

# 評価表現に基づく飲食店評判マルチファセット検索システム

2004MT091 佐藤 誠也 2004MT096 祖父江 達師 2005MT031 稲垣 諭

指導教員 河野 浩之

## 1. はじめに

近年、インターネットの普及に伴い、飲食店検索サイトの有用性は益々高くなっている。しかし、それらのサイトは評判情報の入手において問題がある。

本研究では評価表現に基づく飲食店マルチファセット検索システムを提案する。Blog や評判サイトの記述に基づいて評判情報を肯定的な評判と否定的な評判に分類を行う藤村ら[2]の手法を利用し、また飲食店サイト「グルメびあ」で用いられている、目的指向に基づいたマルチファセット型検索システムを実装する。さらに Google Maps API を用いて検索結果を地図上に反映させる事でより分かりやすくユーザーに店舗情報を提供する。

## 2. 飲食店検索サイトの現状と関連技術

### 2.1. 飲食店検索サイトの現状と課題

飲食店検索サイトは、飲食店の情報を入手する為の Web サービスである。「Yahoo」「goo」等のような、ポータルサイトをはじめ、日本のサイトでは「ぐるなび」「食べログ.com」等、アメリカ合衆国のサイトでは「RESTAURANT.COM」「ZAGAT」等がある。表 1 は主な飲食店検索サイトについての表示方法と評価と登録店舗数、レビュー数を示す。

地図から検索表示できるものは店舗の場所の確認をしや

すく付加情報の比較がしやすい。評価は他サイトとのマッシュアップを利用しているところもある。多くの側面において採点する多軸的採点も特徴的である。その他の特徴として、「グルメびあ」「ZAGAT」は目的指向型であるマルチファセット検索、「EatSpot」「Nile Guide」はタブ切り替えで動的に地図に情報表示、「食べログ.com」は Blog へのリンクが可能である。

その中で評判面においては、サイト内の各飲食店のページ毎にユーザーがレビューや採点を行い、他のユーザーに提供している。この方法だと、具体的に評判を知ろうとする際にレビューを読み比べなくては行けない。レビュー数が多いと、読んで評判を判断することが困難であり、すべて読むことも容易ではない。またレビューと採点に個人差があるので、判断基準にも差が生じる。店舗を比較する際にも複数の飲食店の評価情報を得るには手間が掛かる。また一つのサイトにおいて書かれたレビューのみを扱っていることで、飲食店によってはレビューの数が少ないことや無い場合がある。

### 2.2. マルチファセット検索

マルチファセット検索とは、意味による検索を行うために検索対象の特徴を表したものであるファセットを用いて、階層的にクエリを構築し検索を行う検索方法のことである。この検索方法を使用したシステムとして、ヘルシンキ大学博

表 1: 飲食店検索サイト比較(※J=日本, U=アメリカ合衆国)

飲食店サイト	表示	評価		その他の特徴	登録店舗数 /レビュー数
		レビュー	採点		
グルメびあ (J)	リスト			マルチファセット	
EatSpot (J)	リスト+地図	マッシュアップ (食べログ.com)		条件タブで動的に 地図表示	マッシュアップ (食べログ.com)
食べログ.com (J)	リスト+地図	✓	✓	Blog へリンク	182413 / 567786 (2008年9月)
ぐるなび (J)	リスト/地図	✓			約 60000 / 不明 (2006年12月)
the Menu Pages (U)	リスト	✓	✓		
ZAGAT (U)	リスト/地図	✓	✓	マルチファセット	
Nile Guide (U)	リスト+地図	✓	✓	条件タブで動的に 地図表示	
RESTAURANTS.COM (U)	リスト	✓	✓		

物館のフォトレポジトリに提示されている Ontogator[1], Blog の情報コンテンツの特徴に基づいた Blog 検索システムの BLOGRANGER[3]などが提案されている。

この検索システムは図 1 のように表すことができる。ファセットとは検索対象の特徴を表しており、飲食店であれば「料理のジャンル」などが当てはまる。クエリとはファセットに基づいたデータベースへの問い合わせ項目であり、ファセットが「料理のジャンル」の場合は「和食」などとなる。マルチファセット検索において、ユーザはこの複数の意味により分けられたファセットを選び、各々のファセットに応じたクエリを用いて、データベースに対して検索を行う事ができる。

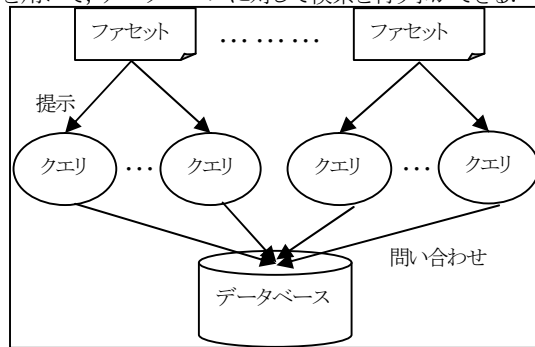


図 1:マルチファセット検索のクエリの構造

### 3. 評価表現抽出における先行研究

藤村らは語のスコアを計算し、電子掲示板の評判情報を文単位で肯定と否定に分類する手法を提案している。彼らの研究では「価格.com」のノート PC に関する掲示板の 2003 年度の書き込み、肯定的な評判 935 件、否定的な評判 551 件を用いている。この評判を基に肯定的な評判には肯定的な概念、否定的な評判には否定的な概念を持った語が多く含まれているはずであるという仮定を置き、肯定的な評判と否定的な評判の差をとっている。

$$Score_p(w_i) = \frac{P_{pos}(w_i) - P_{neg}(w_i)}{P_{pos}(w_i) + P_{neg}(w_i) + k} \quad (1)$$

$$(-1 \leq score(w_i) \leq 1)$$

式(1)は、 $P_{pos}(w_i)$  が肯定的な評判で語  $w_i$  が出現する確率、 $P_{neg}(w_i)$  が、否定的な評判で語  $w_i$  が出現する確率、この確率を利用した語  $w_i$  のスコアを  $Score_p(w_i)$  としている。肯定的な評価表現については正の値をもって、否定的な評価表現については負の値をもって抽出される。  $k$  はスコアが 1/1 となってしまう問題を防ぐための任意の実数である。これにより属性にスコア付けをする。最終的な評価分類は各文書に含まれる語のスコアの総和が 0 より大きければ肯定的とし、0 より小さければ否定的というように分類している。また分類の精度評価を行うために、比較対象の分類アルゴリズムとして SVM (Support Vector Machine) を用いた手法と C4.5 による手法を取り上げている。表 2 では肯定的な評判、否定的な評判ごとに分類精度を測定している。

表 2:分類手法ごとの分類精度(%)

分類手法	肯定的な評判	否定的な評判
C4.5	79.1	60.5
SVM	79.6	71.8
藤村らの手法	83.3	71.3

### 4. 評価表現に基づく飲食店検索システムの構築案

2.1 で述べたように、現状の飲食点検索サイトには便利な機能も多く組み込まれているが、いくつかの問題点も見受けられる。本研究ではそれらの問題点を解決するために、評価表現に基づく飲食店の評判マルチファセット検索システムを構築する。

まず、本研究の飲食店検索システムのデータベースの構築案を図 2 に従って説明する。まず「食ブログ」における飲食店の評判情報を収集し、記述内容に対して形態素解析を行い、3 章で述べた藤村らの分類手法を用い、評価語辞書を作成する。また飲食店に関する情報も収集し、このデータと評価語辞書のデータを格納する事でデータベースを構築する。これらを用いて複数の飲食店情報のファセットを作成し、タブを用いて複数のクエリを提供する。これによりマルチファセット検索を実現し、飲食店の評価情報も含めた検索を行えるようにする。また飲食店の評価情報や他の付加情報を地図上にも反映させ、飲食店の評価情報に加えて位置情報の比較も容易に行う事ができるようにする。

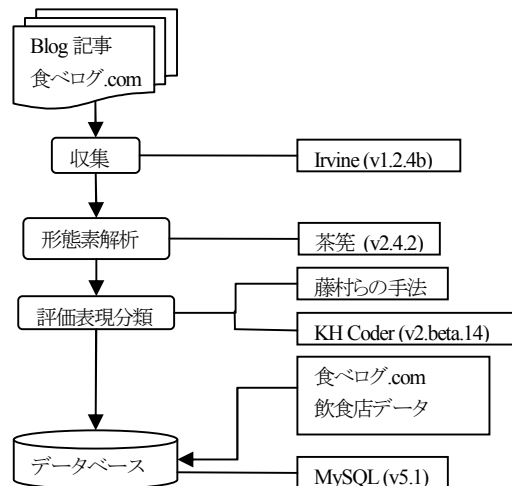


図 2:飲食店評判システムのデータベース構築案

次にシステムとユーザの関係を図 3 に従って説明する。ユーザがファセットタブを選択すると、データベースは複数の飲食店情報のファセットに対応したクエリを提示する。ユーザのクエリ選択によりファセットに応じた検索結果が提示される。これが本システムのマルチファセット検索である。またそれと対応し Google Maps による視覚化を行う。

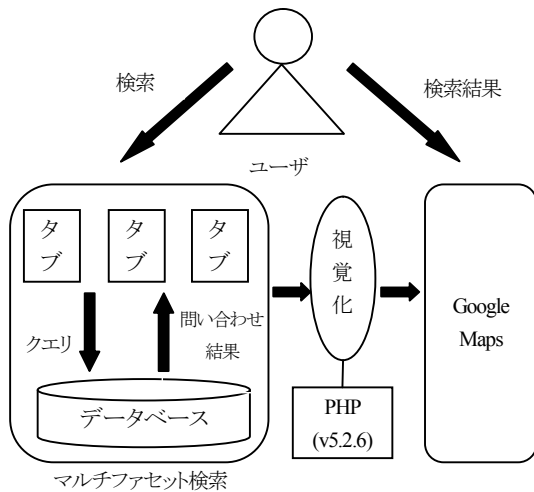


図 3: 飲食店評判検索システムとユーザの関係

## 5. 評価表現に基づく飲食店検索システムの実装

本研究の飲食店検索システムのプロトタイプ作成に当たり、サンプルデータとして「食ブログ」内における愛知県の名古屋駅周辺の飲食店でレビュー記事の数が 20 を超えている店舗を 20 店舗分取り扱った。

### 5.1. 評価表現分類

初めに、評判データは Irvine (v1.2.4) を用いて収集した。20 店舗におけるレビューの総数は 764 個である。収集した評判データに対して、KH Coder (v2.beta.14) を用いて、同ソフトウェア内の茶筌の機能により分かち書きを行い、KH Coder の機能を利用して「おいしい」や「まずい」などの評価語を抽出し csv ファイルに書き出した。それらを Active Perl (v5.8.8.824) で作成したプログラムにより、店舗毎に出力された評価語の集計を、肯定と否定の文書それぞれに対して行う。肯定否定の判断はレビュー記事の採点で 5 段階評価の 3.5 以上を肯定、3.0 以下を否定とした。各文書数は 549 と 215 である。

次に、藤村らの手法を Active Perl で実現し、評価語のスコアリングを行い、評価語辞書の作成を行う。図 4 はプログラム内のスコアを計算する箇所である。肯定的文書と否定的文書をそれぞれ取り込み、肯定的文書で語が出る確率を肯定文書のデータの格納した %tmp\_datas を用いて \$sref\_data へ、否定的文書の語が出る確率である \$sprbb\_n を、語の出現数 \$sword\_n[2] を文書数 \$snum\_n で割る事で求める。肯定的文書内にデータがなかった場合に \$sprbb\_p=0 とし、分子 (0-\$sprbb\_n) を分母 (\$sprbb\_n + \$k) で割りスコア (\$score) とその評価語 (\$shyoukago) を出力する。データがあった場合はそのままスコアと評価語を出力し、分子が 0 の場合はスコアを 0 とし評価語を出力する。

```
while(my $line = <IN>){
  chomp($line);
  my (@word_n) = split(/,/,$line,3);
  my $sref_data = $tmp_datas{$sword_n[0]};
  $sprbb_n = $sword_n[2] / $snum_n;
  if($sref_data == NULL) {
    $score = (0-$sprbb_n) / ($sprbb_n + $k);
    $shyoukago {$sword_n[0]} = [$score];
  }
  else {
    my $sbunsi = $sref_data->[0] - $sprbb_n;
    my $sbunbo = $sref_data->[0] + $sprbb_n + $k;
    if($sbunsi == 0){
      $score = 0;
    }
    else {
      $score = $sbunsi / $sbunbo;
    }
    $shyoukago {$sword_n[0]} = [$score];
  }
}
```

図 4: 評価語分類プログラム

### 5.2. 検索システムの各プログラムについて

システムの構築の際、index.html, map.js, map.php, add\_evl.php, hyouka.php という 6 つのファイルを作成した。図 5 と図 6 の実行例に基づいて説明する。

#### ● index.html

map.js をインクルードし、マルチファセット検索とマップ表示がなされる。マルチファセット検索の部分では各タブに料理ジャンル、予算帯、評価語というファセットを割り当て、Tab Content script v2.2 を使用することにより動的なタブを実現する。それぞれのタブの表示部分にクエリが記述されている。選択されたクエリを map.js を用いて map.php に送る。また add\_evl.php へのリンクがありページを移動する。これが図 5 の検索システムと地図の部分である。

#### ● map.js

Google Maps API を利用し、基礎となる map を構築する。また map.php から受け取った XML データを基に、マーカーを色分けし、地図上に適切な位置に表示し、マーカークリック時に表示される情報ウインドウを作成する。

#### ● map.php

index.html のクエリによりデータベースと接続し、クエリに応じた飲食店 XML データを map.js の関数で利用する。

#### ● add\_evl.php

index.html からリンクされており、map.js をインクルードしている。map.php から map.js により飲食店の ID を受け取る。データベースに接続し、その飲食店の ID を基に評価語によるラジオボタンを作成し、選択されると hyouka.php

から各評価語の飲食店の評判の情報源についての XML データを受け取り、各々のリンクを表示する。図 6 の評判元レビューのリンクと点数の表示部分である。

● hyouka.php

飲食店の評判の情報源についての XML データを、データベースに接続して作成する。



図 5: 実行例 1(検索システムと地図の部分)

評価語を選択してください

- 指定なし
- 絶妙  素晴らしい  大好き  シンプル  美味い
- 控えめ  独特  丁寧  まるやか  濃厚
- 甘い  有名  普通  懐かしい  独特

評判元のレビュー	点数
<a href="#">山本屋本店 エスカ店</a>	-0.124878361821175
<a href="#">おいしい名古屋の味。</a>	0.175188057124815
<a href="#">山本屋本店 エスカ店</a>	0.334374494850636

図 6: 実行例 2(評判元レビューのリンクと点数の表示部分)

## 6. 検索システムの考察評価

本研究では、システムの評価を既存の飲食店検索サイトの「食ベログ」との比較をアンケートによって行う。アンケート回答者は研究室の学生 15 名である。アンケート内容は本システムと「食ベログ」との比較をしてもらい、次のような項目について 10 段階評価を行ってもらった。

1. 評判情報の入手
  - (A) 飲食店の評価を分かりやすさ
  - (B) 飲食店の評判入手の素早さ
  - (C) 情報源文章と比較しての本研究の採点の妥当性
2. システム全体
  - (D) 操作性
  - (E) 満足感
3. 感想

アンケートの結果を平均点で表すと表 3 のようになった。アンケート結果より、評判情報に関しては、マーカーの色や採点などから素早く情報を取得できるという高評価を得られた反面、採点範囲、小数点の位の設定、店舗に関する写真などの情報、悪い評価語の表示、と言った課題も見られた。

表 3: アンケート結果

1. 評判情報	本研究	食ベログ	差
(A)	6.93	6.53	0.40
(B)	7.40	5.87	1.53
(C)	6.21	6.14	0.07
2. システム全体	本研究	食ベログ	差
(D)	7.60	6.87	0.73
(E)	7.00	6.80	0.20

それでも 2 章で挙げたように、私達が目標としてきた点においては達成できたのではないかと思います。

システム全体については、操作面では高評価が得られたが、満足度としてはそれほど高い結果は得られなかった。プロトタイプシステムなので、店舗情報や周辺情報が少ないために、全体としてはあまり満足感を得られているとは言えないが、評判情報についての評価から見ても、店舗数や周辺情報を増やすことで、一般的な飲食店検索システムとしてもおおよそ利用できるのではないかと考えられる。

## 7. まとめ

本研究では、ユーザがより手早く正確に評判情報を取得するために、マルチファセット検索システムを用いた飲食店検索システムのプロトタイプ作製を行った。既存の飲食店検索サイトから 20 店舗分の評判を収集、評価語の抽出を行い、それを利用することで、現状の課題の改善を図った。アンケートの結果、既存の飲食店検索システムと比べて、高い評価を得ることができた。特に、評判情報の入手に関しては、高い評価を得られた。店舗情報を増やしユーザが使いやすいシステムにすることで、実用的な検索システムとしても作動できるであろう。

## 参考文献

- [1] E. Hyvonen, S. Saarela, and K. Viljanen, "Ontogator : Combining View- and Ontology-Based Search with Semantic Browsing," Proceedings of XML Finland 2003, Kuopio, Finland, October 30-31, 2003, [http://www.cs.helsinki.fi/~u/eahyvone/publications/xml\\_nland2003/yomXMLFinland2003.pdf](http://www.cs.helsinki.fi/~u/eahyvone/publications/xml_nland2003/yomXMLFinland2003.pdf) (accessed 2008.9).
- [2] 藤村滋, 豊田正史, 喜連川優, "電子掲示板からの評価表現および評判情報の抽出," 人工知能学会第 18 回全国大会, 3F1-03, 2004.
- [3] 戸田浩之, 藤村考, 井上孝史, 廣嶋伸章, 杉崎正之, 片岡良治, 奥雅博, "目的指向型 Blog 検索システム BLOGRANGER の提案およびユーザ評価," 情報処理学会論文誌: データベース, Vol.48, No.SIG14 (TOD35), pp.132-151, 2007年9月.