
 教えて欲しい。

② 相談内容の係り受け解析（解析処理出力）

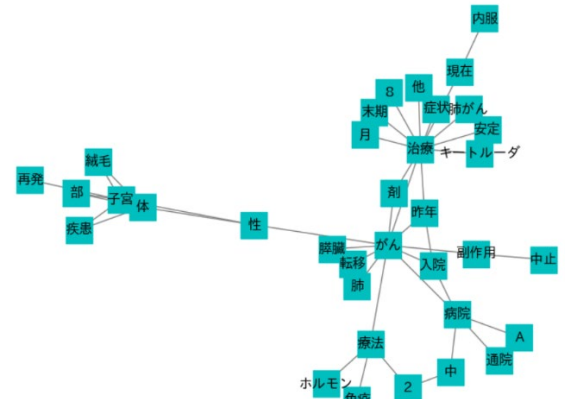
* 0 1D 2/3 2.122599
 子宮 名詞,一般,*,*,*,子宮,シキユウ
 体 名詞,接尾,一般,*,*,*,体,タイ
 部 名詞,接尾,一般,*,*,*,部,ブ
 の 助詞,連体化,*,*,*,の,ノ
 * 1 2D 2/3 2.687666
 絨毛 名詞,一般,*,*,*,絨毛,ジュウモウ
 性 名詞,接尾,一般,*,*,*,性,セイ
 疾患 名詞,一般,*,*,*,疾患,シッカ
 ン
 が 助詞,格助詞,一般,*,*,*,が
 * 2 7D 1/2 -1.598281
 再発 名詞,サ変接続,*,*,*,再発,サイハツ
 し 動詞,自立,*,*,サ変・スル,連用形,する,シ
 た 助動詞,*,*,*,特殊・タ,基本形,た,タ
 。 記号,句点,*,*,*,。 ,。 。
 * 3 7D 0/1 -1.598281
 都内 名詞,一般,*,*,*,都内,トナイ
 で 助詞,格助詞,一般,*,*,*,で,デ
 * 4 5D 0/1 1.291034
 実績 名詞,一般,*,*,*,実績,ジッセキ
 の 助詞,格助詞,一般,*,*,*,の,ノ
 * 5 6D 0/0 2.146836
 ある 動詞,自立,*,*,五段・ラ行,基本形,ある,アル
 * 6 7D 0/1 -1.598281
 病院 名詞,一般,*,*,*,病院,ビョウイン
 を 助詞,格助詞,一般,*,*,*,を,ヲ
 * 7 -1D 0/2 0.000000
 教え 動詞,自立,*,*,一段,連用形,教える,オシ
 エ
 て 助詞,接続助詞,*,*,*,て,テ
 欲しい 形容詞,非自立,*,*,形容詞・イ段,基本形,
 欲しい,ホシイ
 。 記号,句点,*,*,*,。 ,。 。

③ 相談内容の係り受け解析（文節毎出力）

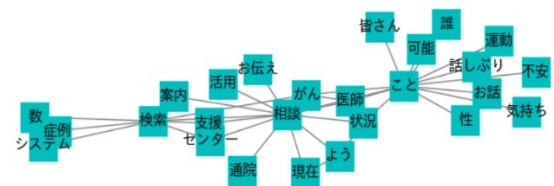
子宮体部 => 絨毛性疾患
 絨毛性疾患 => 再発する
 再発する => 教える
 都内 => 教える
 実績 => ある
 ある => 病院
 病院 => 教える

2. 共起ネットワークによる可視化

① 相談内容



② 対応内容



D. 考察

10件の相談記録情報サンプルデータを用いたテキストマイニングにおいて、形態素解析による単語集計では、相談内容の上位には「がん」「病院」「治療」「療法」「転移」「副作用」など相談者の抱える疾患や治療法、病院などの相談内容のうち最も伝えたい単語が抽出され、対応内容では「相談」「状況」「検索」「情報」「提供」など、相談内容に対する対応や情報提供に関連した単語の数が多く、単語数の修正からある程度の傾向が把握できることが示された。また、係り受け解析および共起ネットワークによる可視化では単語間の繋がりを表現することで、自由記載の相談内容のうち、最も伝えたい文の要約を抽出する事が出来ることが示唆された。今後は、がん相談支援センター「相談のための基本形式」で記録される「がんの部位、相談内容の分類、相談対応」などの情報とともに、分析を行うことにより、FAQの作成支援となる基礎データの生成が期待できる。

E. 結論

本年度はがん相談支援記録のサンプルデータ10件を用いた、テキストマイニング処理のプロトタイプ作成を行い、分析および可視化について結果を得ることができた。次年度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い、プロトタイプを改良し、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

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