

オントロジーを用いた行動目的の抽出に関する検討

Ontology-based extraction of purpose for action from large-scale texts

横山 法子[†] 船越 要[†] 佐藤 隆[†] 鷺崎 誠司[†]
Noriko YOKOYAMA Kaname FUNAKOSHI Takashi SATO Seiji SUSAKI

1. はじめに

情報ポータルサービスにおいて、ユーザの曖昧な要求に答える情報提供のためには、要求の背景にある多様な行動目的を考慮する必要がある。そこで、様々な行動目的を体系化し、その行動目的と達成手段を人々の経験情報から自動で抽出し、曖昧な行動目的を持ったユーザに目的に合った情報を提供する方法を提案する。

行動目的の体系化・抽出に関しては、従来は特定の行動に関する行動目的や、短期的に達成可能な行動目的を扱っていた。例えば、旅行ブログサイトから旅行に関わる行動目的を抽出する研究がある[1]。この中では「世界遺産」「京都っぽい」といった事象に対する形容表現を抽出している。

また、個人の経験情報の抽出に関しては、web 文書から個人の経験情報の抽出、及びデータベース化を行う研究がある[2]。この中では「状況(時間, 空間)」「行動(動作, 対象)」「主観(評価, 感情)」を個人の経験情報として抽出している。

我々は、経験情報の 1 つとして行動目的に着目し、旅行や買い物等複数の行動を取り得る行動目的や、長期的な取り組みを要する行動目的を含めた様々な行動目的の体系化を目指す。行動目的は、「達成感をしたい」といった人間の願望の形で表現する。そして、体系化した行動目的、行動、詳細情報を経験情報から自動で抽出、及びデータベース化して、行動目的に対して、複数の行動を提案する。

2. 提案手法

提案手法は、行動目的の体系化、行動目的・達成手段の自動抽出、情報提供の 3 つの技術から構成される。以下、それぞれの技術について提案手法を述べる。

2.1. 行動目的の体系化手法

ユーザが様々な行動目的を選択することができるように、様々な抽象度の行動目的を体系化する。ここで行動とは、「散歩する」といった動作をとまらう行動、「充実した人生を送る」といった状態、または「達成感を得る」といった心理的活動と定義する。そして行動目的とは、ある行動を行うことで達成したい行動であると定義する。例えば、気分転換するために散歩をしたのであれば、「散歩をする」が行動、「気分転換する」が行動目的となる。

図 1 は系統図[3]の概念を表したものである。行動目的を達成するために行動(手段)が選ばれ、その行動を講じるためにさらに下位レベルの行動(手段)が存在する。本研究では、行動目的もこのように相対的なものであるとし、行動目的は上位から見れば行動(手段)であり、下位から見れば行動目的でもあるとする。

そこで、行動目的同士を目的-手段関係に着目してネットワークを構築することによって、行動目的の体系化を行う。これを行動目的オントロジーと呼ぶ。具体的には、人々の経験情報から様々な行動の行動目的を収集する。そして、様々な行動目的を、手段-目的関係を考慮してネットワーク構造にする。様々な行動目的同士をつなぎ、その際、間に入るふさわしい行動目的があれば適宜追加することによって、暗黙情報の顕在化を行い、行動目的の拡張を行う。また、類似の行動目的はまとめ、行動目的クラスを作る。

[†]NTT サービスエボリューション研究所 NTT Service Evolution Laboratories

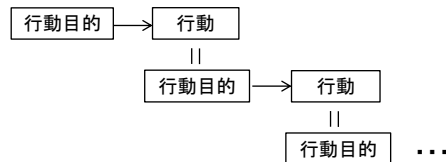


図 1 系統図の概念

2.2. 行動目的・達成手段の自動抽出手法

2.1 で体系化した行動目的を用いて、個人がある行動に関する経験を綴った文書から、行動目的・達成手段を自動で抽出し、行動目的・達成手段データベースに保存する。行動目的としては、行動目的クラスと行動目的詳細文、達成手段としては、行動と詳細情報(文書全体)を保存する。

経験文書から行動目的の抽出を行うために、3.1 で体系化した行動目的クラス毎に、特徴語(その行動目的クラスにおいて出現頻度が高く、他の行動目的クラスには表れない単語または文)を用意する。そして、経験文書に特徴語を含んでいれば、その特徴語が属する行動目的クラスが文書に書かれていると判断する。特徴語を含む文を、行動目的詳細文として抽出する。

達成手段の抽出に関しては、既存の経験マイニング技術[2]を用いて、「動作動詞+対象」または「自動詞」を行動として抽出する。

2.3. 情報提供手法

ここでは、行動目的オントロジーを用いてユーザの行動目的を明確化し、明確化された目的に対して、適切な情報を 2.2 で作成した行動目的・達成手段データベースから抽出して提供する。ユーザに行動目的オントロジーから行動目的クラスを選択してもらい、さらに、情報を絞り込む条件(地域など)があれば指定してもらい、システムは、ユーザが選択した行動目的のクラスや条件をもとに、行動目的・達成手段データベースを参照し、類似の状況における手段情報を抽出する。そして、行動の提案と詳細情報の提供を行う。類似状況を抽出する部分に関しては、様々な手段が抽出できることが考えられる。その中で、もっとも出現することが多い行動を提案することで、ノイズを軽減する。

3. 実験

提案手法の妥当性を確かめるために実験を行った。行動目的の体系化、行動目的・達成手段の自動抽出の際に用いる経験情報としては、大量の経験データがあること、行動目的がはっきりと記述されていることが必要である。経験情報が記述された公開データとして、Twitter、Q&A サイト、ブログが考えられるが、その中で記述量の多いブログを対象にする。対象データは今後拡張予定である。

3.1. 行動目的の体系化

ブログを用いた行動目的の体系化実験を行った。様々な行動目的を洗い出し、まとめることが目的である。

まず、行動目的を、趣味や遊び、日常生活活動等の 10 種類の行動を綴ったブログ記事(100 件×10 種類)から手動で抽出した。抽出する際には、内閣府の生涯学習世論調査[4]で使われていた、「生涯学習を行う理由」から 12 個の行動目的を参考にし、430 個の行動目的を抽出した。これらの中で類似の行動目

的はまとめて行動目的クラスを作り、目的-手段関係となるようにネットワークを構築した。その結果、50 個の行動目的クラスからなる行動目的オントロジーが作成できた。図 2 は作成した行動オントロジーの一部である。ネットワークを考えやすいように、「充実した人生を送る」という最上位の行動目的クラスを設定し、階層構造にした。階層構造にしたため、一部重複して表れる行動目的クラスもある。このようにネットワークを階層構造で表現してもよい。

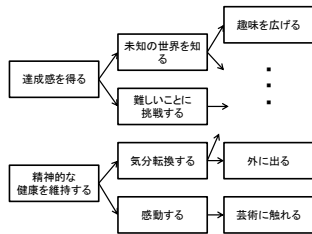


図 2 行動目的オントロジー

3.2. 行動目的・達成手段の自動抽出

行動抽出に関しては、すでにできているものとし、行動目的抽出部分に関する実験を行った。

3.2.1. 実験手法

実験には、3.1 で作成した、10 種類の行動を綴ったブログ記事から、430 個の行動目的を抽出したデータセットを用いた。行動目的が抽出できた文書に対し、50 種類の行動目的クラス(複数可)が正解として付与されている。入力がブログテキスト、出力が複数の行動目的クラスとなる多クラス分類の問題である。

提案手法として、上述のブログ記事を教師データ 221 件、テストデータ 202 件に分け、教師データを用いて行動目的クラスごとに特徴語を抽出した。そして分類実験では、ブログ記事に特徴語を含んでいれば、その特徴語が属する行動目的クラスが、その文書に書かれていると判断し、正解クラスとする。

また、ベースラインとして、ナイーブベイズを用いた分類を行った。ブログ記事を bag-of-words 表現に変換し、提案手法と同様の教師データ、テストデータを用いて行動目的クラスの分類を行う。教師データにおいて、クラス c に単語 $word$ が出てきた回数を $T(c, word)$ 、教師データにおける単語集合を V とする。そしてテストデータにおいて、 $[word_1, word_2, \dots, word_d]$ からなる文書 d の事後確率が最大となる行動目的クラス c_{max} を以下の式で求める。

$$\begin{aligned} c_{max} &= \arg \max_c P(c)P(d|c) \\ &= \arg \max_c P(c) \prod_i P(word_i | c) \\ &= \arg \max_c P(c) \prod_i \frac{T(c, word_i) + 1}{\sum_{word' \in V} T(c, word') + |V|} \end{aligned}$$

3.2.1. 結果と考察

ベースラインと提案手法を用いて分類実験を行った結果を表 1 に示す。それぞれの手法について、各行動目的クラスの適合率 P (=あるクラスと推定した中で正解した数/あるクラスと推定した数)、再現率 R (=あるクラスと推定した中で正解した数/あるクラスの正解数)、F-measure(= $2PR/(P+R)$)を算出し、平均を取った。ベースラインに比べ、提案手法の適合率はわずかしが改善が見られなかった。サイクリングに関するブログであれば、サイクリングに関する行動目的クラスのみを正解とし、他の行動に関する行動目的クラスは正解としていない。そのため、適合率は低い値となったと考えられる。この問題に対処するために、今後は単語間の近さや因果関係表現等をもとに、正しい行動目的-行動ペアを抽出する予定である。また、再現率に関しては、ベースラインと比較して、提案手法では高い値が

出た。しかし、行動目的クラスによっては様々な表現の仕方があるので、特徴語と違う表現を行っている行動目的には分類できなかった。この問題には、シソーラス[5]の類語情報を基に、特徴語の拡張を行うことで対処する予定である。

表 1 行動目的分類結果

	ナイーブベイズ	提案手法
平均適合率: P	0.093	0.130
平均再現率: R	0.119	0.493
平均F-measure	0.104	0.206

3.3. 情報提供

3.1 で使用したブログ記事を用いて、行動目的に対する行動の提案可能性を確認した。3.1 では、10 種類の行動を綴ったブログ記事(100 件×10 種類)から、手動で 430 個の行動目的を抽出し、50 個の行動目的クラスに分類した。そこから、50 種類の行動目的クラス毎に、行動目的クラスに対して行われることが多い行動を抽出した。表 4 に示す通り、例えば「達成感を得たい」という行動目的クラスに対して、サイクリングやピアノ、他人に認められるという行動目的クラスに対して「料理作り」といったように、妥当な行動が提案できることが確認できた。

表 2 目的達成のために行われることが多い行動 (行動 10 種類中)

行動目的クラス	行動目的達成のために行われることが多い行動	行動目的詳細
おいしいものを食べる	旅行	-グルメ旅行、楽しみはやっぱり食だね
自由時間を有効に使う	サイクリング	-家の中でゴロゴロするだけの生活では嫌なので
達成感を得る	サイクリング、ピアノ、パソコン	-上からの景色は、達成感もあり良かったよ! -忙しい状況の中でクラシックの名曲を1曲仕上げるとなるとかなりの努力を必要とされますが、その曲を弾けるようになった時の達成感や努力を重ねた分自信に繋がっていくと思います。 -時間をかけて勉強した甲斐がありました
楽しいことをする	遊園地	-久しぶりに遠出してアクティブな遊びして楽しかった
健康を維持する	サイクリング	-健康的にサイクリング(ママチャリ)しようかなあ。
精神的な健康を維持する	ピアノ、外食	-元気をもらいました(^^) やっぱり音楽はいいね -美味しいもの食べて心も体も栄養補給
興味を広げる	ジャグリング	-インドアの趣味-特技ほしいなあ、と思って...コンタクトジャグリング始めました~
他人に認められる	料理	-パンが作れると、お料理上手に見えるのが嬉しいね~
ボランティアをする	旅行	-旅行に行くことで少しでも復興支援になればいいな

4. まとめと今後の課題

本研究では、オントロジーを用いて様々な行動目的を体系化し、その行動目的と達成手段を自動で抽出し、曖昧な行動目的を持ったユーザに適切な行動を推薦することできる情報ナビゲーション技術を提案した。本手法により、ブログから様々な行動目的を体系化できることを確認した。また、行動目的抽出実験を行い、ベースラインに比べ行動目的抽出精度の向上を確認した。さらに、行動目的に対して、妥当な行動が提案できる可能性を確認した。今後は、対象とする行動を増やし、行動目的のオントロジーの拡張と、行動目的-行動ペアの抽出精度の向上を目指す。

参考文献

[1] 若木裕美, 有賀康顕, 中田康太, 藤井寛子, 住田一男, 鈴木優, “ロコミ情報からの目的情報抽出”, FIT2011(第 10 回情報科学技術フォーラム).

[2] 倉島健, 藤村考, 奥田英範, “大規模テキストからの経験マイニング”, 電子情報通信学会論文誌, Vol.J92-D, No.3, pp.301-310, 2009.

[3] 水野滋, “管理者スタッフの新 QC 七つ道具”, 日科技連, 2000.

[4] 内閣府大臣官房政府広報室, “生涯学習に関する世論調査(平成 24 年 7 月調査)”.

[5] NTT コミュニケーション科学基礎研究所, “日本語語彙体系”, 岩波書店, 1999.