

がん患者やその近親者の闘病ブログに記述される悩みの自動分類

下元 健吾[†] 安藤 一秋[‡]

[†]香川大学大学院工学研究科 [‡]香川大学工学部

1. はじめに

昭和 56 年以降, 日本人の死因の第一位は「がん」であり, 多数の人々ががんで命を落としている. がんには様々な種類が存在し, 病気の部位や病気の段階 (フェーズ) により, その治療方法も多種多様である. そのため, 病気や治療法など様々な情報を書籍や Web 検索などで収集しようとするが, 検索者が求める正しい情報を獲得することは容易ではない.

また, 闘病生活には長期治療が必要なこともあり, 長い闘病生活のためにうつ病などの心の病を併発してしまうことも少なくない. このような状況に対して, 山下らの研究[1]では, 「両親や親友とは異なる, つながりが薄くて境遇の似た人々と交流する場を提供すること」が一つのケアとして重要であると述べている. 心の問題は, がん患者だけでなく, 近親者においても重要となる. そのため, がんを闘うためには, 患者自身および近親者の心のケアを行う必要がある.

Web 上には, がんを闘う人々やその近親者がつづいた闘病ブログが多数存在する. 闘病ブログには, 実際に病気を経験した人や近親者の様々な経験談や心理状態, ケア事例などが記載されているため, 情報検索者と類似した状況にある患者や近親者に有用な情報が存在している可能性が高い. しかし, 現状のブログ検索では, ユーザーが望む闘病ブログを獲得することが困難である.

そこで本研究では, 病気や治療法に関する情報や患者と近親者の闘病経験情報, 告知日や手術, 治療などにおける心理情報など, がん患者やその近親者が必要とする情報を検索するための支援環境の構築を目的とする.

本稿では, 検索支援機能と検索支援に向けたがん患者やその近親者の闘病ブログに記述される悩みを自動分類する手法について検討する.

2. 関連サービス

TOBYO[2]は, 闘病者 (患者・家族・友人) のための闘病体験情報を共有するサービスである. 病名, 病気の部位, 性別, ブログタグ, 発症時の年齢などで絞込みを行い, ユーザーが欲する「闘病ブログ」を検索できる. しかし, 検索結果として得られる情報はブログ記事ではなく, ブログ名やトップページの URL などであるため, 「告知日」や「手術日」, 「がんのフェーズが進行した日」など, 大きな出来事が発生した日について書かれた記事まで絞り込むことができない. そのため, ブログに記載されている内容はタイトルから推測しなければならず, ユーザーの求める情報に素早くアクセスする機能としてさらなる改善が必要であるといえる.

3. 闘病ブログの検索支援

本研究では, 病気や治療法に関する情報や患者と近親者の闘病経験情報, 告知日や手術, 治療などにおける心理情報など, がん患者やその近親者が必要とする情報を検索するための支援環境の構築を目指す. 現在検討している検索支援機能は, 以下の2つである.

1. 患者の属性やキーワードによる絞込み機能
2. タグ付与による絞込み機能

以下, 各支援機能について説明する.

3.1 患者の属性やキーワードによる絞込み機能

患者の年齢, 性別, 病名, キーワードなどでブログを絞込む機能である. これにより, 自分と類似する状況の患者に関するブログの検索を支援する. この機能は関連サービスである TOBYO で一部実装されている. しかし, TOBYO の検索機能は, 投薬名などの任意のキーワードで検索できない. また, 検索結果はブログ名やトップページの URL であるため, 個々のブログ記事を検索結果として得ることはできない. 本研究では, 患者のプロフィール情報に注目することで, 動的に情報を更新する手法を検討している.

3.2 タグによる絞込み機能

ブログ記事にタグを付与し, タグで記事を絞

込む機能である。付与するタグは大きく 2 種類ある。1 つ目は「告知日」や「治療開始日」など闘病生活中の出来事に関するタグである。このタグにより、「がんと告知された日」や「フェーズが次に進行した日」などのブログ記事のみを読むことが可能になる。2 つ目はブログに記述されている悩みに関するタグである。このタグにより、ブログ検索者の悩みと類似する情報が記述されるブログ記事を読むことが可能になる。

以降、悩みタグを付与するための悩みの自動分類について述べる。

4. 闘病ブログに記述される悩みの自動分類

悩みに関するタグを付与するためには、闘病ブログ記事の各文に記述されている悩みを分類する技術が必要となる。そこで、悩みを自動分類する手法について検討する。

悩みの分類には、宮部らの研究[3]を参考にする。宮部らは、静岡分類[4]というがん体験者の悩みの分類体系を用いてブログに記述されている悩みを分類している。静岡分類とは、がんに関する不平、不満、悩み 7,855 件を整理し、16 の大分類とさらに深い階層構造に体系化した分類である。

本研究では、がん患者や近親者の心の問題に焦点をあて、静岡分類の第 12 分類「不安などの心の問題」に注目する。第 12 分類の下位分類の中から体験者の声データの数が多く、表 1 に示す下位分類を対象とする。したがって、本研究では、ブログに記載される心の問題を表 1 に示す 10 分類に分類することを旨とする。

表 1: 対象とする第 12 分類の下位分類

12.1.1.	再発・転移の不安
12.1.2.	再発・転移に伴う動揺やつらさ
12.2.1.	見通しのつかない未来に対する漠然とした不安（治療や身体のこと）
12.2.2.	見通しのつかない未来に対する漠然とした不安（経済面・社会面）
12.2.3.	身体の状態や生活に対する漠然とした不安
12.3.1.	死を意識
12.3.2.	がんを意識
12.4.1.	がん告知（再発・転移・進行の告知を含む）による精神的衝撃
12.4.2.	がん罹患後の精神的不安定感
12.4.3.	持続する精神的な不安定感

闘病ブログに記述される心の悩みを自動分類する手法を以下に示す。

手順 1:

静岡分類の第 12 分類「不安などの心の問題」の下位 10 分類に記述されている体験者の声データを学習データに用いて、one-versus-rest 法による多クラス SVM (Support Vector Machine) で分類器を構築する。

手順 2:

手順 1 で構築した多クラス分類器を用いて、闘病ブログに記述される悩みを自動分類する。

5. 体験者の声データによる性能評価

手順 1 に従って構築する多クラス分類器の性能を確認するため、正答率を評価尺度とし、10 分割交差検証で評価する。学習データは、第 12 分類の下位 10 分類に含まれる体験者の声データを各 100 件利用する。SVM の素性には、体験者の声データの BOW を使用し、素性値には出現頻度を用いる。なお、SVM には、scikit-learn の LinearSVC を利用し、パラメータは規定値である。

表 2 に実験結果を示す。正答率の平均は 0.54 と高い値ではないが、標準偏差が 0.09 となり、クラスによって大差のない分類器が構築できた。しかし、正答率が低いため、素性と学習データについて検討が必要である。

表 2: 10 分割交差検証の結果

分類	12.1.1.	12.1.2.	12.2.1.	12.2.2.	12.2.3.
正答率	0.57	0.54	0.51	0.61	0.57
分類	12.3.1.	12.3.2.	12.4.1.	12.4.2.	12.4.3.
正答率	0.51	0.58	0.44	0.53	0.54

6. おわりに

本稿では、がん患者やその近親者の闘病ブログの検索支援を目的に、患者の属性やキーワードによる絞込み機能とタグ付与による絞込み機能をもつ検索支援環境を提案した。また、闘病ブログに記述される悩みの自動分類法を検討した。体験者の声データを利用した性能評価の結果、クラスに依存しない分類器が構築できたことを確認した。今後は、さらなる分類性能の向上と検索支援機能の実装を進める。

参考文献

- [1] 山下他, “うつ病患者の家族看護者が抱える社会的負担を構成する要素の解明”, 情報処理学会論文誌 Vol.55, No.7, pp.1706-1715, 2014.
- [2] TOBYO : <https://www.toby.jp/>
- [3] 宮部他, “Web テキストから医療の悩みを抽出する: がん体験者の悩みの自動分類”, FIT2014, 13(2), pp.95-96, 2014.
- [4] 「がんの社会学」に関する合同研究班, “がん体験者の悩みや負担等に関する実態調査報告書 がん向き合った 7,885 人の声”, 静岡県立静岡癌センター, 2006.