論　文

有効な組合せの発見による創造活動支援

西原　陽子[†a)]　　砂山　渡[††]　　谷内田正彦[†]

Creative Activity Support by Discovering Effective Combinations

Yoko NISHIHARA[†a)], Wataru SUNAYAMA[††], and Masahiko YACHIDA[†]

あらまし　現在世界中の至るところで組合せによる創造活動が行われている．創造活動が成功するかどうかは用いる組合せによって変わってくるが，世の中には用いる組合せの候補がたくさんあるので，どれを用いればよいのかは一見しただけでは分からない．また，あらゆる組合せに対して成功するか否かを検証することは難しい．そこで，効率良く創造活動が成功する有効な組合せを発見する手法が必要となる．本論文では，創造活動を支援するための有効な組合せを発見する手法を提案する．提案手法では組合せの斬新さと大衆に受け入れられる可能性をキーワードの Web ページヒット数などから測り，それによって組合せの有効度を定める．

キーワード　創造活動，組合せ，流行予測

1. まえがき

現在，組合せに基づく創造活動は，世界中の至るところで行われている．例えば「テレビ電話」が「テレビ」と「電話」の組合せとして近似的に表現できるように，世の中の多くのものは既存のものの組合せで近似的に表現が可能であり，今後世の中で作られる多くのものにもこれがあてはまると考えられる．

組合せに基づく創造活動はだれにでもできることであるが，用いる組合せに関する様々な要因によってその成功と失敗が変わってしまうため，創造活動を行う際には組合せの有効性に関する十分な検証が必要となる．しかし，世の中には数多くの組合せがあるので，一見しただけではどの組合せが有効なのかを知ることは難しく，時間や身体の制約によってすべての組合せを検証することはできない．

また，組合せが創造活動の目的に対して有効でないと創造活動は成功しない．一般に組合せが創造活動にとってありふれたものであれば，それが多くの人に受け入れられる可能性は高い．反対に組合せが斬新であれば，多くの人に受け入れられて広まる可能性は低くなるが，独創的な分その創造活動に対して組合せが既に存在する可能性も低くなる．よって，その組合せには目的達成のための新たなアプローチとしての高い価値を付与できる．すなわち斬新ではあるが，より多くの人に受け入れられる可能性の高い組合せを発見することは非常に有益であるといえる．

そこで，本論文では新しいものを既存のものの組合せとしてとらえ，独創的かつ世の中に広まる可能性の高い組合せを人間に提示し，人間が組合せを検証する効率の改善を図ることで創造活動を支援するシステムを提案する．すなわち本研究は，発想支援や創造性支援として人間の新たなアイデア発生を積極的に支援することを目的とするのではなく，アイデア発生のための有効な材料として既存のものの有効な組合せの出力を目指している．

以下，**2.** で関連研究と本研究の位置付けについて，**3.** で組合せを評価するために行った予備実験について述べる．**4.** で提案する組合せ評価システムについて述べ，**5.** ではシステムの出力と実際に世間に広まり出した組合せとの比較及びシステムと従来手法の比較を行い，**6.** で本論文を締めくくる．

2. 関連研究と本研究の位置付け

人の創造活動の支援を行う研究は，対象とする創造

† 大阪大学大学院基礎工学研究科，豊中市
Graduate School of Engineering Science, Osaka University, 1–3 Machikaneyama-cho, Toyonaka-shi, 560–8531 Japan
†† 広島市立大学情報科学部，広島市
Faculty of Information Sciences, Department of Information Machines and Interfaces, 3–4–1 Ozuka-Higashi, Asa-Minami-ku, Hiroshima-shi, 731–3194 Japan
a) E-mail: yoko@yachi-lab.sys.es.osaka-u.ac.jp

活動や支援方法の違いに応じて多岐にわたって行われている．知識創造の支援に関しては，生成する対話を用いて支援を行う研究 [1] がある．生成される対話はユーザと分身エージェントとの間で交わされる対話である．この分身エージェントとはある人物についての情報を，その本人に成り代わって表現する機能をもち，それを計算機上で実現するエージェントを指す．分身エージェントが生成する対話の応答は深い意味処理などは全く行わない．しかし，対話を行うユーザは違和感を感じることなく，エージェントの発言を意味あるものとして解釈する．そうすることで知識を構築していく．対話創造の支援においては人と計算機の間で自然言語での会話を可能とするプログラム ELIZA がある [2]．ELIZA は入力された文に含まれる単語をもとに，データベースにある返答用のテンプレートを参照して文を作成し返答するが，返答テンプレートの追加機能や文の意味の区切りを学習する機能により，創造的な対話を目指している．

また，創造活動の支援に事物の組合せを用いる研究が従来から行われてきている．例えば，高杉ら [3] はアイデアは雑多な情報をユーザが思考する中で組み合わせることによって生まれるものであると考え，ユーザが入力したキーワードと関連の深いキーワードやテキストを 2 次元平面上に配置したものを示すことで思考支援を行っている．本研究においても，新たなものの創造は既存のものの組合せを基本として，その組合せから連想される新たなものを創造する活動を支援する．

本研究では新しい組合せを作る創造活動の対象として（1）研究活動及び研究活動における論文タイトル作成（2）コンビニエンスストアにおける新商品開発及び販売を扱う．(1）の方では，これから研究活動が活発となっていくと予想されるものに含まれる組合せを予測し，(2）の方ではコンビニエンスストアでの流行商品に含まれる組合せを予測することを目的とする．

人が流行予測を行う際には，さしあたって世の中ではやり始めているものや長期間流行っているものをきっかけにして予測を進めていくことが多い．

注目キーワードを含む重要文を出力することで，未来の流行予測の支援を行う研究 [4] がある．Web 上のデータからキーワードに関する情報の新規性やユーザの観点との関連度を測り，世の中でのニーズが増してきているキーワードを抽出する．出力としては抽出したキーワードとともに Web 上のテキストデータ中における，抽出キーワードを含んだ重要文を出力する．単純にキーワードを示すだけではそこから流行予測をするのは難しいが，キーワードの使用例として重要文を提示することで，ユーザはそのキーワードから流行予測がしやすくなる．しかし，あくまで示されるキーワードは流行予測のきっかけであって，ユーザはそこから更に予測を進める必要がある．本研究では流行する可能性が高いそのものを予測する点が異なる．

対して，流行するキーワード，コンセプトそのものを予測する研究 [5] がある．この研究で用いている手法は，流行の背景には何らかの社会的要因があるという考えに基づいている．社会的要因を「言葉」で表現し，今後流行する可能性のある「言葉」との意味的距離を測る．距離が近いものほど流行する可能性が高いとする．新しく作られたものの予測という点は本研究の類似点である．距離は各「言葉」に付けられた説明文に対して自然言語処理のベクトル空間法を適用して求めるが，説明文が短いために距離に差がつかず，実際に予測を行っているものは将来流行するキーワードが関連する分野である．

また流行予測に関して，文書から主張を得るための手法で地震発生予測を行うものもある [6]．これは過去の地震履歴を文書に見立て，KeyGraph[7] というシステムを用いることでストレスを受けている断層を発見し，地震発生の予測を行うものである．もともと KeyGraph は文書の意味構造を計算によってグラフ表現にし，文書中の主張の込められたキーワードを抽出するための手法として提案された．これに対し，本研究では二つのキーワードの認知度と同時存在確率からのずれの変化を追うことで，その二つのキーワードによって新しく作られるもの自体の流行予測を行うことを目的とする．

本研究においては創造活動に組合せを用いるが，組合せにより近似的に表現される新しいものというのはあらゆる所で見ることができる．先の例でも挙げたテレビ電話などのように身の回りのものの多くは，複数のものの組合せで表現可能といっても過言ではない．企業の経営戦略の一つであるブランド戦略 [8] でも組合せを用いた新商品開発を行っている．ブランド戦略において重要なのは顧客に与える印象に差をつけることであるといわれている．イメージの差別化を図るために企業は自社ブランドの製品と他製品を組み合わせることによって新商品開発を行う．このとき新しく作られるものは自社ブランド製品のライン拡張と他製品

のカテゴリー拡張によって作られる．ライン拡張とは拡張に用いるものの本質はそのままで，そこへ新たな機能を追加して新商品を作ること，また，カテゴリー拡張とは拡張に用いるものに備わっている機能を，別のものに付加して新商品を作ることである．組み合わされてできたものの本質はライン拡張側のものを受け継いでおり，付加価値としてカテゴリー拡張側の機能を備えているとする．したがって，あらゆるものはライン拡張とカテゴリー拡張の両方が可能となっており，拡張の仕方の違いで異なる商品ができることになる．以下にブランド戦略の例を示す．

- ディズニーシー：「ディズニーランド」のライン拡張 ＋「海」のカテゴリー拡張
- プリント倶楽部：「ゲーム」のライン拡張 ＋「写真」のカテゴリー拡張
- カメラ付携帯電話：「携帯電話」のライン拡張 ＋「カメラ」のカテゴリー拡張

ここで，本研究における組合せを定義する．本研究では組合せをキーワードＬのライン拡張とキーワードＣのカテゴリー拡張からなるものとする．

3. 組合せ評価用予備実験

本章では組合せを評価するために行った予備実験について述べる．前章で企業が自社ブランドを確立していくためのブランド戦略でとられる手法の一つとして，自社ブランドの製品のライン拡張とそれ以外の製品のカテゴリー拡張によって新商品を開発する手法があることを示した．世の中で流行した商品の中には二つのものの近似的な組合せで構成されているものが多数存在する．はやった商品ははやるための何らかの要素を備えていたことは間違いない．そこでまず，ある物事に対する注目度を「その物事に対して興味をもっている人の数」として定義すると，注目度はその物事を表すキーワードのヒット数（キーワードを含むデータ数）で近似的に測ることができる．そこでライン拡張側キーワード，カテゴリー拡張側キーワードの各ヒット数に反映された注目度の変化よりその組合せのはやった要因を探ることにした．

実験ではまず，レンタル日記サイト「さるさる日記」[9] 上で書かれた日記から 2,627 個の名詞キーワードを用意して各キーワードのヒット数と異なる二つのキーワードの同時ヒット数を 2000 年 3 月から 2003 年 5 月まで各月ごとに測った．更に世の中ではやった組合せからなる商品について，ライン拡張側キーワードとカテゴリー拡張側キーワードに分けて，その商品が発売された月の前後 3 か月の両キーワードのヒット数及び同時ヒット数(注1)に見られる傾向を探った．

具体例として 8 商品のヒット数の傾向を図 1 と図 2 に示す．図 1，図 2 は縦軸を各商品の発売月後 3 か月間の同時ヒットの上昇率の平均を，全組合せの上昇率の平均に標準偏差を足したもので割った値(注2)とし，図 1 の横軸は各商品の発売月前後 3 か月間のヒット数の平均を全キーワードのヒット数の平均に標準偏差を足したもので割った値であり，図 2 の横軸は各商品の発売月前 3 か月間のヒット数の上昇率の平均を全キーワードのヒット数の上昇率の平均に標準偏差を足したもので割った値としている．

図 1 と図 2 よりはやった商品においては，ライン拡張側キーワードのヒット数，カテゴリー拡張側キーワードのヒット数の上昇率及び同時ヒット数の上昇率のいずれも高い値を示していることが分かる．ここで，あるキーワードのヒット数が平均に標準偏差を加えたもの以上になっているということは，そのキーワードに対する世の中での認知度がある程度高いことを示す．またキーワードのヒット数の比が平均に標準偏差を加えたもの以上になっているということは，そのキーワードが新規的なものであるということを示す．したがって，世の中で流行する組合せからなる商品にはライン拡張側キーワードで表されるものがある程度認知されていること，カテゴリー拡張側キーワードで表されるものに新規性があることが分かった．

4. 組合せ評価システム

本章では提案する組合せ評価システムについて述べる．図 3 にシステム構成図を示す．

3. で述べた予備実験より流行するものを近似的に表現する組合せにはライン拡張側のものがある程度認知されていること，カテゴリー拡張側のものには新規性があることが明らかとなった．そこでユーザは認知されているライン拡張側のものを表すキーワードをシステムに入力し，システムは新規なものを表すカテゴ

(注1)：ライン拡張側キーワードとカテゴリー拡張側キーワードを同時に含むデータ数．

(注2)：平均に標準偏差を足したものが 1 の値をとるように数値の正規化を行っている．実験的にライン拡張側キーワードのヒット数，カテゴリー拡張側キーワードのヒット数，及び両キーワードの同時ヒット数の上昇率の三つは正規分布に従うことを確認している．そこで，(平均＋標準偏差）以上の値をもつ分布の上位のデータ（正規分布においては上位 16%）を 1 以上の値をもつものとして認識しやすくするために正規化を行った．

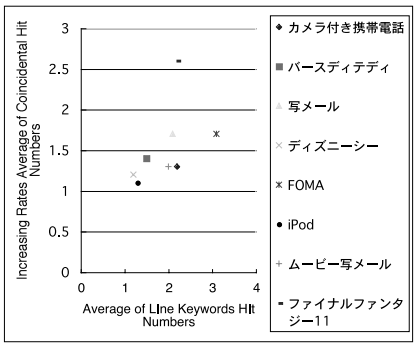

図 1　ライン拡張側キーワードのヒット数と同時ヒット数の上昇率
Fig. 1　Hit numbers of line keywords and rates of coincidental hit numbers.

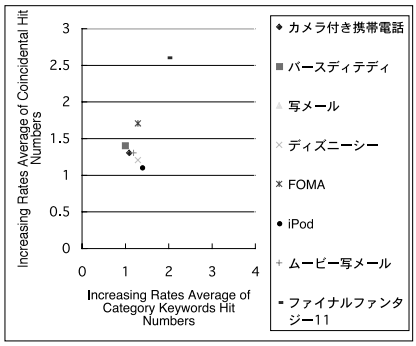

図 2　カテゴリー拡張側キーワードのヒット数の上昇率と同時ヒット数の上昇率
Fig. 2　Hit numbers of category keywords and rates of coincidental hit numbers.

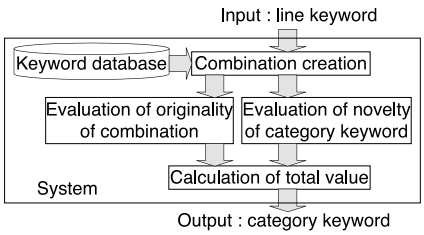

図 3　組合せ評価システム
Fig. 3　Evaluation system of combinations.

リー拡張側のものをキーワードとして出力する．

システムは入力としてライン拡張側のものを表すキーワード L を受け取る．次に，データベース中のカテゴリー拡張側のものを表すキーワード C との組合せを作る．続いて作られた組合せに対して,「組合せの独創性」,「組合せの流行可能性」を示す評価値を算出する．最後に 2 種類の評価値より組合せの評価値を定め，有効な組合せを作るカテゴリ拡張側キーワード C を評価値の高い順に出力する．以下，本システムが用意した 2 種類のキーワードデータベース及びシステムの各モジュールについて述べる．

4.1　キーワードデータベース

本システムがもつ 2 種類のデータベースについて説明する．各データベースには組合せに用いるキーワードと各キーワードのある時点のヒット数と，異なる二つのキーワードの同時ヒット数を示す値が数年分収められている．データベースの一つは論文タイトル予測用，もう一つはコンビニエンスストアの新商品の予測用である．以下，各データベースのキーワード集合を用意した手順及び各データベースの詳細について示す．

4.1.1　論文タイトルデータベース

世の中には多くの研究分野があり，各分野ごとに学会も多数存在する．そのため，研究分野ごとに研究内容を示すキーワードも異なってくる．例えば，人工知能学会においては「データマイニング」を扱った研究を目にすることは多いが,「素粒子」を扱った研究を見かけることは少ない．また研究活動で使用されるキーワードは，日常会話で用いられる言葉とは異なるいわゆる用語であることが多い．したがって，ありとあらゆるキーワードを用意するのではなく，ある程度専門的なキーワードに限定することで出力のノイズも抑えることができ，システムユーザにも分かりやすい結果を出力することができる．

そこで今回は研究分野を人工知能分野に絞り，人工知能学会の全国大会での発表タイトルからキーワード集合を用意した．扱った発表タイトルは第 14 回から第 17 回までに発表された計 930 タイトルである．またタイトルに含まれる 964 個の名詞及びフレーズをキーワードとして用意した．以下にキーワード集合の名詞及びフレーズを用意する手順を示す．

［step1］ 第 14 回から第 17 回までの発表タイトルのテキストファイルを用意する

［step2］ 各テキストファイルを茶筌 [10] にかけ，片仮名表記のキーワードと名詞を取り出し，初期キーワード集合とする

［step3］ 初期キーワード集合中の各キーワードに対して，キーワードを含むタイトル数を数え出現回数がしきい値 T_1 以上であれば，そのキーワードの直前のキーワードをリストアップし，リストアップされたキーワードの出現回数がしきい値 T_2 以上であれば，それら二つのキーワードを結合させたフレーズを作る

表 1　論文タイトルデータベースのキーワード例
Table 1　Examples of keywords in the database of papers' titles.

相関ルール	マイニング	要約
対話	分散協調	ゲーム
管理	音楽	ネット情報
コミュニティ	テキスト	データベース
移動ロボット	インターネット	モデル

表 2　日記データベースのキーワード例
Table 2　Examples of keywords in the database of diaries.

プリン	メロン	ヨーグルト
ワイン	どら焼き	肉まん
サラダ	抹茶	クリーム
ハンバーガー	レーズン	ワッフル
烏龍茶	緑茶	コーラ

［step4］　フレーズと初期キーワード集合をまとめてデータベースのキーワード集合とする

今回，しきい値はそれぞれ $T_1 = 5$，$T_2 = 2$ とした．フレーズをキーワード集合に加えた理由は，論文タイトルには論文のトピックを可能な限り込める必要があるため，キーワードの一つひとつが長くなる傾向がある．しかし，タイトルを茶筌にかけて得られる名詞キーワードはそれよりも短い．隣接する名詞キーワードをつなげることでもとのタイトル中で使用されているキーワードになることからフレーズをキーワード集合に加えた．表 1 は得られたキーワード集合の一部である．

続いてキーワードセット中のキーワードがタイトル中に見られるタイトル数を各回ごとに数え，その時点のキーワードのヒット数としてデータベースに収めた．同様に異なる二つのキーワードが同時にタイトル中に見られるタイトル数を二つのキーワードの同時ヒット数として収めた．

4.1.2　日記データベース

コンビニエンスストアにおいて，次々に新商品が発売されるのはおにぎり，弁当などの食料品分野が多い．新商品が発売されやすいということは流行するものが出る可能性も高いといえる．そこで予測する新商品もおにぎり，弁当，お茶などの食べ物に限定した．キーワードのヒット数は日記レンタルサイト「さるさる日記」[9] 上で書かれている個人の日記で測ることにした．日記には大衆の日々の意見が反映されており，即時性もあることからキーワードの認知度を正確に測ることができると考えたために日記を用いた．扱った日記は 2000 年 3 月から 2003 年 5 月までの 5,906 人の 286,033 件の日記である．キーワードは日記中に含まれる 2,096 個の食べ物を表す名詞を用意した．以下にキーワード集合を用意する手順を示す．

［step1］　「さるさる日記」において 2000 年 3 月から最低 1 か月以上日記を書いている人の日記の HTML ソースを 2000 年 3 月から 2003 年 5 月までダウンロードする

［step2］　各年，各月ごとに日記を分け，HTML タグ等を外したテキストの日記部分だけを切り出す

［step3］　切り出した部分を茶筌にかけ得られる名詞の中から，人手により食べ物を表すキーワードを選んで抜き出し，キーワード集合とする

表 2 は得られたキーワード集合の一部である．続いて 2000 年 3 月から 3 か月ごとに日記を分類した上で，キーワード集合中の各キーワードが現れる日記数を数え，その時点のキーワードのヒット数としてデータベースに収めた．同様に異なる二つのキーワードが同時に日記中に見られる日記数も同時ヒット数として収めた．

4.2　独創性評価部

本章では組合せの独創性の評価方法について述べ，その評価値を与える式を示す．

組合せを作る二つのキーワード L，C があるときに，この二つのキーワードの存在確率を $P(L)$，$P(C)$ で表す．ここで全データ数を all，$keyword$ を含むデータ数は $hit(keyword)$ とし，$keyword$ の存在確率を以下の式で定義する．

$$P(keyword) = \frac{hit(keyword)}{all} \tag{1}$$

二つのキーワード L と C が独立に存在するならば，2 キーワードが同時に存在する確率は $P(L)P(C)$ となる．しかし実際に L と C が同時に存在する確率 $P(L \cap C)$ と $P(L)P(C)$ の間にはずれが生じる．ここで，$P(L \cap C)$ が $P(L)P(C)$ より小さいほど組合せが独創的であるといえる．そこで式 (2) で組合せの独創性の評価式を定義する(注3)．

$$ori(L, C) = \frac{P(L)P(C)}{P(L \cap C)} \tag{2}$$

また，式 (2) による独創性の値が時間経過に伴って小

(注3)：相互情報量に相当する値である．

さくなる組合せを有効とする．なぜなら，独創的でなくなるということは組合せの普及の兆しを見せていると理解されるからである．これを次の式 (3) で評価する．

$$t\_ori(L,C) = \frac{ori(L,C) - ori_{init}(L,C)}{year} \qquad (3)$$

ori_{init} は評価値算出に用いるデータの中で最も古い時点における ori であり，$year$ はその最古のデータまでの年数である．

4.3 流行可能性評価部

あるキーワードに新規性が見られるならば，そのキーワードに対するヒット数の上昇率は新規性がないものに比べてより大きくなる．ここで式 (4) のカテゴリー拡張側キーワードのヒット数の時間変化によりその新規性を評価する．

$$t\_nov(C) = \frac{hit(C) - hit_{init}(C)}{year} \qquad (4)$$

また，同じ上昇率でも評価時点でのヒット数が低いものの方が，これから認知度が増し流行するという意味においてより新規性が強い．したがって，カテゴリー拡張側キーワードの新規性を測る評価式を式 (5) で定義する．

$$nov(C) = hit(C) \qquad (5)$$

4.4 評価値算出部

評価値算出部では，独創性評価部と流行可能性評価部で与えた四つの評価値を統合して，組合せの総合評価値を決定する．

総合評価値は各評価値をその値による順位で正規化したものの総和で，式 (6) のように表す．ただし，$Rank()$ は各評価値の順位を返す関数である．

$$\begin{aligned} &evaluation(L,C) \\ &\quad = Rank(ori(L,C)) + Rank(t\_ori(L,C)) \\ &\qquad + Rank(nov(C)) + Rank(t\_nov(C)) \end{aligned} \qquad (6)$$

システムは $evaluation$ の値が小さい順にカテゴリー拡張側キーワードを出力する．また，ヒット数による足きりを行った上で出力を限定している．足きりを行う条件は次の 3 点であり，以下の条件をどれか一つでも満たす組合せは出力に含めない．

（1） 組合せ評価時点のライン拡張側キーワードとカテゴリー拡張側キーワードの同時ヒット数が 0

表 3 論文タイトル予測用ライン拡張側キーワードと各時点におけるヒット数
Table 3 Line keywords for prediction of papers' titles and their hit numbers.

Line Keyword	14th	15th	16th
情報	17	21	23
知識	21	16	21
発見	11	8	9
ロボット	6	8	8
Average of Hit Numbers	1.2	1.8	1.9

（2） カテゴリー拡張側キーワードの組合せ評価時点でのヒット数が 0

（3） 式 (2)〜(5) の四つの評価値のいずれか一つが平均値以下

(1)，(2) は現在カテゴリー拡張側のものとの組合せが全く存在しないという意味なので，創造活動を行うにはリスクが大きく，流行の兆しもない組合せであり出力としては不適切なためである．(3) は真に有効な組合せを出すために，すべて高評価値がついている組合せのみ取り出すことを目的とした．

5. 実　　験

本章では用意した論文と日記に関する，二つのデータベースに対して行った組合せ評価実験について述べる．本システムの目的は，組合せによる創造を行うユーザが網羅的に組合せを検証することを避けるために，創造活動に有効な材料として今後世の中に現れる可能性の高い組合せを提供することである．以下，各データベースごとの評価実験の詳細と考察を示す．

5.1 論文タイトル予測実験

第 14 回から第 16 回までの人工知能学会全国大会の発表タイトルのデータから，第 17 回のタイトルに現れる注目度が高まってきているキーワードの組合せを **4.1.1** で述べたデータベースを用いて予測した．

ここで，第 15 回から第 17 回まで連続して同時ヒット数が上昇しているならば，組合せに対する注目度が高まってきているとし，それらを正解としてシステムの出力と比較する．注目度が高まっていることは流行るための必要条件である．

表 3 はシステムに入力したライン拡張側キーワードと各時点におけるヒット数（キーワードを含むタイトル数）を表す値を示したものである．表 4 に入力したライン拡張側キーワードとシステムが出力したキーワード例及び全 964 通りの組合せの中からシステムが

出力した組合せの数を示す．

表 5 に本実験の正解となる，第 15 回から第 17 回にかけて注目度が高まってきた組合せと各評価値の順位を示す．表 5 中のカテゴリー拡張側キーワードをシステムはすべて予測し，その出力に含んでいる．実際，第 17 回の発表タイトル中には表 5 に示す組合せが見られた．例として「XML データストリームからの高速な知識発見手法 [11]」,「知能ロボットの自律移動のための実画像からの物体認識 [12]」,「Web からの人間関係ネットワークの抽出と情報支援 [13]」がある．

表 5 中には流行する可能性を評価した新規性の評価値だけでは拾い上げることができない組合せがある．例えば「知識と発見」,「ロボットと知能」などでそれらは世の中での斬新な組合せを表す独創性を評価に組み入れることで拾い上げることができた．本システムでは各値が高くなくても総合的な値が高い組合せ，すなわちカテゴリー拡張側キーワードの新規性と組合せの独創性のバランスがとれた組合せを有効としている．結果，各値が低いにもかかわらず世の中に出てきた「知識と発見」,「ロボットと知能」の組合せを予測することができた．この点から，新規性と独創性の評価値を組み合わせた本システムの組合せの予測は成功しているといえる．

だが，システムは表 5 に含まれないキーワードも多数出力した．表 6 に「発見」を入力したときの，システム出力を示す．表 6 中の「データ」,「知識」以外は同時ヒット数が一定若しくは減少しているため，正解とはされなかったキーワードである．

例えば「ルール」は同時ヒット数が減少している．しかしその独創性は高く，近年データマイニング分野において「相関ルールの発見」というテーマは脚光を浴びている．そのため，このテーマに着目した創造活動も十分に考えられる．しかし，システムの出力には大きくはやった後，少し減衰傾向にある組合せも含まれると考えられ，今後研究テーマが少なくなれば廃れる可能性がある．

一方「知識」「データ」「ルール」以外のキーワードは同時ヒット数が一定である．これらは今後の動向は不明であるが，流行するための必要条件を備えたキーワードとして出力された．例えば「チャンス発見」,「トピック発見」などは最近現れたテーマとして注目されつつあるが，どれだけ流行するかに関しては未知数である．これらの出力はどれも流行するための十分条件ではないため，出力をユーザが吟味する必要はあるが，

表 4　論文タイトル予測実験における本システムの入力と出力
Table 4　Inputted and outputted keywords of this system in an experiment of prediction papers' titles.

Inputted Keyword	Examples of Outputted Keyowrds	Outputted Numbers
情報	知識，マルチ，インタフェース	97 / 964
知識	クラス，音楽，コミュニティ	59 / 964
発見	パターン，モデル，チャンス	24 / 964
ロボット	時間，協調，インタラクティブ	23 / 964

表 5　論文タイトル予測実験において実際に注目度が高まった組合せとその評価値の順位
Table 5　Correct combinations and each evaluation value in an experiment of prediction papers' titles.

Line Keyword	Category Keyword	$ori(L, C)$	$t\_ori(L, C)$	nov	$t\_nov(L, C)$
情報	支援	20	605	1	1
情報	抽出	20	585	34	1
知識	発見	376	605	409	288
発見	データ	20	614	6	1
発見	知識	20	602	4	1
ロボット	移動	400	398	317	82
ロボット	知能	391	389	420	93
ロボット	動作	409	409	295	82

すべての組合せの中から有効なものを選び出すことに比べて，システムは組合せを大きく絞り込んでいるといえる．実際「発見」との組合せに関して今回の正解を対象とした場合，2/964 から 2/24 までに絞り込んでいる．

表 3 と表 4 からライン拡張側キーワードのヒット数が高いほど，システムはより多くの組合せを出力していることが分かる．このことからライン拡張側キーワードの認知度が高いと注目される組合せの数も多くなり，真に有効な組合せを探し出すことがより難しくなると考えられる．

5.2　新商品予測実験

2002 年 6 月から 2003 年 2 月までの日記のテキストデータから，2003 年 3 月から 5 月の間に新発売されるコンビニエンスストア商品に使われている食材の組合せを，**4.1.2** で述べたデータベースを用いて予測した．ここで 2002 年 6 月から 2003 年 5 月まで連続して同時ヒット数が上昇している組合せを正解として，システム出力と比較した．例えばある商品が 2003 年 3 月ごろに新発売されたとすると，それまではその商品中に含まれる組合せは存在しないため同時ヒット数の値は小さい．また商品として発売されると多くの人が注目するため，同時ヒット数の値は上昇する．し

表 6 「発見」を入力したときのシステム出力キーワードとその評価値
Table 6 Outputted category keywords in inputing “discovering” and each evaluation value.

Outputted Keyword	Total of Values	Originality eq.(2)+eq.(3)	Novelty eq.(4)+eq.(5)
ルール	38	3	35
パターン	117	2	115
履歴	258	65	193
電子	259	66	193
モデル	357	343	14
順序	367	252	115
マイニング	368	253	115
決定	373	338	35
会話	392	304	84
要約	392	304	84
知識	660	623	27
データ	641	634	7
構造	653	626	27
物語	1046	621	425
相関	1048	623	425
検出	1049	624	425
視線	1049	624	425
カスケード	1052	627	425
スティング	1052	627	425
チャンス	1052	627	425
トピック	1052	627	425
化合	1052	627	425
掲示板	1052	627	425
類比	1052	627	425

表 7 新商品予測用ライン拡張側キーワードと各時点におけるヒット数
Table 7 Line keywords for prediction of new items and their hit numbers.

Line Keyword	2002/6	2002/9	2002/12
おにぎり	29	32	28
ラーメン	112	113	99
お茶	113	143	127
Average of Hit Numbers	31	35	33

表 8 新商品予測実験における本システムの入力と出力
Table 8 Inputted and outputted keywords of this system in an experiment of prediction papers’ titles.

Inputted Keyword	Examples of Outputted Keywords	Outputted Numbers
おにぎり	うなぎ，つくね，にんじん 目玉焼き，ポテトチップス	24 / 2,096
ラーメン	ポテト，チキン，食パン，チャーハン，ねぎ，肉まん	52 / 2,096
お茶	アイスクリーム，ケーキ，トマト，豆，塩，ワッフル	54 / 2,096

表 9 「おにぎり」を入力したときのシステム出力
Table 9 Outputted category keywords in inputing “rice-ball” and each evalution value.

Outputted Keyword	Total of Values	Originality eq.(2)+eq.(3)	Novelty eq.(4)+eq.(5)
うなぎ	402	158	244
つくね	407	241	166
にんじん	421	202	219
目玉焼き	427	178	249
鯛	443	249	194
マヨネーズ	478	261	217
大根	494	268	226
ポテト	500	279	221
味噌汁	515	282	233
コーン	564	248	316
フレーク	768	217	351
サンドイッチ	584	271	313
牛乳	615	294	321
そば	634	256	378
パン	647	280	367
カレー	668	270	398
ラーメン	677	262	415
漬物	710	231	479
ジュース	712	299	413
ポテトチップス	713	141	572
アイスクリーム	716	294	422
豆	724	281	443
豆腐	734	282	452
おかか	1030	158	872

たがって，同時ヒット数が上昇してきた組合せははやり始めた組合せとみることができる．このような組合せとシステム出力を比較することでシステムの性能を測る．

表 7 にシステムに入力したライン拡張側キーワードと各時点におけるヒット数を，表 8 にライン拡張側キーワードごとにシステムが出力したキーワード例及び全 2,096 通りの組合せの中からシステムが出力した組合せの数を示す．

また，表 9 に「おにぎり」をライン拡張側キーワードとしたときのシステム出力（正解に下線）を示す．これらに対して本実験での正解となる「おにぎり」との同時ヒット数が実際に上昇した組合せを表 10 に示す．同時ヒット数が上昇した 14 個の組合せのうち，システムで予測できたのは「そば」,「つくね」,「アイスクリーム」,「カレー」,「サンドイッチ」,「鯛」,「豆」,「豆腐」,「味噌汁」の 9 個（表中の下線付きキーワード）であった．この 9 個の組合せが予測できた理由は，この 9 個の組合せの独創性と新規性どちらか一方だけに高い評価がついているのではなく．両方にある程度高い値がついているため，システムが 2 種類の評価値を同等に評価して有効な組合せを出力したためである．実際の商品例としては「おにぎりサンド」,「つくねおにぎり」,「鯛飯」などがある．またローソンの「おにぎり屋」シリーズに 2003 年 5 月から味噌汁と緑茶を

表 10 「おにぎり」との同時ヒット数が上昇した組合せ
Table 10 Correct combinations with “rice-ball.”

Category Keyword	Total of Values	Originality eq.(2)+eq.(3)	Novelty eq.(4)+eq.(5)
しょうが	3202	1367	1835
そば	634	256	378
つくね	407	241	166
アイスクリーム	716	294	422
カレー	668	270	398
サンドイッチ	584	271	313
ビール	2919	1367	1552
昆布	3102	1380	1722
醤油	3251	1529	1722
鯛	443	249	194
豆	724	281	443
豆腐	734	282	452
納豆	2610	1529	1081
味噌汁	515	282	233

表 11 新商品予測実験における正解となる組合せが出力に含まれるまでの平均試行回数
Table 11 Trial times average of finding correct combinations.

Line Keyword	Without System Outputs	With System Outputs
おにぎり	150	2
ラーメン	140	3
お茶	116	3

加えておにぎりを核とした商品販売が開始された．これは食材を直接組み合わせたものではなく，創造活動を成功させるには，ユーザ自身も組合せからアイデアを引き出す必要があることが分かる．

表 10 中において，同時ヒット数が上昇したがシステムによって予測できなかったキーワードは「しょうが」,「ビール」,「昆布」,「醤油」,「納豆」の 5 キーワードである．これらには，同時ヒット数の値が周期的に増減を繰り返す傾向が見られた．予測時点においては同時ヒット数の値が減少傾向にあったのが，2003 年 3 月から 5 月にかけては増加していた．過去の同時ヒット数の変化過程をもとに将来流行することが予想されるものは斬新な組合せとはいえないため，本研究で探したい有効な組合せとは異なる．したがってシステムによって予測されないことに問題はない．

また，表 9 中においてシステム出力にはあるが，同時ヒット数は上昇しなかったキーワードは 15 個ある．このうち同時ヒット数が減少したものは「マヨネーズ」,「ポテト」,「ラーメン」の 3 個である．システムはカテゴリー拡張側キーワードの新規性の評価値と組合せの独創性の評価値がともに高い値であったので有効な組合せとして 3 キーワードを出力した．しかし，実際には組合せが生成されず，その後，組合せの同時ヒット数が下がり廃れる結果となったことから，組合せが全く注目されなかったか，気づいた人はいたが実際に採用されるほどのものではなかったと考えられ，新たな流行を作るには不十分なキーワードであったといえる．

一方同時ヒット数が増加も減少もしていないものは「うなぎ」,「目玉焼き」,「漬物」,「ポテトチップス」,「おかか」等があった．このうち「うなぎ」と「目玉焼き」は 2000 年 3 月から 2003 年 2 月まで組合せの評価に用いていない期間を含めて，同時ヒット数が連続して増加も減少もしていないことから，これらとの組合せは将来出てくるかもしれない．しかし「漬物」,「ポテトチップス」,「おかか」は組合せの評価に用いた期間の 2002 年 6 月から 2003 年 3 月までの同時ヒット数は増加も減少もしなかったが，2000 年 3 月から 2002 年 6 月までは同時ヒット数が減少傾向にあった．現時点では評価に用いるデータ量を経験的に決定しているため,「うなぎ」,「目玉焼き」と「漬物」,「ポテトチップス」,「おかか」が評価関数にとっては同じものとなってしまった．後者はシステム出力には含まれるべきではないので，評価に用いるデータ量や時間変化をとらえる評価式をより柔軟かつ正確な表現に改善する必要がある．

表 11 にライン拡張側キーワードごとのシステム出力があるときとないときに正解となる組合せを発見するまでの試行回数を示す．システムがない状態では，全 2,096 個の組合せに対して流行するかどうかの検証を行う必要があり，この場合だと平均して 150 回程度試行を重ねないと正解となる組合せが含まれない．しかし，システムによって検証する組合せが絞り込まれた状態だと，平均 2 回から 3 回で正解となる組合せが含まれる．

本システムを使用する上での最大のメリットは，検証対象の組合せの中にはやりとなる組合せが含まれるまでの平均試行回数を減らすことができる点である．例えば，新商品を開発するためにどの組合せを用いるかを選択する際，システムがない状態では考えられるすべての組合せに対して流行するかどうかの検討をする必要があり，実際に流行する組合せを発見するまでにはより多くの試行を重ねる必要がある．しかしシステムを用いるならば，より少ない試行回数で実際に流行する組合せが出力中に含まれる．したがって，本システムははやりとなる組合せの発見に大いに役立てら

れる．

5.3 ARMA モデルによる時系列予測との比較

本節では提案した組合せの評価式と自己回帰移動平均モデル（ARMA モデル）[14] による同時ヒット数の時系列予測による有効な組合せの予測精度の比較を行った実験について述べる．ARMA モデル（AutoRegressive Moving Average）は AR モデルと MA モデルを組み合わせた時系列分析モデルの一つで，株価や為替レートの時系列データなどの規則性に従って変動する数値データを記述するときに用いられるモデルである．一般的に時系列データの集合 $\{y_t\}$ に対して，p 次の AR(p) モデルは定数 k_{AR}，$\{a_p\}$，誤差項を u_t とすると式 (7) で表される．

$$y_t = k_{AR} + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_p y_{t-p} + u_t \tag{7}$$

対して q 次の MA(q) モデルは $\{y_t\}$ に対して，定数 k_{MA}，$\{b_q\}$ とすると式 (8) で表される．

$$y_t = k_{MA} + u_t + b_1 u_{t-1} + b_2 u_{t-2} + \cdots + b_q u_{t-q} \tag{8}$$

本実験では式 (9) で表される ARMA(2,2) モデルを使用した．

$$y_t = k_{ARMA} + a_1 y_{t-1} + a_2 y_{t-2} + u_t + b_1 u_{t-1} + b_2 u_{t-2} \tag{9}$$

k_{ARMA} は定数である．実験は，**5.1** と **5.2** で行ったものと同様とし，ARMA モデルの時系列データには，組合せを構成するライン拡張側キーワードとカテゴリ拡張側キーワードの同時ヒット数を用いた．ただし ARMA モデルによる出力は，式 (9) による評価値の上位を本提案システムと同数とした．表 12 に論文タイトル予測における再現率（正解となる組合せの数に対する予測できた組合せの数）を，表 13 に新商品予測における再現率を示す．両予測において提案手法が最も良い再現率が得られている．

この理由は，ARMA モデルにおいては 14 回から 16 回までの上昇率がそれほど大きくない場合や，14 回から 16 回までの上昇率が高くても 17 回で減少する組合せが多い場合には，上昇率のみからの予測が困難であったためである．しかし提案手法ではヒット数の上昇に加えて，組合せの独創性の評価値によって組合せの希少性に価値を見出して組合せを予測しており，新しい研究テーマや新商品開発などの新しいものを作り出す創造活動には，独創性の評価が必要であることがわかる．

表 12 論文タイトル予測における ARMA モデルと提案手法の再現率の比較
Table 12 Recall comparison between ARMA model and this system in prediction papers' titles.

Line Keyword	ARMA Model	This System
情報	1/2	2/2
知識	0/1	1/1
発見	2/2	2/2
ロボット	1/3	3/3

表 13 新商品予測における ARMA モデルと提案手法の再現率の比較
Table 13 Recall comparison between ARMA model and this system in prediction of new items.

Line Keyword	ARMA Model	This System
おにぎり	5/14	9/14
ラーメン	6/16	14/16
お茶	6/18	13/18

表 14 四つの評価式と提案手法の再現率の比較
Table 14 Recall comparison among four evaluation formula and proposed formula.

Line Keyword	eq.(2)	eq.(3)	eq.(4)	eq.(5)	eq.(6)
おにぎり	2/14	2/14	5/14	6/14	9/14
ラーメン	3/16	3/16	8/16	7/16	14/16
お茶	2/18	2/18	4/18	5/18	13/18

更に提案した四つの評価式（式 (2) から式 (6)）による再現率と式 (6) による再現率の比較を行った結果を表 14 に示す．予測個数を比較すると一つの評価式だけでは予測できる個数が少ないことが分かる．また各評価式によって予測できた組合せは異なっていたことと，式 (2) と式 (3) は組合せの独創性の一時的な値と時間変化，式 (4) と式 (5) はライン拡張側のキーワードの流行可能性をその一時的な値と時間変化からそれぞれ評価していること，式 (2) と式 (5) はキーワードの少なさを，式 (3) と式 (4) はキーワードの多さを評価しており，各評価式の役割がそれぞれ明確に異なっていることから，すべての評価式が有効な組合せの発見には必要であった．

6. むすび

本論文では世の中で新たに創造されるもののうち，既存のものの近似的な組み合わせによって構成されるものを対象として，独創的かつ世の中に広まる可能性が高い組合せを発見する組合せ評価システムを提案し

た．本システムはキーワードに対する大衆の認知度の変化や組合せの斬新さの変化を測ることで，将来流行する可能性の高いもののもととなる組合せを出力する．

システムはキーワードの組を出力するが，それらを具体的にどのように使えばよいかについては出力はしない．ユーザはシステムによって絞り込まれた選択肢の中からユーザ独自の感覚や好みに応じた組合せを選び，自らの手で深く掘り下げて創造活動を行う．我々は真に知的な作業を行うことできるのは人間のみであると考え，そのような人間の思考の材料として有効な組合せを提供するシステムを構築した．

今後は，各評価式をより実際の変化をとらえられるように改善し，三つ以上の組合せについても考えていきたい．三つ以上の組合せに関しては，二つのキーワードの組合せによって構成されるものが，ある程度の確率で世の中に出て流行すると仮定をすると，その作られたものを未来における新たなカテゴリー拡張側キーワードとしてとらえることができる．そうすることで未来から見た未来の時点における流行予測を行うことを考えている．未来から見た未来の時点における流行商品の予測を行うことで，創造活動が製品開発の場合は，現時点でどの製品の開発に力を入れればよいのかを決定するための 1 指針となり，研究活動の場合は現時点ではどの研究テーマに力を入れてそれをどの分野につなげていけばよいかを考える助けとなると考えられる．

文　　献

[1] 久保田秀和，西田豊明，“ユーザの過去発言を利用した複数エージェントによる創造的な対話の生成，” 信学論（D-I），vol.J84-D-1, no.8, pp.1222–1230, Aug. 2001.

[2] J. Weizenbaum, “Eliza - A computer program for the study of natural language communication between man and machine,” Commun. ACM, vol.9, no.1, pp.36–43, 1966.

[3] 高杉耕一，國藤　進，“スプリングモデルを用いたアイデア触発のための思考支援システムの構築，” 人工知能誌，vol.14, no.3, pp.495–503, 1999.

[4] 砂山　渡，谷内田正彦，“未来の流行を予測する Web からの注目キーワードの発見，” 日本知能情報ファジイ学会誌，vol.15, no.3, pp.309–317, 2003.

[5] 池田定博，金田重郎，金杉友子，“現代用語辞書を用いた流行語予測に関する統計的検討，” 信学技報，vol.100, no.541, 2001.

[6] 大澤幸生，谷内田正彦，“キーワード抽出法 KeyGraph の転用による地震履歴データからの要注意活断層発見支援，” 人工知能誌，vol.15, no.4, pp.665–672, 2000.

[7] 大澤幸生，ネルス ベンソン，谷内田正彦，“KeyGraph：単語の共起グラフの分割・統合によるキーワード抽出，” 信学論（D-I），vol.J82-D-I, no.2, pp.391–400, Feb. 1999.

[8] 中小企業のブランド戦略，中小公庫レポート，no.2003-3, 2004.

[9] 日記レンタルサイト（URL）http://www.diary.ne.jp/

[10] 松本裕治，北内　啓，山下達雄，平野義隆，松田　寛，浅原正幸，“日本語形態素解析システム「茶筌」version2.0 使用説明書第二版，” NAIST-IS-TR99012,（1999），（URL）http://chasen.aist-nara.ac.jp/index.html.ja

[11] 浅井達哉，安部賢治，川副真治，有村博紀，有川節夫，“XML データストリームからの高速な知識発見手法，” 人工知能学会第 17 回全国大会，2002.

[12] 樋口雄一，林　清鎮，渡部広一，河岡　司，“知能ロボットの自律移動のための実画像からの物体認識，” 人工知能学会第 17 回全国大会，2002.

[13] 松尾　豊，友部博教，橋田浩一，石塚　満，“Web からの人間関係ネットワークの抽出と情報支援，” 人工知能学会第 17 回全国大会，2002.

[14] 田中利彦，笹山　茂，坂上智哉，マクロ経済分析ー表計算で学ぶ経済学，中央経済社，1995.

（平成 16 年 1 月 8 日受付，5 月 17 日再受付）

西原　陽子

2003 阪大・基礎工・システム科学卒．現在，同大大学院博士前期課程在学中．

砂山　　渡

1995 阪大・基礎工・制御工学卒．1997 同大大学院博士前期課程了．1999 同大学院博士後期課程中退．同年同大大学院助手．2003 広島市立大学情報科学部助教授，現在に至る．博士（工学）．人間の創造活動を支援する研究に興味をもつ．

谷内田正彦　（正員）

1971 阪大大学院工学研究科修士課程了．同年同大基礎工学部助手，同助教授，教授を経て 1997 より同大大学院基礎工学研究科教授，現在に至る．博士（工学）．画像処理，人工知能，移動ロボットなどの研究を行っている．著書「ロボットビジョン」（昭晃堂），「コンピュータビジョン」（丸善，編著）など．情報処理学会，ロボット学会等各会員．