

# エリアを要約する地図インターフェイス BEIRA

増谷 修†

## BEIRA - Map User Interface with Area Summary

OSAMU MASUTANI†

### 1. はじめに

カーナビや、Web上の地図サービスなどのいわゆる地図インターフェイスでは、Web上の情報を、通常のキーワード検索や、指定された位置（現在地など）の周辺から検索するというようなことができる。このような検索は予め目的が決まっている場合には適しているが、旅の途中や、なんとなくあるエリアの面白いものを探す、などというときにはどうだろうか。

例えば、自分の好みの検索キー（またはジャンルなど）を指定しても、そのエリアでは魅力的な店が全く無い場合や、見つかった場合でも、必ずしも評判の高い店が見つかる保証は無い。また、満足のいく店が見つかったとしても、もしかするとそのエリアには他に独特のジャンルの店があり、結果的にはそちらに行けばよかったと、後悔するという可能性もある。

このような問題に内在する要因として、情報の全体像が見えないまま自分で決めた検索キーによるフィルタリングをしてしまい、上記のエリア独特の店のようなものは排除され、選択肢に上ってこないという点がある。全データを最初から表示することも可能であるが、昨今のPOI（Point of Interest）データベースは巨大になっており、これを行うと、ディスプレイ上が点やアイコンだらけになり用をなさない可能性が高い。

これを防ぐためには、予めそのエリアでどのような特徴のPOIがあるか、あるいは有名かなどの要約情報を提示することが必要と考えた。これにより、そのエリアでのホットなPOIを逃すことが無くなる。

### 2. BEIRA

このアプリケーションを BEIRA (Bird's Eye Information Retrieval Application) と呼ぶ。図 1 に BEIRA のユーザインターフェイスを示す。左側のペインが地図画面、右側が 2次元チャートによる POI 比較画面となっている。

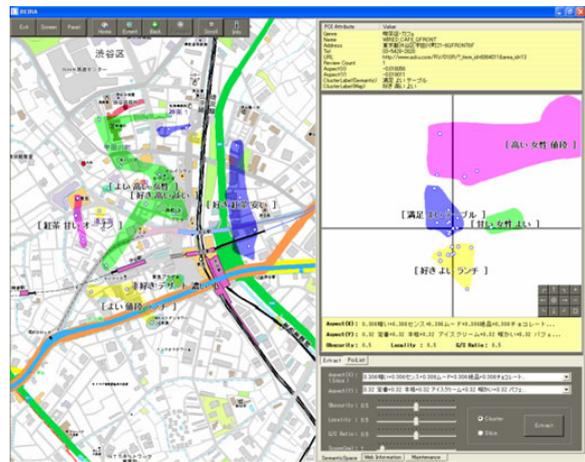


図 1 BEIRA 画面

BEIRA では次のような情報検索が可能である。

- キーワード検索
  - カテゴリ検索
  - 地図上でのエリア要約表示、及びフィルタリング
  - 2次元チャートによる直感的な POI 比較表示
- 要約表示は次節で紹介する BEIRA のコアとなる利用方法で、特徴的なエリアとそのラベルを表示することでユーザの気づきを与える。2次元チャートは、POI を観点空間で表現したものである。基本的には

† (株) デンソーアイティラボラトリー  
Denso IT Laboratory, Inc.

文献<sup>2)</sup>に示される方法に近いもので、観点の自動的抽出、意味(観点)空間上でのクラスタ表示を組み合わせたビューとなっている。観点抽出に、TF/IDFによるベクトル空間モデル、及びLSI(Latent Semantic Indexing)を、クラスタリングにはk-means法を用いている。この2次元チャートにより、ある程度絞り込まれた後のPOI間の比較を一目で行うことが可能になる。通常のPOI検索では最終的には、POIごとの情報ページを一つずつ比較しなければならないのとは対照的である。

### 3. 意味/地理クラスタリング

地図上エリア表示の手順は以下ようになる

- 意味的なまとまりを持つエリアを抽出する(BEIRAエリアとする)
- エリアをラベリングする(BEIRAラベルとする)

通常の地図の場合には行政区画、地勢的区画(平野など)がエリア、地名がラベルとなるが、これらに相当するものを、文書情報から自動的に抽出しようというものである。

この問題を解決するための方法としてクラスタリングを用いた。ここで考えられるクラスタリングには以下の二つがある

- (1) 地図上の位置に基づくG(eo)クラスタリング
  - (2) 文書の意味に基づくS(emantic)クラスタリング
- どちらも確立された技術であるが、そのまま用いたのではそれぞれ以下の問題がある。

- (1) 確かに見た目上まとまったエリアが抽出されるが、意味的なまとまりとは無関係である。
- (2) 意味的なまとまりを捉えることができるが、抽出されるエリアは地理的に相関が無い。

これらの問題を解決するために我々はハイブリッドなアプローチをとった。

地理/意味(G/S)クラスタリングは、地理的な分布と意味的な分布の双方を加味したクラスタリングである。これを行う方法として、地理空間と意味空間を統合したG/S空間(直積空間)を用いる方法を提案する。空間相互の結合比(バイアス)についてG/S比( $r$ )と定義して可変とした。この $r$ を調整することで適切なエリアが得られるのではないかと仮定した。またラベリングには、得られたクラスタ内でTF/IDFによる重要度を基準としたラベル抽出を行った。

### 4. 評価実験結果

このBEIRAについて簡単な評価を行った。POI

データとしてレストラン口コミサイトasku.com<sup>1)</sup>の東京都内のカフェの口コミコンテンツを利用した。カフェを選んだのは、地域の特性を比較的表しやすいと考えたからである。

人手で調査して得た6つの正解クラスタ(渋谷-インターネット、秋葉原-メイドなど良く知られたもの)をどの程度抽出できるかをG/S比( $r$ )を変化させることで計測した。結果のグラフを図2に示す。強いクラスタ("インターネット"、"メイド")は、どのG/S比でも抽出できたが、弱いクラスタ("ラウンジ"、"オーガニック")などは、G/S比が適切で無い場合には抽出できなかった。このデータセットではG/S比10が最も高い性能を持つという結果になった。この値を一般的な適正值ということではできないが、Gのみ、Sのみではうまく捕らえることのできないクラスタを抽出している。

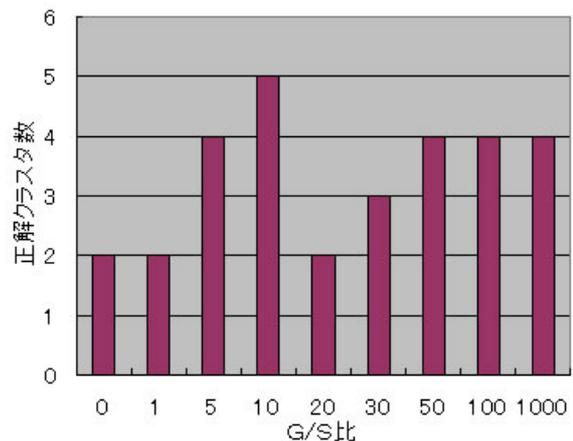


図2 G/S比と正解クラスタ数

今後エリアの精度を上げるために、統計的手法などを用いた改良を行い、評価についても信頼性のある正解データの作成による定量評価と、ユーザテストに基づいた主観評価などを行う予定である。

### 参考文献

- 1) 口コミレストランガイド アスクユー・ドット・コム (<http://www.asku.com>), バンダイネットワークス株式会社.
- 2) 吉田尚史, 関子泰三, 清木康, 北川高嗣: ドキュメントデータ群を対象とした文脈依存動的クラスタリングおよび意味的データマイニング, 情報処理学会論文誌, Vol.41, No.SIG1, (2000).