

感情辞書を用いた提示システムの提案

二瓶 俊昭[†] 呉 一郎[‡] 林 隆史[‡]
 会津大学コンピュータ理工学研究科

1. はじめに

近年、人々はインターネットを通じ、より簡単に自分に関する情報を共有することが出来るようになってきている。この流れは最も初期のインターネットから始まり、Web2.0と言われるようになった現在、多くのシステムにより更に加速している。Web2.0の最たるものがブログ、SNS、Tumblerなどである。我々はこの流れは止まらなると予想している。Web2.0当初では、使用する人がそれぞれ情報を共有することを目的に情報を発信していたがコミュニティ活動などが活発になった。最近だと、システム側では、単にユーザの情報発信を待つのではなく、システムが得た情報を元に積極的な提案をすることによりユーザの活動を促進する流れも生まれている。例えば、Amazonのシステムで見ることが出来る「このページを見た人はこのようなものを買っています」という提案や、Facebook内の友達関係を利用して、既知の友達の存在を知らせたりする機能である。このような、情報発信の簡易化・促進する流れのなかで、感情情報の重要性が増していくとの予測の上で本研究がなされている。感情情報が何故重要性を持つのか、それは感情は人々の生活に密接に関係があるからである。例えば、落ち込んでいるときに活発に活動が出来ないように。それゆえに、もしある言葉や物と感情の関係が明らかになれば、人々をより充実した生活に導くことが出来る可能性がある。

そこで、我々は感情辞書を作成し、それを使った提示方法を提案する。感情辞書の内容は、単語それぞれに感情情報をマッピングして保存する。感情辞書を使う際には、どの単語にどの感情がどの程度結びついているかを容易に知ることが出来る。単語と単語の関係を調査することにより、全体の傾向やその使用者それぞれの傾向も予測できることが可能だと予測している。また、この感情情報システムの内容をAPIで提供することにより幅広いエリアで感情情報を使ったマッシュアップが可能となる。

2. 感情辞書について

通常の辞書とは異なり、この感情辞書は動的な辞書である。辞書で引ける情報は季節や流行、環境などによって変化することがある。感情情報を扱う上では非常にシンプルなモデルを採用し、利便性を考慮した。方法はまず感情情報とそれに付随するテキストコンテンツを取得する。この情報を取得する方法として株式会社

Eyes, JAPANで提供している *Emotion.al* システムを用いた。Emotion.alは簡単に言うとTwitterのようなコンテンツ情報+感情情報を扱うことのできるシステムである。次にテキストコンテンツを処理し、感情情報とマッピングするという方法をとる。感情辞書作成に使用するデータは、ユーザ名、感情情報、単語情報、時間情報である。

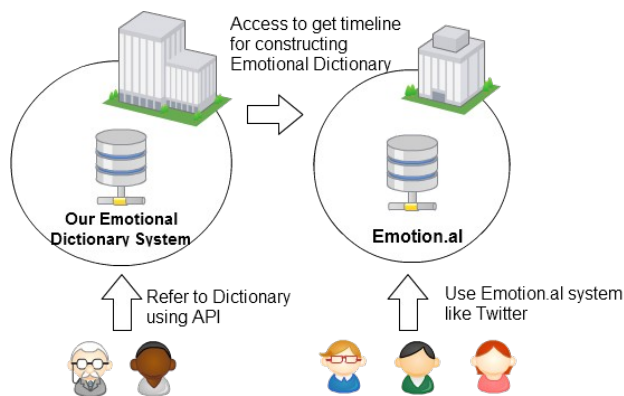


図1 感情情報の取得から提供までのフロー

3. 実装

3.1. データ取得

Emotion.alシステムのAPIを使用しXMLフォーマットで取得する。下記のような種類の情報が提供されている。

- ・テキストコンテンツ
- ・情動情報（快・不快・なしの3種類）
- ・感情情報（怒り、喜び等、約40種類）
- ・登録日時情報
- ・ユーザ情報

3.2. 取得したテキストコンテンツ

3.1で取得したテキストデータを単語毎に分割し感情情報をマッピング。今回使用するテキストコンテンツは日本語を対象としており、単語の分割・抽出の際には形態素解析ツールの一つであるChasenを使用。

3.3. 生成データの保存

今回は感情辞書作成にはRelational Database Management System (RDBMS)を使用。RDBMSはデータの追加、編集、抽出にコストが大きい為、将来的には、Key-Valueストア型のDBを使用することが必要となる可能性がある。今回は実装の容易性の為、にRDBMSを使用。感情辞書構築の為のデータ構造は以下である。

Proposal about Overture system using emotion dictionary

[†] Toshiaki Nihei, Yilang Wu, Takafumi Hayashi

表 1 使用するテーブル一覧

Name	Description
Record	感情情報をまとめる為のテーブル
Object	単語情報用テーブル
Type	単語のタイプ情報用テーブル
Property	付加的な単語情報用テーブル
Record_ Object	レコードとオブジェクトのマッピングに関するテーブル
Object_T ype	オブジェクトのタイプを管理する為のテーブル
Object_P roperty	オブジェクトの付加的な単語情報を管理するテーブル

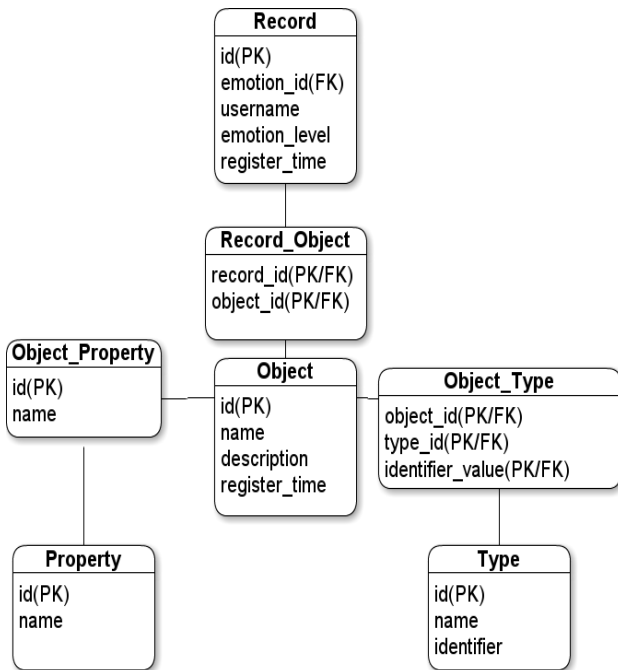


図 2 感情辞書システムのデータ構造

3.4. API としての提供

感情辞書を使い計算された情報を API として提供。本システムが提供する情報は以下。

- ・ 単語情報
- ・ それぞれの感情情報と単語がマッピングされた回数
- ・ それぞれの単語情報が使用された日時情報

感情システムは期間に動的なシステムと言う事で、幾つかの情報抽出指定パラメータを用意することにより使い勝手の向上を図った。それぞれのパラメータは任意で組み合わせて使用することが可能である。現在用意されているパラメータは以下の3種類。

- ・ 期間
- ・ 単語

- ・ 感情情報

4. 感情辞書を使用して情報の使用例

ある期間と単語をキーに、一定の期間の、ある商品名についての感情情報一覧を取得し、最も多い感情を抽出する。次に抽出した感情情報をキーとして情報を取得し、同じ・似た感情を持つ商品名を抽出し新たな提案をすることが可能。また、提案だけではなく商品等について利用すれば顧客満足度の調査等にも役に立つと考えられる。

5. 評価

3500 件程度の情報を元に感情辞書システムを作成し、単語毎にそれぞれの感情がマッピングされた回数を出してみた所、ある程度の偏りが見られる事がわかった。この偏り情報を使用することにより、ある単語とある単語に関連を見つけ出し、人々へ新たな提案をすることにより行動を促進することが可能であると考えられる。今回は数は少ないが、感情の登録の散らばりが大きかった単語の一例を挙げる。

図. 感情辞書に登録されている件数例

単語	喜び	怒り	悲しみ
祭り	4	0	0
雨	1	4	9
腹	2	5	23

6. まとめと今後の課題

今回の調査では単語毎にある程度のどの感情が発生するかについて、偏りが発生することがわかった。今後は、感情辞書システムで提供するユーザ情報も使用して、ユーザの一連の単語登録の流れとどういう順に感情が変化していくかについても研究の余地が考えられる。

また、今回の研究では触れなかったが TF-IDF アルゴリズムを使用することで単に文字を分割するだけでなく重要とみなされる単語を抽出して作成する手法も現在検討中である。

謝辞

本研究は、株式会社 Eyes, JAPAN のシステム、Emotion.al を使用。また、この会社からはその他にも多くの援助を受けており、この場を借り感謝の意を表します。

参考文献

- 1) ChaSen -- 形態素解析器
<http://chasen-legacy.sourceforge.jp/>
- 2) 福原 知宏, 中川 裕志, 西田 豊明
「感情表現と用語のクラスタリングを用いた時系列テキスト集合からの話題検出」
- 3) Emotion.al
<http://emotion.al/>