

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 3 日現在

機関番号：16101

研究種目：若手研究(B)

研究期間：2012～2014

課題番号：24700209

研究課題名(和文) 語彙から受ける印象の知識構築法と曖昧な表現における意図理解への応用

研究課題名(英文) A method of knowledge construction for impression felt from the word and application to intention understanding in vague expression

研究代表者

森田 和宏 (MORITA, Kazuhiro)

徳島大学・ソシオテクノサイエンス研究部・准教授

研究者番号：20325252

交付決定額(研究期間全体)：(直接経費) 2,000,000円

研究成果の概要(和文)：人に優しいマンマシンインターフェースの実現には、人間からの要求を正しく理解する技術が必要であるが、曖昧な表現から意図を理解するには、表層の単語が持つ意味とは異なる解釈が必要となるため、感性や印象などの感覚を知識化する必要がある。本研究では、語彙から受ける印象を知識ベースとして構築する技術を確立し、意図理解への応用に向けた印象知識の有用性を評価した。

研究成果の概要(英文)：The technology that correctly understands the demand of person is necessary for the achievement of a gentle man-machine interface to human. An interpretation different from the meaning of the word on the surface is needed to understand the intention from a vague expression. Therefore, the sensibility, the impression and so on should be built as knowledge. This research established the technology that constructed the impression felt from the word as a knowledge base. The utility of the impression knowledge for the application to the intention understanding was evaluated.

研究分野：総合領域

キーワード：感性情報処理 印象知識 知識獲得

1. 研究開始当初の背景

人に優しいマンマシンインターフェースの実現には、人間からの要求を正しく理解する技術が必要不可欠である。自然言語処理分野においては、概念ベースやオントロジーなどの知識構築によって、語の持つ本来の意味概念を蓄積することで意味理解を行う。例えば「コウノトリが運ぶ」の理解処理では、意味概念の取得によって「鳥」+「運ぶ」となるが、この場合は「幸せ」+「運ぶ」が正しい理解だと考えられる。このような曖昧な表現から意味や意図を理解するには、表層の単語が持つ意味とは異なる解釈が必要となり、そのためには感性や印象などの感覚を知識化する必要がある。

2. 研究の目的

- (1) 語彙と複数の印象情報、その関係を紐付けし、相互に検索可能な辞書構造を考案する。
- (2) 語彙から受ける印象情報を収集し、Webコーパスから印象情報を自動獲得するためのテキストの傾向を調査する。
- (3) Webテキストコーパスからの印象情報獲得法を提案し、印象知識ベースを構築する。
- (4) 曖昧な表現における意図理解へ応用するため、印象知識の有用性を実験により確認する。

3. 研究の方法

- (1) 語彙と複数の印象情報を紐付ける辞書構造の考案と開発

語彙と印象を全て索引語として辞書に保存するために、一般的な索引語の格納形式であるトライ構造と、その効率的なデータ構造であるダブル配列の研究成果を利用する。索引語同士を紐付ける手法は、共起情報を効率よく格納する辞書構造の研究成果を発展させる。この研究成果の辞書は、共起する二つの索引語に関係する情報を、付属情報として保存できる構造であり、付属情報の記憶領域を拡張することで、本研究における印象知識（複数の索引語に関係する情報や、一つの関係間にある複数の付属情報）の保存が可能となる。しかし、紐付けられた語彙（索引語）群の相互検索には非効率である（ほぼ線形の探索になる）ので、この問題を解決する辞書構造を考案し、辞書エンジンのプロトタイプを開発する。

- (2) 語彙から受ける印象情報収集とWebテキストにおける傾向の調査

まずは人手による印象情報の収集を実施する。収集は、作業者が自由に選択した名詞に対して、その名詞から受ける印象や、連想する表現を書き出すことにより行う。例えば「マーチ」であれば「車」「行進曲」「3月」「コアラ」「大学」、「インフルエンザ」であれば「注射」「痛い」「鳥」「感染」「タミフル」のように書き出し、1万語程度を目標に収集する。次に、収集した表現に対してWeb検索

を行い、語彙（名詞）と収集表現間の共起度などを調査して関連性を探る。例えば「マーチ」を検索すると、ほとんどが「車」に関する結果となり、「行進曲」などは少ない。このことからJaccard係数などを利用して関連性の強さを数値化することで印象の傾向を探る。また「インフルエンザ」のように検索しても印象表現が結果に表れない場合もある（主に最新情報や治療法が結果に表れる）ので、どうすれば印象表現を含む検索結果となるのかを調査する。

- (3) Webテキストからの印象情報獲得法の提案と知識ベースの構築

Webテキストの傾向は、「小さいおサイフケータイ」「ケータイは軽いのが一番」「ケータイ電話」など、係り受け関係、あるいは意味的共起関係にあることが予想される。そこで、共起関係からの概念構築手法や、分野連想語獲得手法などの研究成果を応用し、テキストから意味的共起関係を抽出・解析することで印象情報を獲得する手法を提案する。この手法を印象知識構築エンジンとして実装し、Webテキストからの知識ベース構築を実施する。

- (4) 曖昧な表現における意図理解への応用に向けた印象知識の有用性評価

例えば「ケータイが大きい」の場合、主体語「ケータイ」の印象「小さい」「軽い」などと、対象語「大きい」との違いを検出することによって「不評」など意図の取得が可能になると考えられる。そこで、この違いの有無を主体語と対象語の関連度を計測することで取得する手法を研究する。また、印象知識の有用性確認のため、関連度取得手法を評価する。具体的には質問応答や感性コミュニケーションへの応用を念頭に置き、評判情報や感性検索へ関連度取得手法を適用することで評価する。

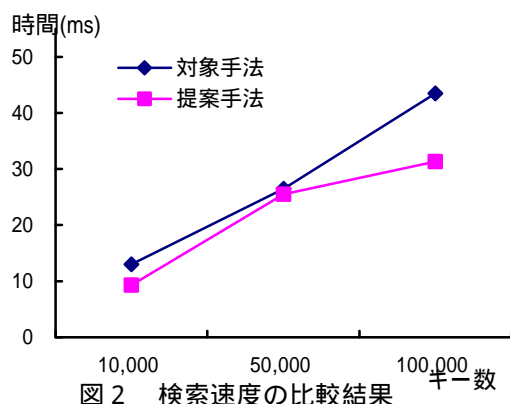
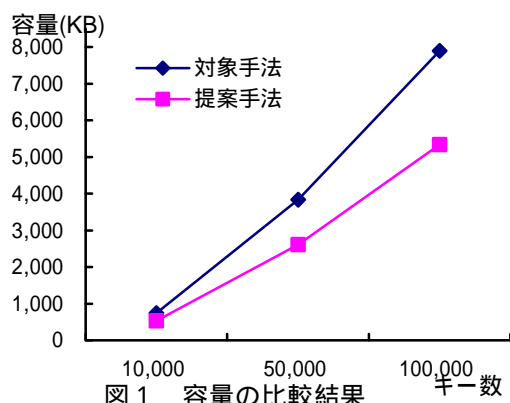
4. 研究成果

- (1) 語彙と複数の印象情報、その関係を紐付ける辞書構造の開発

まず、一般的な索引語の格納形式であるトライ構造と、その効率的なデータ構造であるダブル配列の研究成果を活用し、語彙と印象を全て索引語として辞書に保存する手法を考案した。次に、索引語同士を紐付けるために、共起情報を効率よく格納する辞書構造の研究成果を発展させ、複数の索引語に関係する情報を、付属情報として保存できる構造を考案した。この手法は、紐付けられた索引語群の相互検索を可能にするとともに、辞書サイズを抑える工夫を施すことで、登録される情報の増加に対処された構造となっている。また、この辞書構造を実現する登録・検索エンジンのプロトタイプを開発した。

プロトタイプを用いて、提案する辞書構造

の評価をおこなった。結果を図1, 図2に示す。対象手法には従来の共起情報を格納する辞書に, 提案辞書構造と同様の相互検索に必要な情報を付与したものを使用した。図1は辞書容量の比較結果であるが, 登録キー数が増加するほど効率よく格納できていることがわかる。また, 図2は索引語に紐付けられた印象情報を全て取り出す際の検索速度の比較結果であり, 対象手法と同等の速度を維持していることがわかる。



(2) 語彙から受ける印象情報収集と Web テキストにおける傾向の調査

人手による印象情報の収集を実施し, 約3万表現を収集した。表1に収集した印象情報の例を示す。収集した印象情報には, 連想する表現として「ケータイ」に対する「小さい」「軽い」「電話」など, 係り受けや共起関係にあると考えられるものの他, 「WEB」「ポケット」など, 直接的な関係がわかりにくいものや, 「カメラ」「電卓」など別の語から「ケータイ」を連想するなど向きがあるもの, などの特徴が見られた。

この収集した印象情報について, 語彙と印象の組み合わせで主にマイクロブログを Web 検索し, 得られたテキストを調査した。語彙と印象の組み合わせによっては検索結果を得られない場合も存在した。マイクロブログでは現在話題の事柄がより多く拡散する, つまり検索結果を得られやすい傾向にあるため, 結果を得られない組み合わせは印象として適切ではないと考えられる。

表1 印象情報の収集例

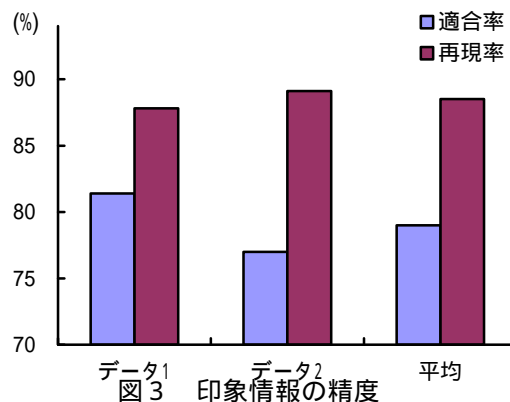
語彙	印象情報
イヤリング	アクセサリー, ピアス, 痛い, 疲れる, 落とす, 無くす, かわいい, ファッション, おしゃれ, 綺麗, キラキラ, 女性
カメラ	写真, 携帯電話, ケータイ, スマートフォン, 現像, ビデオ, フラッシュ, レンズ, アルバム, 写メ, フィルム
コウノトリ	鳥, 白い, 綺麗, 幸せ, 野生, 黒
たけのこ	里, 美味しい, 中華, 八宝菜, きこの山, 竹
ドライブ	自動車, 車, 遠出, 気晴らし, 回転
流行	ファッション, 髪型, 言葉, 人気, インフルエンザ, 風邪, トレンド

(3) 大規模テキストからの印象情報獲得法の考案と知識ベースの構築

収集した印象情報と Web テキストの調査をもとに, 共起関係からの概念構築手法の研究成果や分野連想語獲得手法の研究成果を応用して, 印象情報獲得法を考案した。

手法は起点となる語彙から Web テキストを取得し, 多属性照合技術を用いて関連する印象情報を抽出する。ここで, 抽出した印象情報を起点にさらに Web テキストを取得し, 印象情報を抽出することで, 元の語彙に対する印象情報を拡充する。

このプロセスが予想以上に時間を要することとなったため, 小規模知識として観光地名から受ける印象を利用した評価表現の収集, 知識ベース化をおこない, 印象知識獲得法の評価をおこなった。図3に獲得した印象情報の精度を示す。特に再現率が高いことから, 漏れなく収集できていると考えられるため, 効率のよい知識構築手法が実現できているといえる。



(4) 意図理解への応用に向けた印象知識の有用性評価

“ケータイが大きい”の場合の主体語「ケータイ」と対象語「大きい」との関連度を計測する手法を考案した。関連度は「ケータイ」の印象「小さい」「軽い」などを次元に対応させたベクトルを生成する。「大きい」も同様にベクトルを生成し、両者の類似度から計測する。関連度が高いと似た印象を持つことになるため、ベクトル成分の類似性を基に意図の取得ができる。

関連度取得手法を評価するため、印象情報獲得法を利用して楽曲名から受ける印象を小規模知識ベースとして構築し、印象語との関連度から楽曲の検索を試みた。知識ベース構築、検索の過程では楽曲のサウンド情報は全く用いていないので、語彙から受ける印象の検索への応用ともなっている。

表2は、楽曲名に対する印象付与精度を示している。良好な結果を得られているため、印象付与の対象が制限されることなく、印象情報獲得法の応用性が期待できる。また、図4は印象をもとに楽曲検索し、提示(推薦)された結果の満足度(5段階評価)を示している。良好な結果であることから、関連度取得法の有用性が示せたといえる。

表2 印象付与の精度

楽曲	件数	正解数	精度(%)
1	189	163	86.2
2	245	180	73.5
3	133	123	92.5
4	419	354	84.5
5	114	102	89.5
6	89	81	91.0
7	249	169	67.9
8	107	94	87.9
9	139	134	96.4
10	61	53	86.9
合計	1745	1453	83.3

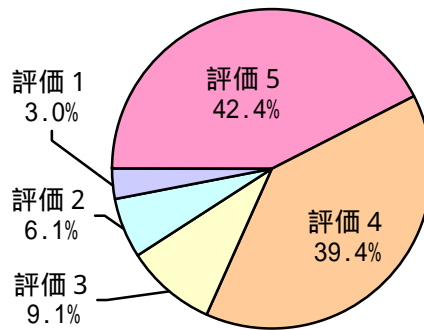


図4 検索結果の満足度

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表](計 1 件)

(1) Kazuhiro Morita, A Method of Music Retrieval Using Impression Keywords, 7th International Conference on Computer Science and Information Technology (ICCSIT2014), 2014.12.23, Barcelona, Spain.

6. 研究組織

(1) 研究代表者

森田 和宏 (MORITA, Kazuhiro)

徳島大学・大学院ソシオテクノサイエンス
研究部・准教授

研究者番号：20325252