

評判に基づくモバイル端末向け施設検索システムの設計と実装

Design and Implementation of Reputation based Facility Locator System

大野木 碧
Midori Onogi

内田 渉
Wataru Uchida

山田 直治
Naoharu Yamada

磯田 佳徳
Yoshinori Isoda

NTTドコモ サービス&ソリューション開発部
Service and Solution Development Department, NTT DOCOMO, Inc.

Recently, the number of users searching facilities based on reputations has been increasing. When a user searches facilities based on reputations, he/she should chose the closest reputation to his/her preference from a large number of expressions. In ordinary, users have to clarify their preferences and find the closest reputation by checking many reputations which are associated with each facility page. However, under the scene that the users should search outside using mobile phone, it becomes difficult because of the constraints of mobile phone's specs (e.g. screen size) and time pressure. Therefore, we designed and implemented the system to recommend the reputations whose concept are similar to the reputation selected by the user in order to let him/her explore and find the closer reputation to their preferences without searching facilities and examining all reputations associated with them. In this paper, we define two aspects of the recommendation and propose the methods to extract the suitable reputation depending on the users' search experience and knowledge of the area are described. The evaluation results show the validity of our methods.

1. はじめに

近年、ブログや口コミサイトなどのCGM(Consumer Generated Media)に投稿された評判を参考に施設を検索するユーザが増加している。例えば、CGM サイトの一つである食べログ[食べログ]は2010年4月現在、57万件以上の飲食系の施設に関して、133万件以上の評判記事を保有しており、月間総ページビューは約3億に達し、過去2年間で3倍以上の伸びを記録している。施設検索における評判は、駅からの距離、価格などの基本的な情報に加えて、施設検索において重要な情報としてとらえられていると考えられる。

駅からの距離、価格、などの基本的な情報に関する条件は、そのときのユーザがいるエリアや訪問予定のエリア、予算などに応じて制約条件として決まってくるケースが多い。しかし、評判に関しては、評価の観点が「味」の場合は「激辛で美味しい」、「デザートが美味しい」、評価の観点が「雰囲気」の場合は「大人の雰囲気」、「気軽に入れる雰囲気」、評価の観点が「同行者」の場合は「家族向け」、「友達同士に向いている」など、表現が多岐にわたる。そのため、評判に基づいて施設を検索する場合は、これらの評判の中でどの評判を重視するかを決めることが重要なタスクとなる。

我々は、ユーザが評判に基づき施設を探索する際に、施設につけられた評判を比較検討することで、自分の嗜好に最も合う評判を決定していると考え、施設に共起する評判を連想評判候補としてユーザに提示し、自分に合った施設の検索を支援するシステムを設計・実装した。本稿では、連想評判に求められる条件として、特定の評判がつけられる施設での共起しやすさを表す特異性と、ユーザに発見を与える性質を表す発見性を定義し、ユーザの特性に応じて適した連想評判候補を提示する方式を提案する。

以下、2章でユーザが重視する評判の探索プロセスについて述べる。3章では関連研究について述べる。4章では実装した連想評判候補提示システムについて述べる。5章で連想評判の

要件と、連想評判候補の提示方式について述べる。6章で提示方式の評価の結果について述べる。7章では、結論を述べる。

2. ユーザが重視する評判の探索プロセス

一般的な施設検索サービスでは、施設ごとに評判が投稿されている。そのため、自分が重視する評判がまだ明確になっていないユーザは、いくつかの施設につけられた評判を確認する中で、自分が重視する評判を決めていくことになる。図1に一般的な施設検索サービスの施設と評判の情報構造と、ユーザが施設検索において最も重視する評判を決定するステップの例を示す。

ユーザが自分の嗜好にマッチする評判を探索する場合、以下のようなステップをたどると考えられる。

- S-1. ある施設の評判を確認する。
- S-2. 上記の施設の評判の中から、自分の嗜好に近い評判を選択する。
- S-3. 同じ評判が投稿されている他の施設を、【S-2】で得た評判を仮の検索条件として使うなどの方法で発見し、その施設に投稿された評判を確認する。
- S-4. 前回選択した評判より自分の嗜好に近い評判があった場合はその評判を選択し、それが投稿された施設を再度探していく。以下、ユーザは自分の嗜好に最も合う評判を見つけるまで【S-3】、【S-4】の手順を繰り返す。

上記のステップを、図1を使って例示する。施設①を選択し、施設①に投稿された評判A(センスがいい)、評判B(大人の雰囲気)、評判C(いい雰囲気)を確認する【S-1】。施設①に投稿された評判の中で、評判B(大人の雰囲気)が最も自分の嗜好に近いと判断する【S-2】。同じ評判B(大人の雰囲気)が投稿されている施設②を発見し、そのほかに投稿された評判D(高級感がある)を確認する【S-3】。最初に選択した評判B(大人の雰囲気)と評判D(高級感がある)を比較して、評判Dの方が自分の嗜好にマッチしていると判断し、評判Dが投稿された別の施設③を探していく【S-4】。施設③に投稿された評判D、E、Fを確認し【S-3】最も自分の嗜好に近い評判Eを選択し【S-4】、評判の探索を終了する。

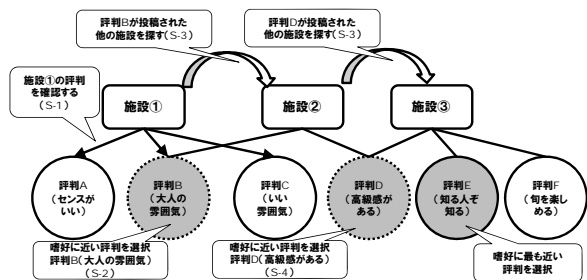


図1 ユーザが重視する評判の探索プロセスの例

このように、ユーザは評判と、その評判が投稿された施設を繰り返したり、少しずつ自分の嗜好により近い評判に更新していくことで、多種多様な評判の中から自分が重視する評判を絞り込むと考えられる。

上記のプロセスにおいてユーザが自分の嗜好に最も近い評判に近づくことができるのは、同じ施設につけられる複数の評判には、その施設のある側面についての様々な角度からの意見が含まれるため、今まで最も自分の嗜好に近いと思っていた評判とそれらを比べて、どちらがより「近い」か「近くない」かが判断できるためと考える。図1の例では、ユーザが「大人の雰囲気」から「高級感がある」、「知人ぞ知る」と比較しながら評判を選び、段階を追って自分の嗜好に近い「知人ぞ知る」という評判に近づいている。本稿ではこのように共通した側面に対する異なる2つの評判の関係を連想評判と呼ぶことにする。

一方、評判に基づく施設検索は、家の中だけではなく、外出先で携帯電話などのモバイル端末を利用して行われる場合もある。現在普及している施設検索サービスでは、前述のとおり施設に対して書き込まれた評判を確認することができるが、モバイルで施設に投稿されている数多くの評判を確認し、確認した評判に基づいて他の施設の評判を確認する、という手順を繰り返すのは、ユーザにとって負担が大きい上、時間もかかるため、外出先というシーンには適していない。

そこで、ユーザの操作の負担軽減と検索時間の短縮のため、対象の評判が投稿されている施設と同じ施設に投稿される傾向がある評判を、「学生に人気」といえば「行列ができる」というように直接連想評判の候補として提示し、S-3のような施設の確認を行わずに、ユーザの嗜好にマッチした評判を効率的に検索できるようにするシステムを設計、実装した。本稿では、本システムにおいて、ユーザが選択した評判に対する連想評判候補を提示する方式を提案する。

3. 関連研究

従来にも、関連する情報を繰り返し提示し、ユーザの情報探索を支援する理論やシステムが提案されてきた。高野らは、文書中のキーワードの共起回数から関連するキーワードや文書を検索するエンジン GETA[高野 04]を開発している。GETAはキーワードが文書に出現する回数を単語ベクトルとして保存し、キーワードの共起関係に基づいて文書ごとに単語ベクトルを圧縮することで、関連するキーワードや文書を高速に検索できるシステムである。しかし、ここで共起回数に基づいて単語間の関連性を計算すると、どの単語とも共起する単語が繰り返し出やすくなり、本稿の想定環境のように、ある特定の単語への絞込みを効率的に行いたい場合には適していない。

ある条件のもとで発生しやすい事象を算出するために、リフトとよばれる指標が提案されている[Geng 06]。リフトを用いることで、どこにでも良く現れる事象を除去し、元の事象と特に同時に発生しやすい事象を抽出できる。施設検索に適用した場合は、ある評判がつけられた施設に対してのみつけられることの多い評判を抽出することができる。

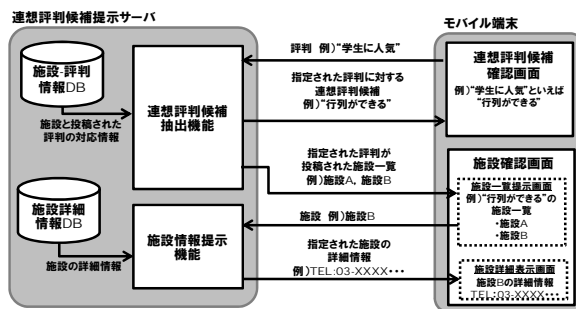


図2 連想評判候補提示システム

一方、上記の共起回数やリフトのように、事象間の関係性を評価するのではなく、推薦が行われるユーザにとっての推薦効果を評価するための観点として、発見性を導入した方法も提案されている。Zieglerらは、推薦されるアイテムのジャンルが多様であることを発見性が高いこととみなし、ジャンルの多様性を考慮する手法を提案している[Ziegler 05]。本やCDなどの商品のように、ジャンル別に分類された膨大なアイテムの中からユーザに発見を与えるアイテムを推薦するようなシーンでは有効な手法であるが、本稿で想定する評判のように系統だったジャンルが定義されていないデータには適していない。

Hijikataらは、ユーザが知らないアイテムは発見性が高いとし、アイテムに関する既知・不既知のプロファイルとユーザの嗜好プロファイルを組み合わせてアイテムの推薦を行う方法を提案している [Hijikata 09]。ユーザの既知・不既知のプロファイルを入手できるような環境では、有効な手法である。

4. 連想評判候補提示システム

図2に実装した連想評判候補提示システムを示す。施設を特定するIDとその施設に投稿された評判情報のセットを施設-評判情報DBに、施設の詳細情報は施設詳細情報DBにそれぞれ格納している。これらのデータは施設検索サービス「ドコイク?」[ドコイク?]のWebAPIを利用して収集した。サーバの連想評判候補抽出機能で、ユーザが指定した評判に対する連想評判候補を抽出し、モバイル端末に提示する。さらに施設情報提示機能で、ユーザが選んだ評判に対応する施設の詳細情報を、モバイル端末に提示できるシステムになっている。

5. 連想評判候補の提示方式

5.1 連想評判の要件

ある評判について重視しているユーザに対して、そのユーザが最も重視する評判により近い評判の候補としての提示が有効な評判は、以下の2つのうち、両方もしくはどちらか一方の特性をもつ評判と考えられる。

1. 特有性をもつ評判

2章で述べたとおり、ユーザは同じ施設に投稿された評判を確認しながら評判を繰り返したり、自分の嗜好に最も近い評判を探索していく。ここで、元の評判が投稿された施設につけられた他の評判でも、自分の嗜好に最も近い評判に近づくために必ずしも有効ではないものも存在する。どのような施設にもつけられる傾向がある「おいしい」といった評判は、例えば「大人の雰囲気」、「大衆的な雰囲気」など、相反する嗜好に基づいていると考えられるどちらの評判についても、関連性があるととらえられてしまう可能性がある。「大人の雰囲気」に対しては「フレンチがおいしい」、「大衆的な雰囲気」に対しては「ラーメンがおいしい」など、対象の評判にのみ関連性のある連想評判候補を提示する必要がある。そのため、どの施設にも出現する一般的な評判ではなく、元の評判が出現する施設で特に共起しやすくなる評判(特有な評判)が前述の要件を満たすと考えられる。

2. 発見性をもつ評判

一方で、毎回同じ連想評判が出ていると、ユーザは同じ評判からしか施設を検索することができない。評判に基づく施設検索に慣れていないユーザや、過去に対象のエリアに訪問したことがあるユーザに対しては、ユーザがまだ検討したことがない発見性のある評判が効果的になると考えられる。

これらの特有性、発見性の観点からユーザの施設検索の習熟度やエリアに関する知識、土地勘によって重視される度合いが異なる。例えば、評判に基づく施設検索の経験やエリアに関する知識が少ないユーザは、評判の種類やエリアの評判の傾向を把握できていない可能性がある。このようなユーザに対しては、特有性の高い連想評判候補を提示することで、ユーザが重視する評判に効率的に到達できるようになると考えられる。一方、施設検索経験やエリアの知識がある程度あるユーザには、特有性のある評判に加えて、新しい発見が期待できる発見性のある連想評判候補が推薦されることが求められると考えられる。

そこで本稿の提案システムでは、上記のようなユーザの特性に応じて使い分けられるような、特有性と発見性、もしくはどちらか一方を考慮した連想評判のスコアを複数定義し、連想評判候補を提示する方式を採用する。

5.2 連想評判候補の提示方式の提案

方式1. 特有性を重視した連想評判候補の提示方式

評判に基づく施設検索の経験やエリアに関する知識が少ないユーザには特有性を重視した連想評判候補を提示する。本稿では、共起関係の特有性を考慮した指標であるリフトを用い、その値が上位の評判を、特有性の高い連想評判候補として提示する。評判 A と評判 B のリフト $lift_{AB}$ は評判 B の出現確率 $P(B)$ と確信度 $P(B|A)$ の比 $P(B|A)/P(B)$ として表される。 $P(B)$ は、全施設数 N に対して評判 B が出現する施設数を $n(B)$ とすると $P(B)=n(B)/N$ となる。また、評判 A と評判 B が共起する施設数を $n(A,B)$ とすると、 $P(B|A)=n(A,B)/n(A)$ として表される。

方式2. 特有性と発見性を重視した連想評判候補の提示方式

評判に基づく施設検索の経験者や訪問するエリアに対してある程度知識があるユーザに対しては、特有性のある評判に加えて発見性のある連想評判候補を提示する。

3章で述べたとおり、Hijikataらは既知・不既知のプロファイルに基づき発見性の高いアイテムを推薦する手法を提案しており、本稿の評判に基づく施設検索システムの場合でも、ユーザがそれまでに検討したことがない評判を推薦することで、ユーザの新しい発見につなげられる可能性があると考えられる。しかし、施設検索システムでは検索対象のエリアでユーザがそのシステムを何度も利用しているとは限らないため、ユーザの評判の既知・不既知のプロファイルを入手することは難しい。そこで、アイテムの逆出現頻度を表す IDF [Jones 72]を用いる。

IDF は特定の文書に出現しやすい特徴的なキーワードを抽出する指標として、主に自然言語処理の分野で用いられている。施設における評判 A の IDF は $idf_A = \log(N/n(A))$ と定義できる。これを利用して、 IDF が低い、つまりどの施設にも投稿される傾向が強い評判は、ユーザが既知である可能性が高いとみなし、アイテムの既知・不既知のプロファイルがない状況でも評判の発見性を推定できると考えられる。

本稿では、下記のように特有性の指標リフトおよび発見性を考慮した指標 IDF を組み合わせた評判 A と評判 B の連想スコア $score_{AB}$ を定義する。

$$score_{AB} = \alpha \cdot \frac{lift_{AB}}{N} + (1 - \alpha) \cdot idf_B \quad (0 \leq \alpha \leq 1) \quad (1)$$

表 1 評価データの概要

対象エリア	関東(東京都, 神奈川県, 千葉県, 埼玉県, 茨城県, 栃木県, 群馬県)
施設数	1,112,760 件
評判数	142,398 件

表 2 集約した評判一覧

いい雰囲気	ふらっと立寄る	行列ができる
いつか行きたい	ヘルシー	高級感漂う
オトナの雰囲気	ホッと一息	子供 OK
ガールズトーク♪	ボリューム満点	旬を楽しむ
かわいい	また行きたい	接客◎
こだわり	ランチ	知る人ぞ知る
これはうまい	わいわい	駐車場あり
スイーツ	一度は行くべき	眺めが良い
センスがいい	温泉	入りやすい
タバコあり	家族でお出かけ	評判の店
テイクアウト	学生に人気	品揃え豊富☆
デート♪	個室あり	夜遅くまで営業
とにかく安い	交通の便◎	

それぞれ $lift_{AB}$, idf_B が 0 から 1 の値をとるように正規化を行っている。 α はパラメータであり、発見性を重視したい場合は小さな値を、特有性を重視したい場合は大きな値を設定することで、方式の動作を調整することができる。

6. 評価

本章では 5.2 章で述べた連想評判候補の提示方式に関して、特有性・発見性の関係性を評価する。

6.1 評価データの概要

評価には施設検索サービス”ドコイク?”[ドコイク?]の施設・評判データを用いた。ドコイク?では施設名に対してユーザが評判を表すタグを複数付与することができるようになっており、これらのデータを表 1 のとおり Web API を用いて収集した。また、本評価で収集した評判は、同じ意味でも異なる表現がある場合も含んでいる。そこで、表 2 のとおり類似する評判を 38 個の評判に集約した。集約された評判数 69,110 件であり、それ以外の評判はノイズとして除去した。

6.2 連想評判の提示方式の評価

評価1. 特有性を重視した連想評判候補の提示方式

方式 1 において、リフトが特有性を考慮した連想評判候補の抽出において有効かどうかを評価した。比較対象として、共起回数のみを考慮した確信度の高い順に連想評判候補を提示する方式を用いた。本評価では、被験者 10 人に対して特と思われる評判の組をアンケートにより調査し、正解データを作成した。 i 人以上が指定した評判を特有性の高い連想評判とし、それぞれの方式の推薦結果に含まれる数を測定し、その適合率と再現率を評価した。

図 3 に $i=1, 4, 10$ についてリフトと確信度の値で順位付けした連想評判候補の適合率、再現率を示す。 $i=1$ の場合を除いて、リフトを用いて算出した方が適合率、再現率ともに高くなっている。 $i=1$ の場合は、1 人でも特有性がある支持した場合も正解となるので、多くの評判が特有な連想評判に指定される。そのため、どちらの指標で算出しても上位に正解の評判が含まれやすくなり、リフトと確信度で傾向の違いが出なかったものと考えられる。

方式 1 においては、一定の人数が特有性があると支持する評判について、リフトは確信度よりも、特有な連想評判の提示に有効であるといえる。

評価2. 特有性と発見性を重視した連想評判候補の提示方式

方式 2 の連想スコアについて評価する。 $\alpha=1, 0.99998, 0.99996$ について推薦された上位 x 件の連想評判候補の idf_B の平均値を比較した結果を図 4 に示す。 α を減少させることで発見性を考慮した指標 idf_B の平均値が上昇していることが分か

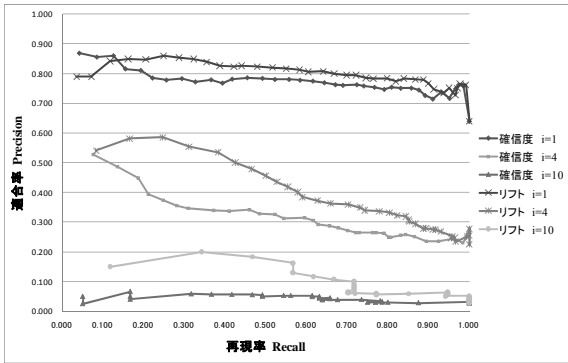


図3 特有性のある連想評判の適合率/再現率

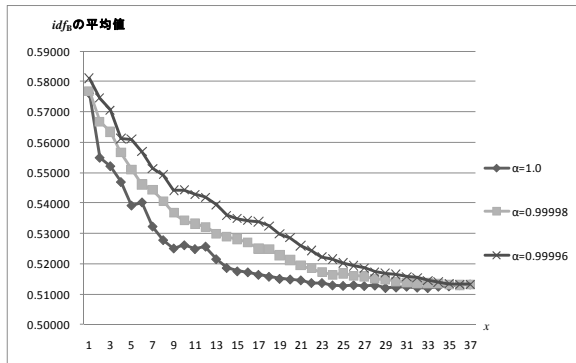


図4 パラメータ α と idf_B の平均値の関係

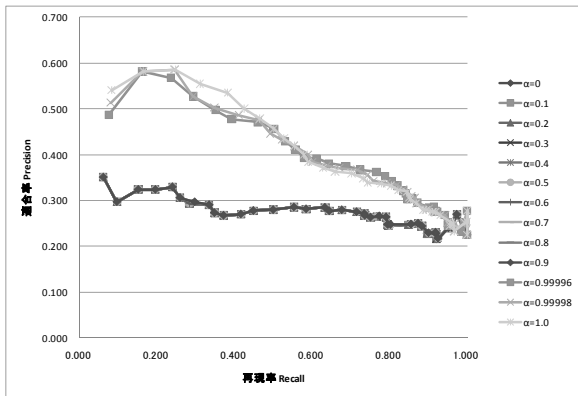


図5 特有性の連想評判と α の関係

る。また、同様に α を変化させたときの連想評判の具体例を表3に示す。表3は、評判「こだわり」に対する連想評判候補を式(1)の連想スコアが高い順に並べたものである。「ガールズトーク」「学生に人気」などの idf_B の高い連想評判の順位が上がっていることが分かる。さらに、発見性を考慮した IDF を連想スコアに反映することによる、特有性のある連想評判の再現率・適合率への影響を検証した。図5に評価1で特有の評判と判定した人数 $i=4$ の場合の正解データについて、 α を0.1刻みで変化させた場合の適合率・再現率を示す。 $\alpha=0.9$ までは $\alpha=0$ の場合とほとんど値が変わらず、 $\alpha=0.99998, 0.99996$ 付近で $\alpha=1$ のリフトのみで連想評判候補を出した場合と適合率・再現率がほぼ同じ値になる。これは、式(1)の連想スコアにおいて、特有性を考慮した $lift_{A,B} / N$ の値に対して発見性を考慮した idf_B の値が大きい数値に分布しやすいためである。以上のことから、 α を変化させることで、特有性を考慮した連想評判候補を提示しながら、同時にユーザーに新しい発見を与える可能性の高い評判を連想評判候補として提示することが可能になることがいえる。

表3 α と評判「こだわり」に対する連想評判候補の比較

$\alpha=1$		$\alpha=0.99998$		$\alpha=0.99996$	
連想評判候補	idf_B	連想評判候補	idf_B	連想評判候補	idf_B
オトナの雰囲気	0.568405	オトナの雰囲気	0.568405	知る人ぞ知る	0.572210
知る人ぞ知る	0.572210	知る人ぞ知る	0.572210	オトナの雰囲気	0.568405
また行きたい	0.580694	また行きたい	0.580694	また行きたい	0.580694
行列ができる	0.582604	行列ができる	0.582604	行列ができる	0.582604
評判の店	0.640582	評判の店	0.640582	評判の店	0.640582
デート♪	0.494300	デート♪	0.494300	ガールズトーク♪	0.680898
一度は行くべき	0.486500	一度は行くべき	0.486500	デート♪	0.494300
接客◎	0.483896	ガールズトーク♪	0.680898	一度は行くべき	0.486500
ランチ	0.341105	接客◎	0.483896	接客◎	0.483896
ガールズトーク♪	0.680898	ランチ	0.341105	高級感漂う	0.585679
高級感漂う	0.585679	高級感漂う	0.585679	ランチ	0.341105
品揃え豊富☆	0.474237	品揃え豊富☆	0.474237	品揃え豊富☆	0.474237
これほうまい	0.349887	これほうまい	0.349887	いつか行きたい	0.509748
いつか行きたい	0.509748	いつか行きたい	0.509748	これほうまい	0.349887
わいわい	0.512046	わいわい	0.512046	わいわい	0.512046
センスがいい	0.454845	家族でお出かけ	0.532495	学生に人気	0.685252
家族でお出かけ	0.532495	センスがいい	0.454845	家族でお出かけ	0.532495
ホッと一息	0.560873	学生に人気	0.685252	ホッと一息	0.560873
ボリューム満点	0.499991	ホッと一息	0.560873	センスがいい	0.454845
個室あり	0.521775	ボリューム満点	0.499991	個室あり	0.521775
学生に人気	0.685252	個室あり	0.521775	ボリューム満点	0.499991
交通の便◎	0.413671	交通の便◎	0.413671	いい雰囲気	0.631535
入りやすい	0.462179	いい雰囲気	0.631535	スイーツ	0.560083
ヘルシー	0.490264	入りやすい	0.462179	ヘルシー	0.490264
スイーツ	0.560083	ヘルシー	0.490264	かわいい	0.610027

7. 結論

連想評判の候補を提示することでユーザーの施設検索を支援する手法について述べた。連想評判に求められる要件として、特有性と発見性を定義し、ユーザーの評判検索の経験値や訪問エリアの認知度に応じた連想評判候補の提示方式を提案した。評価では、ドコイクの施設データを用いて、特有性を重視した提示方式および特有性と発見性を重視した提示方式について有効性を検証した。今後は実際のシステムで評判に基づく施設検索プロセスの目標達成までの時間や操作性などのユーザービリティの評価を実施する予定である。

参考文献

- [食ベログ] [食ベログ]ランキングと口コミで探せるグルメサイト, URL: <http://tabelog.com/>
- [ドコイク?] ドコイク? 地図・店舗情報の検索エンジン, URL: <http://www.doko.jp/>
- [高野 04] 高野 明彦, 西岡 真吾, 丹羽 芳樹: 連想に基づく情報アクセス技術 汎用連想計算エンジン GETA を用いて, 情報の科学と技術 Volume 54 Issue 12, 社団法人情報科学技術協会, 2004.
- [Geng 06] Liqiang Geng and Howard J. Hamilton : Interestingness Measures for Data Mining A survey, ACM Computing Surveys Volume 38 Issue 3, ACM, 2006.
- [Ziegler 05] Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan and Georg Lausen : Improving Recommendation Lists Through Topic Diversification, Proceedings of the 14th International Conference on World Wide Web, ACM, 2005.
- [Hijikata 09] Yoshinori Hijikata, Takuya Shimizu and Shogo Nishida : Discovery-oriented Collaborative Filtering for Improving User Satisfaction, Proceedings of the 13th International Conference on Intelligent user Interfaces, ACM, 2009.
- [Jones 72] Karen Sparck Jones : A Statistical Interpretation of Term Specificity and its Application in Retrieval, Journal of Documentation Volume 28 Issue 1, MCB UP Ltd., 1972.