

テキストマイニングによる成功経験の抽出

中島正人¹ 本村陽一¹ 三浦未生² 武澤友広³ 石渡利奈³

産業技術総合研究所¹ 産業技術大学院大学² 国立障害者リハビリテーションセンター研究所³

1. はじめに

今までできたことが、なぜかできない。認知症初期に見られる能力の喪失感、認知症者の自己評価を低下させ、活動の意欲を低下させる[1]。この傾向は、特に若年性認知症者において深刻である。介護のみで対応しきれないと認識されており、自立支援を行う機器の開発が求められている。われわれは初期の認知症者が書いた日記の中から成功経験を自動的に抽出し、その経験を想起させるようフィードバックを行う支援機器の開発を目指す(図1)。

この支援機器の開発には、ポジティブ感情が生じた経験の想起によって、抑うつ傾向が低減する効果があるという認知心理学的知見が基礎となっている[2]。筆者らは、この知見を初期の認知症者に活用できると考える。その際、手段として日記の活用を提案する。日記は①手軽に日々の活動を記録でき、②読み返すことでフィードバックになるという二つの役割を果たすためである。

本稿では、その手始めとして、成功経験に関するキーワードをテキストマイニングによって抽出し、辞書化するまでの方法について2つの点から検討する。①成功体験を達成と親和に分類し、それぞれにあてはまるキーワードを抽出すること、②抽出されたキーワードを用いてより精度の高い辞書作成するために、ICFを用いてコード化することである。

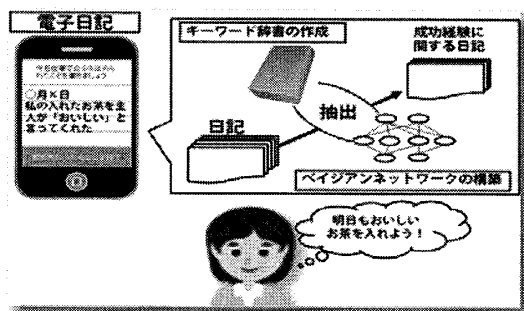


図1 電子日誌システム

Extraction of positive experience by text mining.

Masato Nakajima¹, Yoichi Motomura¹, Miki Miura², Tomohiro Takezawa³ and Rina Ishiwata³, ¹National institute of advanced industrial science and technology, ²Advanced institute of industrial technology, ³National rehabilitation center for persons with disabilities

2. 成功経験記述データ

2-1 記述データ収集

成功経験に関連するキーワードを抽出し、辞書を作成するために、「自分をほめてあげたい」過去の経験に関する記述を収集した。

調査対象者は、中高年男女 76 名であった。記述の際、だれが、いつ、どこで、何をしたという点を明確に記すこと、思い出した順番に最高で3つまでの経験を書くよう求められた。最終的に、全 154 件のデータ(文章)が収集された。各データの分量は 1~5 行程度であった。

2-2 ラベルの分類

各データは、達成、親和、その他のラベルに分類された。分類の基準は以下である。達成は「自分で成し遂げたこと」、「他者からの評価」などの記述であり、業績、評価、達成、手柄、などの語を含むものとされた。親和は「ひととの共生」、「他者への施し」などの記述であり、愛、協力、友情、理解などの語を含むものとされた。さらに、両者を同時に含むものを達成&親和とし、どちらも含まないものは、その他とラベル付けされた。

その結果、154 件の全データに関して達成、親和のどちらか、または、どちらも含む場合に分類され、その他とされるものはなかった(表1)。

3 テキスト処理

3-1 構文解析

154 件のテキストデータに対して Text Mining Studio[TMS]を使用して構文解析を行った。解析結果として、行 ID、文章 ID、単語 ID、見出し語(テキストの構文解析結果)、原型(見出し語の原型)、置換語(辞書によって変換された語)、品詞、係り先、述語属性が出力された。その結果 1438 語が抽出された。

3-2 評判分析

評判分析とは、テキストデータが肯定的内容か、否定的内容かを判定するために用いられるテキストマイニングの手法の一つである。評判分析には TMS を用いた。評判分析によって 154 件のテキストデータを肯定的内容と否定的内容に分類された。これはラベル付けの結果がどれくらい再現されるかを評価

するものであった。達成、親和にラベル付けられた内容は成功経験であるため肯定的であると判定されることが期待される。また、評価が正しかったデータの語は、各ラベルで抽出されたキーワードとしてより適切なものであると考えられた。

達成、親和に該当する肯定的内容には 29 件、否定的内容として 125 件が割り振られた (表 1)。ラベルの分類と適合していないことが明らかになった。

判定率が低かった点について、各データで記述の仕方に個人差が大きかったこと、文中にある複数の話題が存在することに問題があったと考えられる。

表 1 ラベル分けと評判分析の結果

ラベル分け		評判分析	
達成	79 件	肯定的内容	29 件
親和	46 件		
達成&親和	29 件		
その他	0 件	否定的内容	125 件

3-3 コード化による辞書化

評判分析には限界があった。係り受けの関係よりも、共起頻度の方が抽出できる話題が多い (図 2 左)。しかしながら、TMS では共起頻度による判定において文中の複数の話題に対応できない。また、評判分析では達成と親和を分けて抽出できないため改善する必要があった。

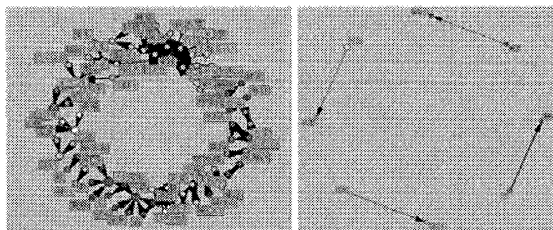


図 2 共起頻度 (左) と係り受け頻度 (右)

共起頻度の問題と達成・親和を分類して抽出する方法として、国際生活機能分類 (ICF) を利用したコードによる辞書を作成し、その評価を行った。ICF とは、WHO が作成する日常の生活活動を心身機能と身体構造、活動と参加の点でコードによって記述したものである。コード化することで自由記述のテキストを体系化できるというメリットが見込める。

方法として、まず、達成・親和に対応するコードを ICF から選出した (表 2, 左 1, 2 列)。その結果、達成 8 コード、親和 6 コードを選出した。続いて、構文解析によって抽出された語を対応するコードへと振り分けた。その際、共起する語の組み合わせで辞書化した (表 2, 「共起する語の例」, 「登録数」*登録数は共起語の組み合わせ数を示す)。親和での登録数が多いのは、共起する語に関連するひとを表す語の数が多いためである。

表 2 ICF コードとそれに分類された語の例と数

コード	内容	共起する語の例			登録数
		免許	取得	できる	
d1551	複雑な技能の習得				8
d1750	単純な問題の解決				0
d1751	複雑な問題の解決	目標	達成	できる	12
d177	意志決定	自分	意志	取り組む	11
d2201	複数課題の達成				0
d2302	日課の達成	毎日	出来る		14
d2400	責任への対応	業務	やり遂げる		19
d2402	危機への対応	ストレス	克服		9
d6600	セルフケアの援助	家族	世話		66
d6601	移動への援助	救急車	手配		30
d6603	対人関係への援助	地域密着化	貢献		2
d7100	敬意と思いやり	相手	立場	理解	31
d7101	感謝	家族	協力	感謝	22
d7102	寛容さ	家族	受け入れる		41

《ICF コード辞書の評価》

154 データに ICF コードによって達成、親和を分類した辞書を用いて適切な分類ができるかを確認した。その結果、コードによる分類によって、評判分析よりも判定率が向上した (表 3)。達成では判定率が 70% となり、達成&親和では 59% となった。親和は 43% と全体と比べ、やや低い判定率であった。

表 3 ICF コード辞書の評価結果

	正解	不正解	計	判定率
達成	55	24	79	70%
親和	20	26	46	43%
達成&親和	17	12	29	59%
	92	62	154	60%
評判分析	26	125		17%

4. まとめ

本稿では、テキストマイニングによって成功経験、特に、達成と親和の記述を抽出できることを確認した。その際、抽出されたキーワードを用いてより精度の高い辞書作成するために、ICF を用いてコード化することが判定率を高めることが確認できた。

今後さらに精度を高めるため、適切なキーワードの抽出とより詳細なコードの分類を行う必要がある。

謝辞

本研究は、経済産業省事業「サービスと IT の融合による新市場創出促進事業 (サービス工学研究開発事業)」および文部科学省科学研究費補助金 (課題番号: 21500536) の助成を受けた。

参考文献

- [1] Clare, Rowlands, Bruce, Surr, & Downs. 'I don't do like I used to do': A grounded theory approach to conceptualising awareness in people with moderate to severe dementia living in long-term care. *Social Science & Medicine* (2008)
- [2] Seligman, Rashid, & Parks. *Positive Psychotherapy*. *American Psychologist*. (2006)