論　文

情報ナビゲーションへの連想ルールの適用

小山　聡[†]　　石田　亨[†]

Applying Association Rules to Information Navigation

Satoshi OYAMA[†] and Toru ISHIDA[†]

あらまし　本稿では，モバイル環境における WEB 情報の活用に適した新しいインタフェースとして，情報ナビゲーションエージェントを提案する．これまでの情報検索システムは，検索結果のユーザへの表示とそれに基づくクエリの変更を通して多くの情報をいかに絞り込んでいくかが中心的課題であり，デスクトップ環境での利用に適した方法であった．それに対して我々のエージェントは，検索対象の属性の制約条件を用いて近似解を生成し，ユーザとの対話を通して情報を得ることにより，漸近的によりよい情報へとナビゲートしていくことができる．その際に，エージェントがキーワード間の連想ルールを用いて検索条件の解決と新たな検索条件の追加を行う手法を提案した．また，連想ルールを不用意に用いるとナビゲーションの失敗を導くため，統計的検定とグラフ構造の解析を用いた連想ルールの精錬を行うことを提案し，その有効性を確認した．

キーワード　エージェント 情報ナビゲーション 情報検索 連想ルール デジタルシティ

1. まえがき

World Wide Web に大量の情報が蓄積されるに従い，情報過負荷 (information overload) が利用者にとって大きな問題となってきている．現在の WEB における情報検索エンジンでは検索結果の一覧を画面に表示し，それをユーザが一つ一つ確認し，キーワードを追加して再度絞り込み検索を行う必要がある．研究の課題としても，適合フィードバック [1] などを用いて，「いかに情報を適切に絞り込んでいくか」が研究の中心であった．

しかし、これは、典型的には「デスクトップ環境での直接操作インタフェース」を想定した場合に有効な方法である．我々は、例えば日常生活のための情報基盤であるディジタルシティ [2] において，街中での旅行者などへの情報案内システムを構築することを目指している．そこでは，地理的関係演算などを用いて，「○○交差点から 200 メートル以内にある中華料理店で日曜日も空いている店は?」といった問い合わせを可能にするような検索システムも開発されている [3]．このような複雑な検索をモバイル環境において直接操作インタフェースで行うには限界があり，自然言語での問い合わせを取り扱える対話型のエージェントインタフェースが求められる．携帯端末を用いたインターネット利用の急速な拡大にみられるように，今後 WEB 情報のこのような利用形態はますます重要になると考えられる．

そこでは，大量の情報をユーザに提示し，それを絞り込んでいくのではなく，まず具体的な解を 1 つ提示し，ユーザとの対話を通して，「よりよい情報へとナビゲーション」していく研究が必要である．適合フィードバックの一種である漸次的適合フィードバック (incremental relevance feedback) [4] では，検索されたドキュメントの中から，最も適合すると思われるものをユーザに 1 つだけ提示し，それ対してユーザが適合／不適合の判定を行い，システムが検索キーワードの重み付けを変更して再検索を行うという処理を繰り返していく．この方法は，ユーザとのインタラクションが通常の検索システムと比べて単純であり，モバイル環境にも適していると考えられる．ただし，ユーザからのフィードバックが適合／不適合の 2 値しかなく，クエリ全体に対して変更が行われ，個々の検索条件を独立に変更させることができないため，利用者との対話を通じた詳細なナビゲーションを行うことは困難である．

[†] 京都大学大学院情報学研究科社会情報学専攻
Department of Social Informatics, Kyoto University, Kyoto 606-8501, Japan

我々は検索対象となるオブジェクトの属性に対応する検索アスペクト間での連想ルールを利用して，ユーザを要求に近い解へと導いていく方式を導入した．ただし，従来提案されてきた連想ルールは意味的に関連のないキーワード間にも検出される場合があり，不用意に用いるとナビゲーションの失敗を招く．そこで，連想ルールを精錬して用いることとした．

以下，2 章では連想ルールを用いた情報検索について述べ，3 章では情報ナビゲーションエージェントによる連想ルールの利用方法を，4 章では連想ルールの精錬方式を示す．5 章で本手法の実験評価を行い，6 章で関連研究について紹介する．

2. 連想ルールを用いた情報検索

連想ルール (association rule) [5] とはデータマイニングの分野で POS データの解析などに利用されているものである．以下にその定義を示す．

$\mathcal{I} = \{i_1, i_2, \ldots, i_m\}$ をアイテム全体の集合，$\mathcal{D} = \{t_1, t_2, \ldots, t_n\}$ をトランザクションの集合 (データベース) とする．各トランザクション t_j はアイテムの集合 $\mathcal{I}$ の部分集合 $t_j \subseteqq \mathcal{I}$ として表される．アイテム集合 X のすべてのアイテムが，トランザクション t_j に含まれるとき，t_j は X を含む ($X \subseteqq t_j$) という．X の支持度 (support) とは，$\mathcal{D}$ 全体に対する X を含むトランザクションの割合とする．例えば，$X = \{$ビール, おむつ$\}$ としたとき，全体のトランザクションの中で，この 2 つのアイテムを同時に含むものの割合が X の支持度となる．ここで，連想ルールとは，$X \subset \mathcal{I}, Y \subset \mathcal{I}$，$X \cap Y = \emptyset$ であるアイテム集合 X, Y 間に成り立つ，

$$X \Longrightarrow Y$$

の形のルールであり，確信度

$$\mathrm{confidence(X \Longrightarrow Y)} = \frac{\mathrm{support(X \cup Y)}}{\mathrm{support(X)}}$$

および支持度

$$\mathrm{support(X \Longrightarrow Y)} = \frac{\mathrm{support(X \cup Y)}}{|\mathcal{D}|}$$

で評価される．確信度と支持度の下限が与えられたときに，データベースからこれを満たす連想ルールを発見することが，データマイニングにおける問題となる．

データベースを WEB ページの集合，トランザクションを各 WEB ページ，アイテムを WEB ページに含まれるキーワードとすると，連想ルールを情報検索の問題に対応づけることができる [6].

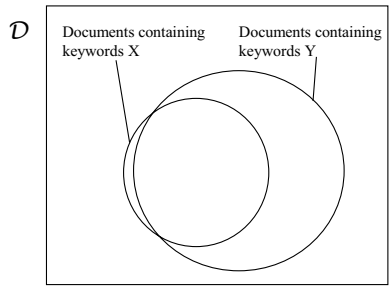

図 1　キーワード間の連想
Fig. 1　Keyword Associations

その場合，図 1のように，キーワード集合 X を含むページの多くがキーワード集合 Y も含む場合，$X \Longrightarrow Y$ の形の連想ルールがなりたつことになる．

X = {“チャーシュー”,“スープ”}, Y = {“ラーメン”} で

$$\mathrm{confidence(X \Longrightarrow Y)} = 0.83$$
$$\mathrm{support(X \Longrightarrow Y)} = 0.1$$

であるとき，$X \Longrightarrow Y$ という連想ルールでは “チャーシュー” と “スープ” というキーワードを含む WEB ページが与えられた場合，“ラーメン” というキーワードを含む可能性が 83 %あり，“チャーシュー” と “スープ” と “ラーメン” というキーワードを含むページが元の WEB ページの集合の中に 10% あることを示している．我々は，このような連想ルールを情報のナビゲーションに用いることを試みる．

また，連想ルールのデータマイニングにおいては，外部の分類階層を併用する方法 [7] も提案されている．この方法を用いれば，概念階層を反映した連想ルールの抽出も可能である．以下では，もっとも単純なキーワードとキーワードの間の連想ルールを，情報ナビゲーションに用いる方法について説明する．その後この手法の評価を行い，今後の拡張の方向について議論する．

3. 情報ナビゲーションエージェントによる連想ルールの利用

3.1 連想ルールを用いたユーザとエージェントのインタラクション

我々の目標としているユーザとエージェントの間の

典型的なインタラクションは、以下のようなものである．

ユーザ:	宴会の店を探しているのですが.
エージェント:	宴会でしたら，しゃぶしゃぶはいかがでしょうか?
ユーザ:	いいですね．どこがありますか?
エージェント:	○○はどうでしょう．宴会ももちろんできます.
ユーザ:	他にはありませんか?
エージェント:	△△はどうでしょう．ここもしゃぶしゃぶの店ですので宴会もできると思います.
ユーザ:	他には?
エージェント:	居酒屋ですが××ではどうでしょうか? ここはお座敷があるので，宴会はできると思います.
ユーザ:	鍋がありますか?
エージェント:	はい，寄せ鍋がメニューにあります.
ユーザ:	詳しく教えてください.
エージェント:	場所は○○交差点の近くで…

上の例のような対話をエージェントが行うには，「お座敷があれば宴会ができるだろう」といったドメイン固有の知識を用いることが必要となる．そこで本論文では、連想ルールを検索知識としてWEBページから抽出し，その知識を用いてユーザに対して明示的に説明を行いながら検索を行う方法を提案する．

3.2 検索アスペクトへの連想ルールの適用

検索対象となるオブジェクトは，検索の場面で特に考慮されるいくつかの属性をもつ．たとえば，飲食店の情報を検索する場合，場所，メニュー，雰囲気，サービスといった条件などを考慮するであろう．我々は，このように特に検索の場面で意識されるオブジェクトの属性を検索アスペクトと呼ぶ．

情報検索を，検索アスペクトに対する制約問題ととらえると，ユーザの検索目標は，検索対象を制約するアスペクトと値の組で表される．「祇園で宴会のできる店をさがす」という例を考えてみる．これは，検索対象となるWEBページの中から，場所，利用状況という二つの異なるアスペクトに対して，[場所]=“祇園” かつ [利用状況]=“宴会” いった制約を満たすものを求める問題である．

問題領域におけるアスペクトの集合を $\mathcal{A}$ とし，それぞれのアスペクト $A \in \mathcal{A}$ に対して $D(A)$ を各アスペクトの取りうる値 (キーワード) の集合とする．検索要求はアスペクトへの値の割り当て

$$A_i = x_i \quad (A_i \in \mathcal{A}, x_i \in D(A_i))$$

の組として与えられるとする．エージェントは現在満たされているアスペクト，まだ満たされていないアスペクトを意識し，ユーザとの対話を通してなるべく多くのアスペクトの制約を満たそうと努める．そのためにエージェントはユーザと対話を行いながら連想ルールを適用して検索を実行する．

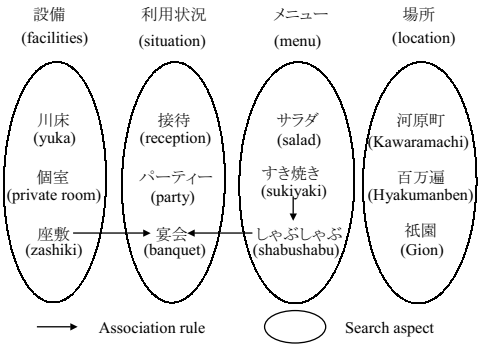

図 2　クエリへの連想ルールの適用
Fig. 2　Applying association rules to a query

例として，{“しゃぶしゃぶ”, “宴会”, “祇園”} という初期クエリが与えられた場合を考える．“しゃぶしゃぶ” は [メニュー]，“宴会” は [利用状況]，“祇園” は [場所] に関する，それぞれ異なった 3 種のアスペクトに属する．ここで，図 2のような連想ルールが存在しているとする．

次の場合が考えられる．

（1） 3 種のアスペクトを満たすページが適量存在する．

（2） 3 種のアスペクトを満たすページは存在しない．

（3） 3 種のアスペクトを満たすページが必要以上に多数存在する．

(1) の場合には，エージェントはユーザに該当のページを推薦すればよい．(2) の場合には，エージェントは検索条件を緩和して近似解を求めるために連想ルールを用いる．そして，その近似解をどのようにして求めたかについて，ユーザに説明を行う．(3) の場合には，エージェントは検索結果の絞り込みのために，新

しいアスペクトの制約の追加をユーザに求めるが，その候補を連想ルールを使って導き出す．

以下では (2)(3) の場合について，それぞれエージェントの動作を説明する．

3.3 アスペクトの制約の解決に用いる場合

WEB ページの場合，記述方法の不統一などから，アスペクトに対応するキーワードがページの中に存在していない場合においても，必ずしもアスペクトが満たされていないとは結論できない．例えば，飲食店の WEB ページに “宴会” というキーワードがなかったとしても，その店で宴会ができないとは必ずしもいえないであろう．例えば，そのお店に座敷があることが分かれば，宴会もできると推測することが可能である．すなわち “座敷”$\Longrightarrow$“宴会” という連想ルールを適用することにより，まだ満たされていないアスペクトを満たすことが可能となる．

エージェントは連想ルールを，以下の順に探索する．

（1） 既にクエリに現れているキーワード間のルール

（2） 既にクエリに現れているキーワードへの，未だクエリに現れていないアスペクトに属するキーワードからのルール

（3） 既にクエリに現れているキーワードへの，同一アスペクトに属する別のキーワードからのルール

現在のクエリを $A_1 = x_1, \ldots, A_n = x_n$ としたとき，(1) の場合は，$x_j \Longrightarrow x_i (1 \leqq i \leqq n, 1 \leqq j \leqq n)$ を適用して $A_i = x_i$ という制約を外して検索を行う．図 2でクエリが [利用状況]=“宴会”，[メニュー]=“しゃぶしゃぶ”，[場所]=“祇園” であったとき，“しゃぶしゃぶ”$\Longrightarrow$“宴会” という連想ルールを用いて，[メニュー]=“しゃぶしゃぶ” と [場所]=“祇園” という条件だけで検索するような場合である．この時のエージェントの説明としては，「祇園の○○があります．しゃぶしゃぶをやっているので，宴会もできると思います」という形になる．

(2) の場合には，未だクエリに含まれていないアスペクトに対応するキーワード x からのルール $x \Longrightarrow x_i$ を利用し，$A_i = x_i$ の条件を外して，$A_{n+1} = x$ という新しい制約を加えて検索を行う．これは，初期クエリが [利用状況]=“宴会” かつ [場所]=“祇園” であるような場合，“座敷”$\Longrightarrow$“宴会” というルールを適用し，[設備]=“座敷” かつ [場所]=“祇園” の組合わせで検索するような場合である．このときエージェントはユーザに対して，「祇園の○○があります．お座敷があるので宴会もできると思います」といった形の説明を行うことになる．

(3) の場合には，$x \in D(A_i)$ である x からの連想ルール $x \Longrightarrow x_i$ を適用し，このアスペクトの制約を $A_i = x$ へと変更して検索を行う．(1) と同様の初期クエリの場合，“すき焼き”$\Longrightarrow$“しゃぶしゃぶ” という連想ルールを用いて，メニューに関するアスペクトの制約を [メニュー]=“すき焼き” に変更して検索し，「祇園の○○があります．宴会ができます．すき焼きがあるので，しゃぶしゃぶも食べられるかもしれません．」といった対話を行う形になる．

3.4 追加のアスペクト制約を提案する場合

初期クエリによる検索結果が多すぎる場合，検索制約を強化して絞り込みを行う必要がある．まだ制約条件が指定されていないアスペクトを加えることにより，検索結果を絞り込める可能性がある．その際に，エージェントは連想ルールを使って，制約の候補を提示することができる．現在のクエリを $A_1 = x_1, \ldots, A_n = x_n$ としたとき，この中に現れないアスペクトに属するキーワード x からの連想ルール $x \Longrightarrow x_i$ を適用して，

$$A_{n+1} = x$$

を新たに検索条件に追加することを提案する．例えば，[サービス]=“宴会” かつ [メニュー]=“しゃぶしゃぶ” という条件で検索して結果が多すぎる場合，“座敷”$\Longrightarrow$“宴会” というルールが成立していれば，[設備]=“座敷” を追加の制約として提案することができる．このとき，エージェントはユーザに対して，「お座敷があった方がいいでしょうか?」と提案する形で尋ねることになる．

4. 連想ルールの精錬

情報ナビゲーションの成功には，連想ルールの質が重要である．WEB ドキュメントの集合においては，意味的に関係のないキーワード間にも連想ルールが多く導かれるため，情報ナビゲーションに有効なルールを抽出する目的で，以下の方式を採用した．

4.1 統計的検定によるルールの絞り込み

ここでは，統計的な有意性の観点から不要なルールを削除することを行う．例えば，ドキュメント集合全体 $\mathcal{D}$ のうち，キーワード x を含むものが 30%，キーワード y を含むものが 100% であったとする．ここで，連想ルール $x \Longrightarrow y$ において確信度が 1.0 となるが，

x と y との出現に関係があるとは言えない．すなわち，

$$\mathrm{confidence}(x \Longrightarrow y) \simeq \mathrm{support}(y)$$

のような場合には，このルールはナビゲーションに用いることはできない．

このような観点からルールの有意性を確認するには，ドキュメント集合全体 $\mathcal{D}$ におけるキーワード x の出現と y の出現が独立であるかどうかを，以下に示す統計学の χ^2 検定の手法を用いて確認すればよい [8]．これは，決定木の学習において，不適切な属性の採用を防ぐ χ^2 枝刈り [9] と呼ばれる方法と同じ考えに基づいている．統計的検定を用いる考えは，詳しい手法は定かではないが，文献 [7] にも述べられている．そこでは，POS データに対してはこの方法が有効ではないと述べられている．我々は，この方法が WEB において効果的であることを実験によって示す．

母集団が 2 つの性質 A, B の両方において，互いに背反な m 個のクラス $A_1, \ldots, A_m$ と n 個のクラス $B_1, \ldots, B_n$ に分けられており，大きさ N の標本における，各クラスの観測度数が，$a_1, \ldots, a_m$ および $b_1, \ldots, b_n$ であるとする．ここで，クラス「A_i かつ B_j」の観測度数が $x_{ij} (i = 1, \ldots, m; j = 1, \ldots, n)$ であったとする．（ここで，$\sum_{j=1}^{n} x_{ij} = a_i, \sum_{i=1}^{m} x_{ij} = b_j$．）このとき

$$\chi_0^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(x_{ij} - a_i b_j / N)^2}{a_i b_j / N} \qquad (1)$$

は自由度 $(m-1)(n-1)$ の χ^2 分布に従う．

表 1の例を考える．ここでは，confidence(“時間” $\Longrightarrow$ “料理”) $= 0.64$ という連想ルールが成り立っている．このルールの有意性を確認するために，「ドキュメント集合 $\mathcal{D}$ におけるキーワード “料理” の出現とキーワード “時間” の出現は独立である．）」という帰無仮説をたて，検定を行う．

表 1の分割表で式 1に従って計算を行うと $\chi_0^2 = 0.84$ となる．ここで，自由度は $(2-1)(2-1) = 1$ であり，危険率を 0.05 としたときには，$\chi_{0.05}^2 = 3.84 > 0.84$ であるので，帰無仮説は棄却されない．よって，この例においては，ドキュメント内のキーワード “時間” と “料理” の出現頻度は関係がある（独立でない）とは言えない．このように，連想ルールに対して統計的検定を行うことで，ナビゲーションに用いることができないルールを取り除くことが可能になる．

表 1　統計的に意味のないキーワード間の共起
Table 1　Keyword coocurance with no significance

	“料理” を含む	“料理” を含まない
“時間” を含む	488	273
“時間” を含まない	178	75

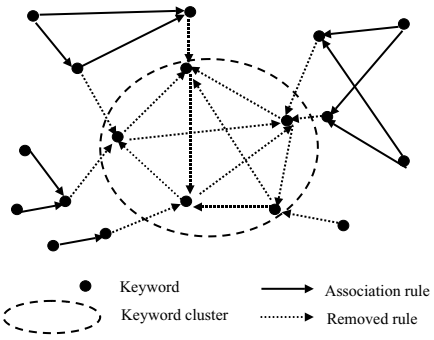

図 3　キーワードクラスタリングを用いたルールの精錬
Fig. 3　Refinement of rules by keyword clustering

4.2　クラスタリングを用いた不要なキーワードの除去

WEB ページのヘッダやフッタのようなフォーマットに含まれる語は，検索の際には有効でないため，これらを前処理として除去することを考える．これらの語はお互いに強い共起関係にあると考えられる．キーワード間の 1 対 1 の連想ルールのグラフを用いてキーワードのクラスタリングをし，そのようなキーワードのセットを抽出することができる．

ここで有向グラフ $G = (V, E)$ の強連結成分 (Strongly Connected Component, SCC) とは，G の部分グラフ $G' = (V', E')$ であって，$u, v \in V'$ である任意の節点間に E' 上での有向路が存在するもののうち，極大であるものをいう [10]．

最小クラスタサイズを c，クラスタ内最小確信度を $\alpha (\simeq 1)$ とした際の，キーワードのクラスタリングを用いた連想ルールの精錬手順は以下のとおりである．

（1） ルールの集合の中から，確信度が α より大きなルールを抽出し，グラフを構成する．

（2） 節点数が c 以上の全ての強連結成分をキーワードクラスタとして抽出する．

（3） キーワードクラスタ内のキーワードを不要キーワードとし，これらのキーワードを含む連想ルールを除去する (図 3)．

5. 実験評価

5.1 実験の設定

インターネットの利用者層が拡大するに従い, そこで流通する情報もローカルで日常生活に直結するものが増加していくと考えられる．デジタルシティとは，そのようなコミュニティネットワークのプラットフォームを都市をモデルにして構築する試みであり，世界各地で行われている [2]．日本においても，デジタルシティ京都プロトタイプ(注 1)を開発中である．デジタルシティ京都における GeoLink は WEB ページから抽出した住所情報をもとにその XY 座標を決定し，地図上でのブラウジングを可能にしたシステムである．現在までに京都市内の住所を持つ約 2600 件の WEB ページが掲載されている [11]．地図インタフェースの他に，キーワードと地理的関係や WEB のリンク関係を組み合わせて検索できる検索言語も開発されている [3]．我々は，このようなデジタルシティにおける情報検索を支援するエージェントを構築することを目指している [12]．

そこで，本稿で述べた各手法の評価対象として，京都市内の住所を持つ飲食店の約 1000 件の WEB ページを用いることにした．html ファイルから html タグを除去し，形態素解析プログラム『茶筌』[13] を用いて形態素解析と，単語への品詞別のタグ付けを行った．その中から，名詞（ただし代名詞や数詞など検索キーワードとして通常用いられないと思われるものは除く）をキーワードとして抽出した．だだ，『茶筌』には登録されておらず，うまく抽出ができないキーワードが存在する．そこで，約 120 語を辞書に登録した．

その結果として，

html ドキュメント数: 1014
抽出されたキーワード数: 10901
延べキーワード出現数: 76243
1 ページあたり平均キーワード数: 76 語

のデータセットが得られた．

このデータセットにおける，キーワードの出現頻度（出現するドキュメントの数）の分布を図 4に示す．

- ドキュメント集合中に 1 回しか出現しないキーワードが半数以上を占める．
- ドキュメント集合中に非常に高い割合で含まれるキーワードが存在する．

といった特徴が現れている．

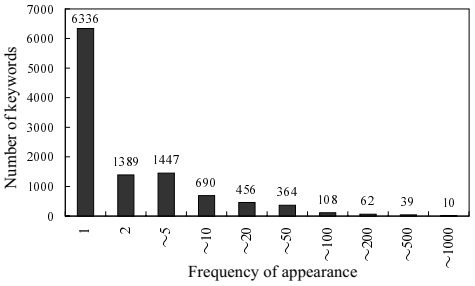

図 4 キーワードの出現頻度
Fig. 4 Distribution of keyword appearance

これらのキーワード間の 1 対 1 の連想ルールとして，確信度 50% 以上，支持度 0.5% 以上で 30397 個を生成した．例として，キーワード “イタリア” からの導かれるキーワードとして以下が抽出された．

“京都” “料理” “時間” “営業” “メニュー”
“店” “定休” “予約” “住所” “最寄り駅”
“パスタ” “予算” “問い合わせ” “情報” “電話”
“平均” “質問” “グルメ” “滋賀”
“Copyright” “取材”

5.2 ルールの精錬

キーワードのクラスタリングの方法で連想ルールの精錬を行った結果，ルールの数を 30%(9159 個) に削減することができた．上の “イタリア” の例に対して適用した結果，以下のルールが残された．

“イタリア” $\Longrightarrow$ “料理”
“イタリア” $\Longrightarrow$ “予約”
“イタリア” $\Longrightarrow$ “パスタ”

レストラン案内に関するページで定型的に使われる語へのルールが取り除かれていることが分かる．

さらに，上の結果に対して統計的手法によって有意水準 0.01 で連想ルールの精錬を行った結果，ルールの数は最初の 12%(3651 個) に削減された．

POS データに関して連想ルールの抽出を行った研究 [7] では，χ^2 検定を使ったところルールの数は 1%

(注 1)：http://www.digitalcity.gr.jp/

表 2　検索アスペクト
Table 2　Aspects of information retrieval

アスペクト	キーワード数	キーワードの例
1. メニュー	284	すき焼き サラダ ワイン カキ
2. スタイル	27	コース セット 定食 懐石 弁当
3. 利用状況	24	宴会 パーティ ディナー 接待
4. 雰囲気	38	気軽 カジュアル 伝統 風情
5. 設備	30	個室 座敷 川床 町屋
6. 国	22	ヨーロッパ イタリア 中国 日本
7. 利用客	12	グループ 女性 大人 カップル

表 3　アスペクト間での抽出ルール数
Table 3　Number of extracted rules between aspects

	1	2	3	4	5	6	7
1. メニュー	158	61	17	2	10	7	0
2. スタイル	16	16	0	1	2	0	0
3. 利用状況	3	9	20	3	5	0	4
4. 雰囲気	1	2	0	1	0	2	0
5. 設備	9	1	5	1	1	0	0
6. 国	6	1	0	0	0	3	0
7. 利用客	0	0	0	0	0	0	2

程度しか削減できなかったと報告されている．情報検索においては非常に高い頻度で現れる語が存在し，それらのキーワードに関する連想ルールが多く抽出されるため，この方法の適用が特に有効であると考えられる．

上の“イタリア”の例に対して統計的手法を用いてルールの精錬を行った結果，以下のルールが残された．

“イタリア” $\Longrightarrow$ “料理”

“イタリア” $\Longrightarrow$ “パスタ”

意味的に関連のあるキーワード間のルールだけが抽出されていることがわかる．

5.3　情報ナビゲーションにおけるルールの評価

将来的には，検索アスペクトに対応するキーワードを WEB ページから自動的に抽出することを目指すが，今回は手法の評価のため，京都市内のレストランの WEB ページの中で出現頻度の高いキーワードを人手で表 2のようなアスペクトに割り当てた．

各アスペクトごとの連想ルールの抽出数を表 3に示す．各行が連想ルールの条件部，各列が連想ルールの結論部を表している．これから，連想ルールが多く抽出されているアスペクトと，[雰囲気] や [利用客] のように，連想ルールがほとんど抽出されていないアスペクトが存在していることが分かる．

例えば，[設備] アスペクトから [利用状況] アスペクトへの連想ルールとして，以下が抽出されている．

“大広間” $\Longrightarrow$ “宴会”

“床” $\Longrightarrow$ “納涼”

“旅館” $\Longrightarrow$ “宴会”

“川床” $\Longrightarrow$ “食事”

“お座敷” $\Longrightarrow$ “宴会”

これらのルールのナビゲーションにおける有効性を評価するためには，ユーザどのような頻度で検索キーワードを入力するかを知る必要があり，今後ユーザを使った実証実験を行う必要があるが，上の例からは，以下のような考察を得ることができる．

- 例えば“お座敷” $\Longrightarrow$ “宴会” のルールのように，人間が見て意味があると判断できるルールをいくつか抽出することに成功している．
- アスペクトに割り当てられたキーワード数に対して，ルールに現れるキーワードの数が少ない．上の例では結論部の [利用状況] アスペクトには 24 個のキーワードが割り当てられているが，ルールに現れているのは 3 個である．

抽出されているルールが少なくなる理由としては，データ数の不足が考えられる．外部シソーラスを利用し，同義語や類義語などをクラスタリングし同一視して連想ルールの抽出を行う [7] ことにより，この問題を解決することが可能であると考えている．

以下では，抽出された連想ルールのうち，特徴的なものを例示し，情報ナビゲーションに用いる際の得失について考察する．

1.　“スパゲティ” $\Longrightarrow$ “パスタ” ([メニュー] $\Longrightarrow$ [メニュー])

この例は，上位概念と下位概念の関係を表現している．この連想ルールを用いることで，上位概念のキーワードでアスペクトを指定した場合に，下位概念のキーワードを含むドキュメントにも適合させることが可能となる．すなわち，シソーラスを用いたクエリの拡張と同様の働きをすることができる．

2. “イタリアン” $\Longrightarrow$ “イタリア” ([国] $\Longrightarrow$ [国])

この例はキーワードの関係が同義語にあたる場合である．この例でも，シソーラスを用いた場合と同様の効果を示すことができる．

表 4　アスペクト間での有効なルールの割合
Table 4　Rate of effective rules between aspects

	1	2	3	4	5	6	7
1. メニュー	0.33	0.28	0.59	0.00	0.30	1.00	-
2. スタイル	0.00	0.88	-	0.00	0.00	-	-
3. 利用状況	0.00	0.33	0.70	0.33	0.80	-	0.25
4. 雰囲気	0.00	0.50	-	1.00	-	0.50	-
5. 設備	0.00	1.00	0.80	0.00	0.00	-	-
6. 国	1.00	1.00	-	-	-	0.33	-
7. 利用客	-	-	-	-	-	-	1.00

3. “すきやき”$\Longrightarrow$“しゃぶしゃぶ”([メニュー]$\Longrightarrow$[メニュー])

同義語ではないが，関連の深いキーワード間の連想ルールの例である．このルールを使うことにより，アスペクト制約を解消できる可能性があるが，上の 2 つの例と違い，ナビゲーションの失敗を導く可能性も存在する．

4. “コロッケ”$\Longrightarrow$“サラダ” ([メニュー]$\Longrightarrow$[メニュー])

“サラダ” は非常に多くの WEB ページに現れるキーワードである．そのため，一見関連のなさそうな “コロッケ” というキーワードからの連想ルールが生じている．そもそも，結論部のキーワードの出現頻度が大きいため，このようなルールを適用する機会は稀であると考えられる．

5. “座敷”$\Longrightarrow$“宴会” ([設備]$\Longrightarrow$[利用状況])

異なるアスペクトに属するキーワード間の連想ルールの例である．このような関係はシソーラスなどから直接得ることができず，連想ルールの採用によって始めて明らかになる．このルールを用いて，アスペクト制約の解消や追加のアスペクトの提案に用いることが可能となる．

6. “川床”$\Longrightarrow$“食事” ([設備]$\Longrightarrow$[利用状況])

この例も上と同様であるが，連想ルールの適用によって，特殊化されすぎる例である．このようなルールは，制約の追加に用いると失敗に陥る可能性が高い．そのため，制約を追加する場合にはユーザとの対話を通して確認を行いながらルールの適用を行う必要がある．

上の例の 1,2,3 および 5 に当てはまるものを有効なルールとし，それぞれのアスペクト間でどれだけの割合の有効なルールが抽出されているかを示したものが表 4である（もともとルールが抽出されていないアスペクト間の値は計算していない．）．

この結果を見ると，各アスペクト間で有効なルールの占める割合が大きく異なっている．エージェントは情報ナビゲーションを行う際に，どのアスペクト間のルールを優先的に適用するかを考慮することで，ナビゲーションの効率を高めることができると考えられる．

6. 関 連 研 究

従来の情報検索においては，検索性能の向上のため，クエリ拡張と呼ばれるさまざまな手法が提案されてきた [14]．これらは，ドキュメントに含まれるキーワードとユーザの指定したキーワード間のミスマッチを解消するために，初期クエリに関連キーワードを追加して検索を行う方式である．キーワードの追加方法としては，外部シソーラスを利用したり，ドキュメント集合からシソーラス自動的に作成する方法などが提案されている [15] [16].

キーワードの追加を連想ルールを用いて行ったシステムとしては， Mondou [6] がある．これは，連想ルールを用いてユーザに関連語を提示し，ユーザがその中からキーワードを選択をして検索の絞り込みを行う検索エンジンである．

これらの先行研究では，ユーザに提示される多数の検索結果の集合の質をいかに向上させていくかに焦点が当てられており，デスクトップ環境での直接操作インタフェースを前提としている．それに対し，我々の手法は具体的な解を 1 つづつ提示し，順次よりよい解へ変更していくもので，モバイル環境に適した手法となっている．解の変更の方法は無数に存在し，その中には解の品質の低下を招くものもあるが，我々の手法はエージェントがユーザとの対話を通して検索アスペクトを考慮しながら連想ルールを適用していくことで，効果的なナビゲーションが可能となっている．

ユーザに解の候補を提示し，それに対するフィードバックを反映して解を洗練させていくインタフェースとしては， Automated Travel Assistant [17] が存在する．このシステムも，検索要求を検索対象の属性と値に関する制約と捉えている点が我々のシステムとの共通点である．ただし，検索対象は旅客機のフライト情報のデータベースであり，データの完全性は保証されている．それに対して，我々は WEB から抽出した連想ルールを精錬して用いることにより，不完全な WEB 情報での利用を可能としている．

7. む す び

本稿で我々の提案した情報ナビゲーションエージェ

ントは，ユーザとの対話に連想ルールを用いることにより，漸近的によりよい情報へとナビゲートしていくことができる．このエージェントは，モバイル環境における WEB 情報の活用に適した新しいインタフェースである．

エージェントがユーザとの対話の中で連想ルールを用いる手法として，

- アスペクトの制約の解決に用いる場合
- 追加のアスペクトの提案を行う場合

の二つを示した．

また， WEB ページから連想ルールを抽出する際に，意味的に関係のないキーワード間に連想ルールが抽出される問題を解決するため，

- 統計学の χ^2 検定を用いて，不要なルールの削減を行う方法
- キーワードのクラスタリングにより不要なキーワードを除去する方法

の二つの手法を提案した．今後は，ナビゲーションに用いることができるキーワード間の連想ルールをより多く抽出するための外部のシソーラスの利用と，対話機能を含めたエージェントシステム全体の開発を行う予定である．

文　　献

[1] G. Salton and C. Buckley, “Improving retrieval performance by relevance feedback,” Journal of the American Society for Information Science, vol. 41, no. 4, pp. 288–297, 1990.

[2] 石田亨, “デジタルシティの現状,” 情報処理, vol. 41, no. 2, pp. 163–168, 2000.

[3] 平松薫, 石田亨, “地域情報サービスのための拡張 web 空間,” 情処学論：データベース, vol. 41, no. SIG6(TOD7), pp. 81–90, 2000.

[4] I. J. Aalbersberg, “Incremental relevance feedback,” Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR ’92), Copenhagen Denmark, pp. 11–22, June 1992.

[5] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules,” Proceedings of the 20th VLDB Conference, Santiago, Chile, September 1994.

[6] 河野浩之, 長谷川利治, “WWWデータ資源検索におけるデータマイニング手法,” 情処学 DBS 研報, vol. 1996, no. 45, pp. 33–40, 1996.

[7] R. Srikant and R. Agrawal, “Mining generalized association rules,” Proceedings of the 21st VLDB Conference, Zurich, Swizerland, 1995.

[8] 薩摩順吉, “確率・統計,” 岩波書店, 1989.

[9] J. R. Quinlan, “Induction of decision trees,” Machine Learning, vol. 1, pp. 81–106, 1986.

[10] F. Harary, R. Z. Norman and D. Cartwright, “Structural Models: An Introduction to the Theory of Directed Graphs,” John Wiley & Sons, 1965.

[11] 平松薫, 小林堅治, B. Benjamin, 石田亨, 赤埴淳一, “デジタルシティにおける情報検索のための地図インタフェース,” 情処学論, vol. 41, no. 12, pp. 3314–3322, 2000.

[12] S. Oyama, K. Hiramatsu and T. Ishida, “Cooperative information agents for digital cities,” International Journal of Cooperative Information Systems, vol. 10, no. 1&2, 2001.

[13] 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 浅原正幸, “日本語形態素解析システム『茶筌』 version 2.0 使用説明書 第二版,” Technical Report NAIST-IS-TR99012, 奈良先端科学技術大学院大学, December 1999.

[14] R. Baeza-Yates and B. Ribeiro-Neto, “Modern Information Retrieval,” Addison-Wesley, 1999.

[15] J. Xu and W. B. Croft, “Query expansion using local and global document analysis,” Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR ’96), Zurich Switzerland, pp. 4–11, August 1996.

[16] L. M. de Campos, J. M. Fernández and J. F. Huete, “Query expansion in information retrieval system using a bayesian network-based thesaurus,” Proceedings of 14th Conference on Uncertainty in Artificial Intelligence, Wisconsin, USA, pp. 53–60, 1998.

[17] G. Linden, S. Hanks and N. Lesh, “Interactive assessment of user preference models: The automated travel assistant,” Proc. 6th International Conference on User Modeling (UM97), Sardinia, Italy, pp. 67–78, June 1997.

（平成年月日受付）

小山　　聡　（学生員）

平成 6 年京都大学工学部数理工学科卒業，8 年京都大学大学院工学研究科数理工学専攻修士課程修了．平成 8-10 年日本電信電話株式会社勤務．現在，京都大学大学院情報学研究科社会情報学専攻博士後期課程在学中．情報検索，人工知能に興味を持つ．

石田　　亨　（正員）

昭 51 年京都大学工学部情報工学科卒業，53 年京都大学大学院修士課程修了．同年日本電信電話公社電気通信研究所入所．現在，京都大学大学院情報学研究科社会情報学専攻教授．工学博士．人工知能，社会情報学に興味を持つ．