



| | |
|------------------|---|
| Title | つばやきロボット：インターネット資源を利用したユーザの要求発見 |
| Author(s) | 高木, 慧佑; ジェブカ, ラファウ; 荒木, 健治 |
| Citation | ファジィシステムシンポジウム講演論文集, 27, 916-919 https://doi.org/10.14864/fss.27.0.211.0 |
| Issue Date | 2011-09-12 |
| Doc URL | http://hdl.handle.net/2115/63618 |
| Type | proceedings |
| Note | 第27回ファジィシステムシンポジウム(27th Fuzzy System Symposium). 2011年9月12日～14日．福井大学文京キャンパス, 福井市. |
| File Information | FSS27-2011_916-919.pdf |



[Instructions for use](#)

つぶやきロボット - インターネット資源を利用した ユーザの要求発見

Tweeting Robot - Using Internet Resources to Discover User Needs

○高木 慧佑¹, ジェプカ ラファウ¹, 荒木 健治¹

○Keisuke Takagi¹, Rafal Rzepka¹, Kenji Araki¹

¹北海道大学大学院情報科学研究科

¹Graduate School of Information Science and Technology Hokkaido University

Abstract: In this study, we describe using text information from Twitter to understand context for task selection of a robot. An intelligent robot of the future is required to discover users' discomfort and improve it whenever possible. However, the level of image or speech recognition is far from satisfying, which does not help robots to understand users' needs from context when an order is not explicit. Therefore, we decided to utilize the text information to automatically discover situations when robot's help is needed. Users share their information about their daily matters on Twitter. A vacuum-cleaning robot acquires users' "tweets", and starts cleaning if it is recognized as a description of situation requiring action. Additionally robot can reply user when there is an explicit request but the robot is not able to perform the task. Moreover, a chat module is triggered on to answer tweets which were recognized as unrelated to the cleaning task.

1. はじめに

将来の知的システムやロボットはユーザの要求や不快な状態を自動的に発見して、その要求を解決する手段や、不快を改善する手段を選択し、提供することが求められる。そのためにはユーザの生活している実世界のコンテキスト情報をシステムが得て、それを理解する必要がある。しかしながら、現状のカメラやマイクから得られる情報のみでは、ユーザの状態を理解するための情報とするには不十分である。そこで我々の研究ではインターネットのような集合知から得られる情報と、実世界の情報を組み合わせることで、システムやロボットのコンテキスト理解を助けることを目的としている。

関連研究として、ショッピングモール内のユーザの移動の軌跡情報を蓄積し、収集した情報から道案内を求めているユーザを発見して、そのユーザに対しロボットが道案内を行うという研究が行われている[1]。また我々の従来研究では、複数のユーザがいる環境において、ロボットが実行したタスクに対するユーザの評価と、タスクを行った時間と場所の情報を収集することで、その時間と場所に依存したユーザの評価を推定することが可能であることを明らかにした[2]。しかしながら、これらの研究では言語情報を、システムがコンテキストを理解するための情報として扱ってはいなかった。

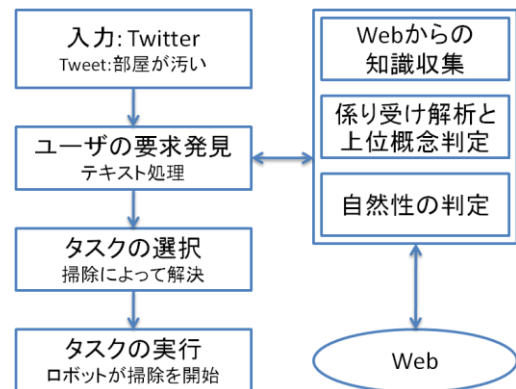


図1 処理過程

本稿では、インターネット資源を知識源として利用し、ユーザが行った Twitter¹の書き込みからユーザの要求を自動的に発見するための言語処理手法を提案する。要求を解決するタスクとして、自動掃除ロボット Roomba²による「掃除」を設定し、「部屋が汚い」などの掃除によって改善可能なユーザの不快な状況を発見することを目的とする。本手法の処理過程の概要を図1に示す。

言語情報を使用する理由として、掃除ロボットの特徴の問題がある。掃除ロボットが発する動作音によって音声認識が困難であることや、掃除ロボットの高さが低いことにより、ロボットに設置したカメ

¹ www.twitter.com

² www.irobot.com

ラからの顔認識が困難であることが考えられる。また、複数のユーザが生活している環境ではユーザの要求は対立することがある。例えば一方のユーザは汚い部屋を掃除することを望んでいるが、もう一方のユーザは赤ちゃんが寝ているのでうるさい掃除機を動かして欲しくないと思っている場合、それらの情報を言語情報から得ることで、システムが倫理的な判断を行う際の判断材料とすることが可能になると考えられる。

2. 手法

● Web からの知識収集

我々の従来研究において、掃除の実行をシステムが選択するための知識となる文章を Web から収集する手法の提案を行った[3]。検索エンジンを利用し、「*ので掃除した」というクエリで完全一致検索を行うことで、スニペット中の「ので」の前に含まれる、掃除の理由となる文を収集することが可能となる。例として、「部屋が汚いので掃除した」というスニペットの文から、「部屋が汚い」という文を掃除の理由として抽出する。そのようにして収集された文のセットに含まれている名詞、形容詞に対し、OkapiBM25[4]を利用した解析を行う。これにより、単語の出現頻度が高いほど、かつ検索エンジンでのヒット数が低い、つまり一般性の低い単語ほど高いスコアを得ることになる。結果として、掃除の理由として特徴的な単語を抽出することが可能となる。それらの単語を本稿では「特徴語」と呼ぶこととする。OkapiBM25により高いスコアを得た特徴語と、そのスコアを表1に示す。

この手法により得られた単語をクエリとしてテキストの検索を行うことで、掃除する状況の可能性のある文章を判定することが可能となる。例として、「部屋が汚い」という入力や「教室が埃っぽい」という入力に対して掃除する状況であると判断し、システムがユーザに掃除を行うことを提案することが可能となる。しかしながら、単語のみをクエリとした判定の場合、「字が汚い」という入力に対しても反応してしまうという問題が発生する。このような入力を掃除する状況ではないと判別するためには「何

表 1 特徴語の抽出

| 特徴語 | Okapi BM25 によるスコア |
|-----|-------------------|
| 掃除 | 11.0050506147475 |
| 汚れ | 7.22616194426290 |
| 埃 | 7.02808877680245 |
| ホコリ | 5.96823084475785 |
| 汚い | 5.62837247255482 |
| ゴミ | 5.26486776818813 |

が」汚いのかを調べ、それが掃除可能なものであることを判別する必要がある。そこで次項では、本稿で新たに追加した手法である、係り受け解析と日本語 WordNet[5]を利用した上位概念の検索によりこの問題を解決する手法について述べる。

● 係り受け解析と上位概念の判定

上項で述べた手法によって掃除する状況の可能性があると判定された文に対して CaboCha³を利用した係り受け解析を行うことで、特徴語の前後に係る名詞を抽出する。例として「部屋が盛大に汚い」という入力に対し、特徴後である「汚い」に係る名詞「部屋」と「盛大」が抽出される。

さらに、上述した方法によって得られた名詞に対し、シソーラスである日本語 WordNet を利用することで上位概念の検索を行う。本稿では、その名詞の上位概念に「営造物」、「調度品」、「平坦」、「通路」、「勤務場所」が含まれる場合、掃除することが可能な状況であると判断する。例えば「営造物」を上位概念に持つ名詞として「教室」が存在するため、「教室が汚い」という入力に対して掃除が可能な状況であると判定することが可能となる。一方「字が汚い」という入力の場合は「字」の上位概念に設定した上位概念が含まれていないため、掃除する状況ではないと判断することが可能となる。

上位概念の選択は現段階ではヒューリスティックに決定しているため、掃除可能である単語の抜けが発生する場合がある。今後は上位概念の選択について Web を利用した判定やユーザからのフィードバックの利用が必要になると考えられる。

³ <http://chasen.org/~taku/software/cabocha/>

● 文脈に依存する自然性の判定

掃除する状況ではあるが、掃除機は使わない場合が考えられる。例えば「部屋が汚い」という状況は掃除機を使った掃除が自然だと考えられるが、「壁が汚い」という状況の場合は、掃除機の使用は不自然であると考えられる。このような状況の判別を行うことで、システムはユーザの要求について理解しているが、その手段を持っていないことをユーザに伝えることが可能となる。

掃除機で掃除することが可能であるかについての判定を行うため、検索エンジンを利用し「”(名詞)に掃除機をかける””というクエリで完全一致検索を行う。名詞の部分には係り受け解析で得られた名詞を挿入する。「部屋」と「壁」をクエリとした場合のヒット数を比較すると、「”部屋に掃除機をかける””は822ヒットに対し、「”壁に掃除機をかける””は1ヒットとなるためヒット数による判別を行うことが可能である。しかしながら、一般性が低い単語の場合は掃除機をかけることが自然な場所であってもヒット数が少なくなってしまう。例えば「”アトリエに掃除機をかける””はヒット無しとなってしまう、壁とアトリエの区別を付けることが不可能となる。よって、単語の一般性を考慮したヒット数のしきい値を設定するか、上位概念を利用してより一般的な単語に変換するなどの工夫が必要になると考えられる。

3. 評価実験

係り受け解析と上位概念の判定をシステムに実装し、システムが判断した掃除可能な状況の精度を明らかにするため評価実験を行った。Twitterの検索を利用し、「汚い」をクエリとして検索した結果得られた100文に対し、システムと被験者がそれぞれ、掃除することが可能な状況であるかそうでないかの判定を行う。被験者の判定を正解と設定することでシステムの適合率、再現率、F値を評価する。被験者は、本研究室の大学院生の男性3名と大学生の女性1名である。システムは100文のTwitterの書き込みを解析した結果、10文を掃除可能な状況であると判断した。システムが正解と判断した文章の例を表2に示す。過半数より多い3人以上の被験者が掃

表2 システムの判定例

| |
|------------------------|
| 和子の部屋初登場。 けっこう汚い |
| 部屋が汚いけれど片付ける気が沸かないので放置 |
| 汚い部屋なう |
| 風呂とキッチンが汚いのは本当嫌だねえ |

表3 3人以上の評価が一致した文についての結果

| | 3人以上が正解 | その他 |
|----------|---------|-----|
| システムが正解 | 10 | 0 |
| システムが不正解 | 8 | 82 |

| 適合率 | 再現率 | F 値 |
|------|------|------|
| 1.00 | 0.56 | 0.72 |

除可能な状況と判断した文を正解とした結果を表3に示す。システムが正解と判断した10文はすべて3人以上の被験者によって正解と判定された。また、3人以上が正解と判定した18文のうち8文を、システムは不正解と判定した。

4. 考察

今回の結果から、係り受けを解析し係り受け先の名詞の上位概念を調べることで、高い精度で掃除可能な状況の判定が可能であることが明らかになった。提案手法により、「先生、字が汚いです!!」や「大人って汚い」といった入力を弾くことが可能となったため、適合率が上昇したと考えられる。

一方、ユーザは正解と判断したがシステムは不正解と判定し、失敗した文例を表4に示す。失敗した原因として、係り受け解析の失敗が挙げられる。例えば「部屋汚い」などのように助詞落ちが発生する場合は係り受け解析の失敗が起こる。また、「研究室が汚い」等を解析した場合「研究」と「室」で分かれてしまうため、連続する名詞はまとめる処理をしているが、そのために「十中八九部屋が汚い」の場合に「十中八九部屋」でまとってしまうため上位概念の獲得に失敗してしまう場合が存在する。他には、「部屋がカオス! (汚い)」のような、単純に係り受け解析だけでは対処できない問題も存在する。これらの問題を解決することが、より再現率を向上

表 4 システムが判定に失敗した文例

| |
|------------------------|
| 部屋汚い |
| まあ十中八九部屋が汚いせいですが |
| ギャー！部屋がカオス！（汚い） |
| 風呂は汚いんで、時間も無いしとりあえず... |

させるために必要になると考えられる。

また、掃除可能な上位概念の選択がヒューリスティックなため、掃除可能な単語の抜けが発生する場合や、単語自体が WordNet に未登録の場合がある。例えば「風呂は汚いんで」の場合の「風呂」は现阶段では WordNet に登録されていない。その場合は WordNet で検索可能な単語（「風呂」の場合は「風呂場」）に置き換えるか、ユーザからのフィードバックを得て学習するなどの機能が必要となると考えられる。また、掃除すべき状況だが「汚い」等の特徴語を含まない場合がある。例として、「タマゴが割れた」という文からは掃除する状況を連想することが可能であるが、現状のシステムでは発見することが不可能である。解決のためには、「汚い」という概念に関する知識が必要となるため、ConceptNet[6]等のオントロジーの利用が考えられる。

5. まとめ

本稿では、インターネット資源を利用してユーザの要求を Twitter の書き込みから自動的に発見し、解決する手段を提供するシステムの手法について述べた。掃除を望む状況の判断のため、掃除する状況を表す文を Web 上から収集し、特徴語の抽出を行った。特徴語の係り受けを解析し、得られた単語の上位概念を、WordNet を利用して調べることで、単語のみでは誤りが発生する文の判断が可能となった。また、掃除機による掃除が自然な状況かどうかを調べるための手法と、問題点について述べた。係り受け解析と上位概念の判定を実装したシステムの精度を調べるための評価実験を行い、提案手法によって高い精度での判定が可能であることが明らかになった。さらに、システムが発見した文、提案手法により弾くことが可能となった文、システムが発見することが不可能であった文について考察を行った。

今後の予定として、提案手法を実装したロボットによる、複数のユーザがいる環境内での動作実験を計画している。また、提案手法における係り受け解析の問題点や、上位概念の設定の見直しを行うことで、システムの精度の向上を目指す予定である。

参考文献

- [1] 佐竹聡, 神田崇行, ディラン グラース, 塩見昌裕, 石黒浩, 萩田紀博: 環境情報を理解してサービス提供を行うロボットの実現, 情報処理学会論文誌, Vol. 51, No. 2, pp. 290-300, 2010
- [2] Keisuke Takagi, Rafal Rzepka, Kenji Araki: Method for Social Behavior Acquisition by Robots Using Feedback from the Environment and Users, IWMST2010, pp. 164-166, 2010
- [3] Keisuke Takagi, Rafal Rzepka, Kenji Araki: Just Keep Tweeting, Dear: Web-Mining Methods for Helping a Social Robot Understand User Needs, AAAI 2011 Spring Symposia (SS-11-05), pp. 60-65, 2011
- [4] Robertson, S.E., Walker, S., Jones, S., Hancock-Beaulieu, M.M., Gatford, M.: Okapi at TREC-3, 3rd Text Retrieval Conference, pp. 109126, 1995
- [5] Kow Kuroda, Francis Bond, Kentaro Torisawa: Why Wikipedia needs to make Friends with WordNet, 5th International Conference of the Global WordNet Association (GWC-2010), pp 9-16, 2010
- [6] Havasi, C., Speer, R. and Alonso, J.: ConceptNet3: a Flexible, Multilingual Semantic Network for Common Sense Knowledge, Recent Advances in Natural Languages Processing, pp. 277-293, 2007

連絡先

北海道大学大学院情報科学研究科

言語メディア学研究室

高木慧佑

E-mail: eyu.8055@media.eng.hokudai.ac.jp