

分担研究報告書

医療広報におけるソーシャルメディアの可能性

（共起ネットワークによるツイート解析手法の開発）

分担研究者 田中 敦 山形大学大学院理工学研究科 准教授

研究協力者 津谷 篤 山形大学大学院理工学研究科 研究員

研究要旨

プロフィール上に「胃がん」「腸がん」「子宮がん」「肺がん」「乳がん」「白血病」の記述があるアカウントのツイートを名詞の共起ネットワークとして表し、がんに関係するアカウントから発せられるツイートのうち頻度が高い話題を明らかにした。共起ネットワーク上で各話題は一つにまとまって表される。平成24年度に行った共起ネットワーク作成では無駄な語の除去は人の判断で行っていたが、今年度はそのようなことも行うことなしにどのような話題がされているか把握できるようになった。共起ネットワーク上で見られた話題として、各がんの「診断」「症状」「治療」に関するもののほかに、介護、健康保険の適用不適用、ワクチン接種公費助成、啓発活動といったものがあった。ただし、子宮がんでは、高頻度に現れる話題の中に「診断」「治療」の話題が見られなかった。

A．研究目的

Twitterはユーザが「ツイート=つぶやき」として140文字以内の短い文を投稿するマイクロブログサービスである。Twitterのユーザ数は世界で1億9000万人と言われており、その中には何らかの病気を患うユーザの投稿も見られる。そのツイート中では投稿者の闘病の様子や、治療の様子、治療を受けた感想、その病気に対する自身の考えが記されている。さらに、その病気の治療にあたる医師の患者に対する情報提供をねらいとしたツイートも存在する。それら投稿は、現在同じ病気にかかっているユーザの参考となっていると考えられるだけでなく、今後の応用を見越したビッグデータとしても貴重なものである可能性がある。本研究では、プロフィール上に特定のがん名が記されているツイッターユーザのツイート中で、そのがんに関してどのようなツイートがされているかを、語の共起ネットワーク作成を介して明らかにする。

B．研究方法

まず初めにツイッターのアカウントプロフィールにがんに関しての記述があるアカウントを検索する。ツイッター自体にアカウントプロフィールのみを検索する機能は無い。そこでアカウントプロフィールのみを検索する機能を提供するウェブサービス「16プロフィール検索」を用いることで該当アカウントを集める。

続いて収集したアカウントのツイートを取得する。ツイートの取得にはツイッターAPIを用いる。ツイッターAPIとはツイッターを運営する組織がツイッ

ターアプリケーション制作者に公式に提供している機能のことである。ツイッターAPIを組み込むことで、アプリケーション制作者は自分が制作しているアプリケーションにツイッター運営組織が提供している様々な機能を加えることができる。ツイッターAPIを用いることで指定アカウントのその時点までのツイート最大200ツイートを取得し、全アカウントツイートを一つにまとめる。

（倫理面への配慮）

本研究は、インターネット上に公開されている情報を対象とする。ヒトや動物を対象とはしない。本研究については平成24年度に山形大学医学部倫理委員会の承認を得た。

次に得られたツイート各行を単語（形態素）に分解する。単語への分解には形態素解析ソフト「茶筌」を用いる。そしてそこから名詞のみを抽出する。文を構成する単語（形態素）の品詞は様々ながテキストマイニングでは通常、名詞、形容詞、動詞が用いられる。しかし、本研究では1)否定文の扱いの難しさ、2)抽出された語のその品詞の割合の少なさ、から名詞のみを扱うこととする。「茶筌」には一般的な単語であるならば品詞を判別可能とするデフォルト辞書が備わっているが、今回は医学専門用語も取りこぼすことなく処理したいためウェブサイト「がん情報サービス<http://ganjoho.jp/>」「gooヘルスケア <http://health.goo.ne.jp/>」の記述から得られた品詞判別不能の単語のうち医師ががんに関係が深いと判定したものを辞書に加え形態素解析を行った。

続いて句点「。」や「?」「!」で区切るために

「。」や「?」「!」で改行する．同じ行中出现する単語の各組合せをカウントし，全単語組合せの同じ行中に同時に出現する頻度（共起回数）を調べる．

次に共起回数が多かった組合せ上位100位までを求め，それら組合せの単語同士をリンクさせたネットワークを図示する．このとき各単語はノードとして表される．

高頻度である日常的なツイート，趣味に関するツイートを共起ネットワークから除去する方法として2つの方法を採用する．

日常的なツイート除去に関しては，がんに関係がない一般アカウントの日常的ツイートと共通する内容を除去する方法を採用する．日常的ツイートは一般のアカウントもがんに関係するアカウントも同様ならずである．まず特徴のないアカウント名をいくつか取得し，さらにそれらアカウントのツイートの共通に出現した単語を取得する．そしてその単語を共起ネットワークから除去する．特徴のないアカウントは「16プロフィール検索」サービスで地名をキーワードとして検索して得た．具体的には今回使用したキーワードは「山形県米沢市」である．

趣味に関するツイート除去に関しては，趣味に関する内容を頻繁にツイートするアカウントのツイートを抑制する方法を採用する．趣味に関する内容が共起ネットワークに示されることの原因は，同じ趣味を持つアカウントがたくさんいるからではなく，趣味に没頭しているアカウントが趣味に関して頻繁にツイートすることが原因と考えられる．そこで，同一アカウントに複数回ツイートされた共起語は+1しかカウントしないことにする．つまり共起頻度の定義は「ツイート内同一行で1度でも共起したアカウント数」となる．

C．研究成果

成果1．高頻度である日常的なツイート，趣味に関するツイートを共起ネットワークから除去

高頻度である日常的なツイート，趣味に関するツイートを共起ネットワークから除去する前（図1）と除去後（図2）を示す．除去後，共起ネットワークの多くが医学に関係するものになっていることがわかる．

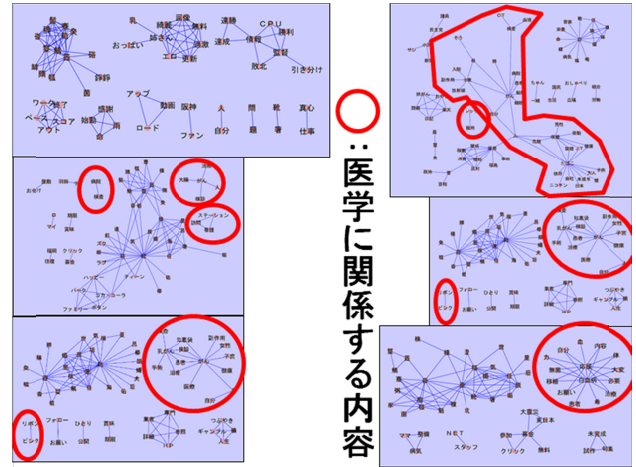


図1．高頻度である日常的なツイート，趣味に関するツイート除去前

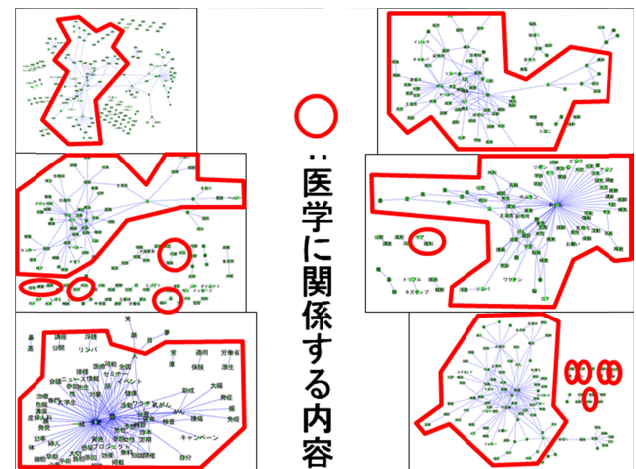


図2．高頻度である日常的なツイート，趣味に関するツイート除去後

成果2．各共起ネットワーク

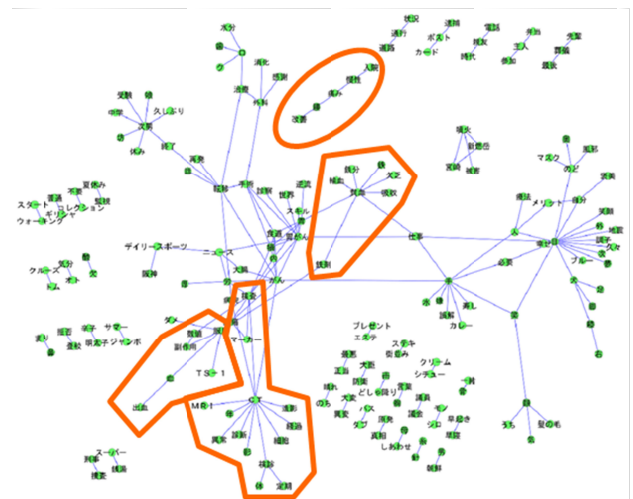


図3．胃がんの共起ネットワーク

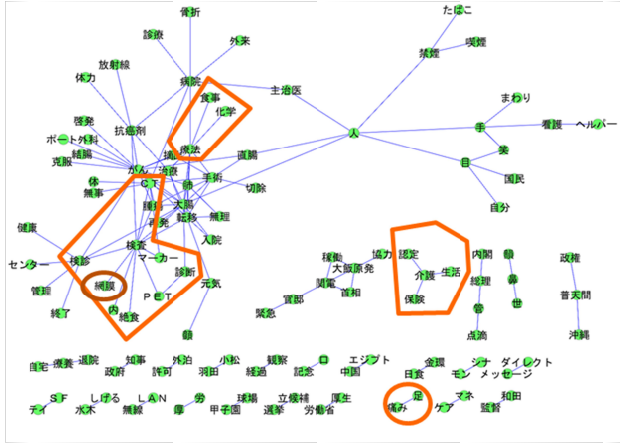


図 4. 腸がんの共起ネットワーク

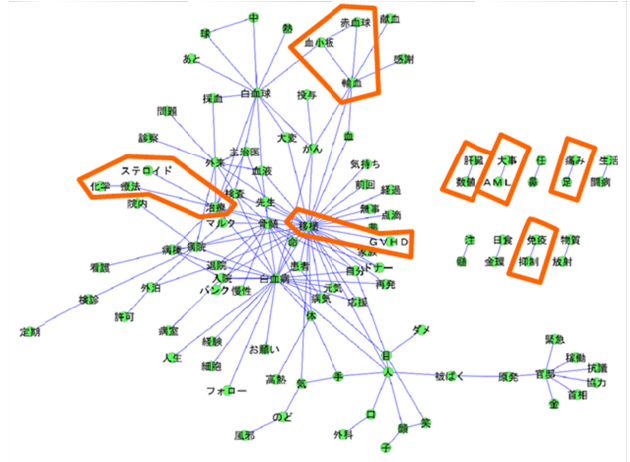


図 8. 白血病の共起ネットワーク

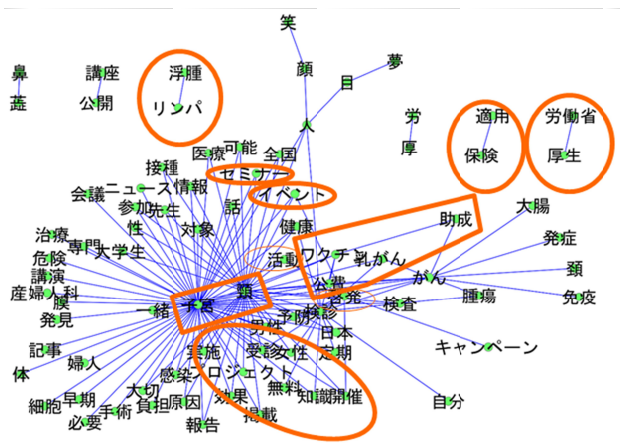


図 5. 子宮がんの共起ネットワーク

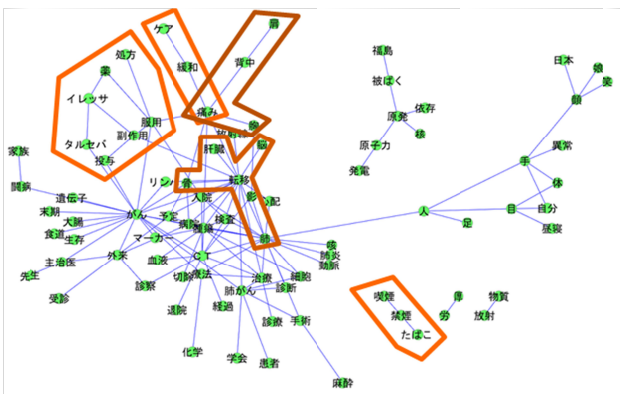


図 6. 肺がんの共起ネットワーク

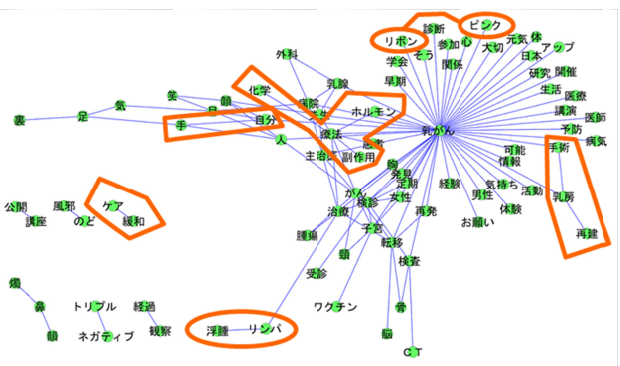


図 7. 乳がんの共起ネットワーク

D. 考察

共起ネットワーク上で見られた話題は、各がんの「診断」「症状」「治療」であったが、子宮がんでは、高頻度に現れる話題の中に「診断」「治療」の話題が見られなかった。これは子宮がんが頻繁につぶやかれる内容が「子宮がんはワクチンで予防できる」といったことの啓発、または啓発活動に関する内容であることがほとんどであることが理由として挙げられる。

ほかには、介護、健康保険の適用不適用、ワクチン接種公費助成、啓発活動といった話題が見られた。

表 1. 各共起ネットワークの内容

	診断	症状	治療	その他
胃がん	CT, MRI, 腫瘍マーカー	腰痛, TS-1副作用	抗がん剤TS-1 鉄剤服用	
腸がん	CT, PET	エルプラット副作用	化学療法 食事療法	介護
子宮がん	×	リンパ浮腫	×	・啓発活動 ・検診健康保険適用外 ・ワクチン接種公費助成
肺がん	CT	転移, 肩痛, 背中痛, イレッサ副作用	抗がん剤イレッサ, アルセバ	緩和ケア
乳がん	自分の手で診断可能	転移 リンパ浮腫	化学療法 ホルモン療法	緩和ケア ピンクリボン
白血病	肝臓数値	肝臓数値 足痛 免疫抑制 GVHD	化学療法, ステロイド治療 赤血球輸血 血小板輸血	AML 造血幹細胞移植

E. 結論

プロフィール上に各がんの記述があるアカウントのツイートをもとに名詞の共起ネットワークとして表し、がんに関係するアカウントから寄せられるツイートのうち頻度が高い話題を明らかにした。平成24年度に行った共起ネットワーク作成では無駄な語の除去は人の判断で行っていたが、今年度はそのようなことも行うことなしにどのような話題がされているか把握できるようになった。共起ネットワーク上で各話題は一つにまとまって表される。

F．研究発表

1.論文発表

なし

2.学会発表

津谷篤、田中敦、成松宏人「語の共起ネットワークから見たがん患者のTwitterツイート」第29回
ファジィシステムシンポジウム 平成25年9
月11日 於 大阪国際大学

津谷篤、菅原祐也、田中敦、成松宏人「語の共起
ネットワークから見たがん患者のTwitterツイ
ート」第17回 日本感性工学会春期大会 平成26
年3月23日 於 北海道大学工学部

G．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