

うつ病患者の周辺看護者支援に向けたブログ記事へのタグ付けと分類

齋藤 俊英†

平田 圭二†

荒牧 英治‡

宮部 真衣‡

† 公立はこだて未来大学

‡ 京都大学学際融合教育研究推進センター

1 はじめに

うつ病患者を身近で手助けする家族、恋人、友人や職場の人々(以降、周辺看護者)は、直面している悩みを解決したり、近未来を予測するために、自分自身の置かれた状況により近い人物から情報を求めている。類似した状況を検索する際の項目は、患者やその他の人物との関係性、家族構成、年収、ある人物の行動や発言といった意味的なものである。しかし一般的な情報検索方法である Web のキーワード検索では、似ていない状況のページを多く含んでしまう。山下らは、うつ病患者の看護経験がある 15 名の家族看護者にインタビューを行い、コミュニケーション環境の改善や精神的および社会的ストレスを軽減するために有効と思われる技術に関する示唆を述べた [1]。本研究の目的は、その 1 つである探究と逃避のジレンマにおける、探求の支援である。

意味的な情報を検索する上で適切なタグやカテゴリの条件を得るため、OKWave[2] に投稿された文章を対象に分類とタグ付けを行った。OKWave とは匿名で質問や相談が出来る投稿サイトである。状況説明のために意味的な情報の一部公開が必要な周辺看護者の投稿も多数見られる。一方、誰でも投稿が可能のため起承転結・時系列順・口語的・回想を多く含むなど、さまざまな形式の記述が存在する。特定の形式や主張が明確に記述されていない文章から、周辺看護者の求める文章を見つけ出すことが本研究に求められる。

2 機械学習を用いた分類とタグ付け

自由に記述された投稿の中からある投稿に類似したものを見つけるために、Support Vector Machine (SVM) と Conditional Random Field (CRF) による用語の抽出と文章の分類という 2 つのアプローチにより周辺看護者の状況により近い投稿を見つけ出す。SVM を用いた自動分類では、候補全体の数を A、カテゴリの数を C とすると、投稿を均等に分類できる適切なカテゴリを設定することで、全体の候補を A/C まで絞ることができ

る。そのため分類により検索候補を削減できるという意味で、キーワード検索など他の検索手法と併用しやすく、検索精度の向上が可能である。CRF による自動タグ付けでは、文章中の登場人物の情報や他の人物との関係性をタグで表現することで意味的な情報の抽出が期待できる。

2.1 用語の抽出

人物間の関係性を投稿から探索する際のインデックスとして、それらを表現するタグを導入する。a, p のタグは文章において存在が保証される周辺看護者の筆者 (author) と患者 (patient) を表す。その他の人物 (other) は、存在しない場合や複数人存在する場合など人数にばらつきがあるため、o のタグとして統一し表現した。a, p, o を組み合わせることにより人物の関係性を、ap, pa, do, od のタグで表現する。do, od は学習効率向上のため、本来 po, ao, op, oa であるところをまとめて表現した。対象のデータは 2014 年 8 月 26 日時点で OKWave の夫婦家族カテゴリに分類される投稿 3577 件中 1000 件である。学習データは 1000 件の中から周辺看護者の投稿であると判断した 449 件 (平均文字数 1656 字) にタグ付けを行い作成した。

表 1 タグの種類

タグ	文章中の対象となる箇所
p	患者自身の情報
a	筆者自身の情報
o	その他の人物の情報
ap	筆者から患者への行動や心情
pa	患者から筆者への行動や心情
do	筆者・患者からその他の人物への行動や心情
od	その他の人物から筆者・患者への行動や心情

2.2 実験計画と実施

10 分割交差検定を行った。まず用意した学習データを均等に 10 のグループに分割する。それぞれのグループを A, B, ..., J とする。A 以外の 9 つのグループの文章を全て学習し、タグを除去した A に新たに自動タグ付けを行う。この作業を A から J まで行い、除去前のタグの数 (GOAL)、システムが付けたタグの数 (SYS)、除去前のタグとシステムが付けたタグで一致するタグの数 (MATCH) を算出する。

Automatic Classification and Tagging of Blog Articles towards Supporting Caregivers for Depressed Family Member

†Toshihide Saito, Keiji Hirata, Future University Hakodate

‡Eiji Aramaki, Mai Miyabe, Center for the Promotion of Interdisciplinary Education and Research

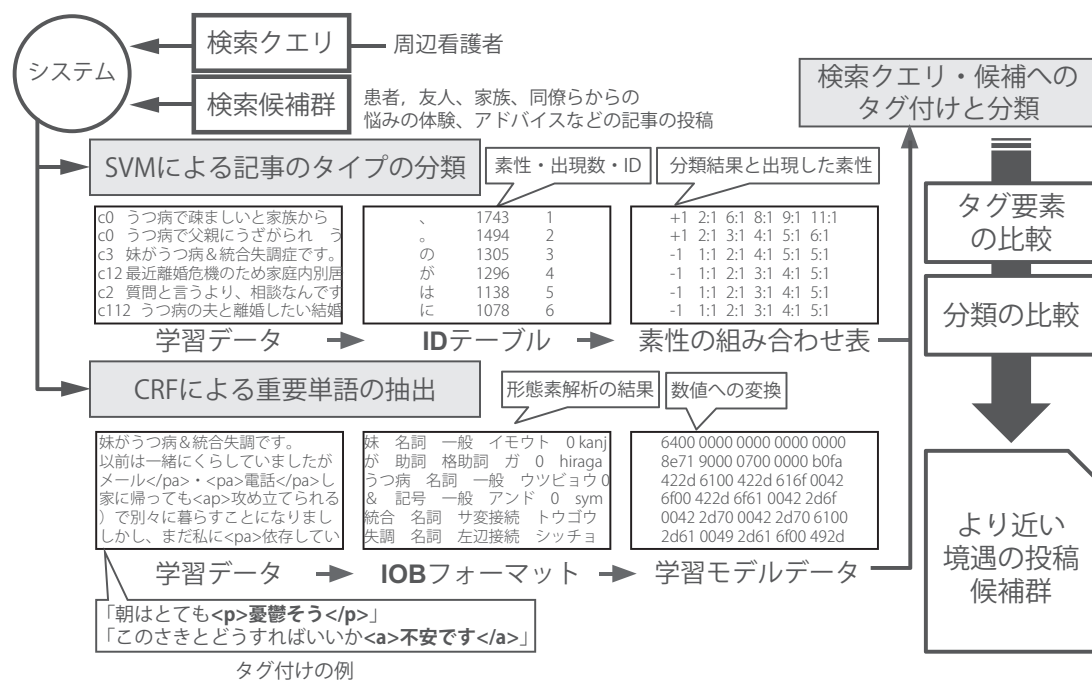


図. 1 検索手法概要

2.3 結果と分析

実験の結果は表2の通りとなった。それぞれの値の

表2 実験結果

	A	B	C	D	E
GOAL	1181	1517	1471	1356	1198
SYS	210	216	216	214	219
MATCH	27	36	24	30	40

	F	G	H	I	J
GOAL	1389	1281	1431	1169	1368
SYS	221	213	219	218	264
MATCH	46	35	34	44	40

平均をとり、適合率、再現率、F値を算出した。結果は適合率 25.3 % (= 139.2/1336.1), 再現率 10.4 % (= 139.2/1747.2), F値 0.147 である。GOAL に対して SYS の値が低く、SYS に対して MATCH の値が低いことから、<a> ショックを受け と <a> ショック のように学習データのタグ付けが統一されていなかったため、学習効率が低下したこと、また学習データの不足が精度の低い要因として考えられる。そのためさらなる学習データの作成とタグ付けの範囲の統一により、精度の向上が期待される。

3 おわりに

本研究では、周辺看護者が検索を行う際の問題点と課題を説明し、検索手法に用いる CRF によるタグ付けに対して 10 分割交差検定を行い結果を報告した。今後、

検索手法を実現するために、投稿内容の時間展開に関するタグ付け、抽出した用語の利用方法、タグ付けと分類の組み合わせの検討が必要だと考えられる。状況を適切に捉えるために、文中の過去の出来事と現在の状況を表現する必要がある。抽出した用語の利用に関しては、タグの個数や中身の比較による類似度の表現などが考えられる。タグ付けと分類では、視点（筆者と患者との関係の情報）が重複して含まれるため、複数の組み合わせが考えられ、場合によってはタグの設定を改めて検討する必要がある。

謝辞

本研究は JSPS 科研費 26280116 の助成を受けたものです。

参考文献

- [1] 山下直美, 葛岡英明, 平田圭二, 工藤喬: 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] 質問・疑問に答える Q&A サイト OKWave, <http://okwave.jp> (2014).