

典型的なオブジェクト選定のための カテゴリタイプにおける重要な観点の抽出

藤坂 達也[†] 湯本 高行^{††} 角谷 和俊^{†††}

[†]兵庫県立大学大学院環境人間学研究科 〒670-0092 兵庫県姫路市新在家本町 1-1-12

^{††}兵庫県立大学大学院工学研究科 〒671-2280 姫路市書写 2167

^{†††}兵庫県立大学環境人間学部 〒670-0092 兵庫県姫路市新在家本町 1-1-12

E-mail: [†]nd10u025@stshse.u-hyogo.ac.jp, ^{††}yumoto@eng.u-hyogo.ac.jp, ^{†††}sumiya@shse.u-hyogo.ac.jp

あらまし 認知心理学において、典型性はカテゴリ上で定義され、またその観点が複数存在すると提唱されている。しかし、カテゴリは多様であるため、どんな種類のカテゴリにおいてどの観点で典型度を算出すれば、人間の考える典型的なオブジェクトを抽出できるか定かではない。そこで、本研究ではカテゴリと典型性の観点の関係性を明らかにする一途として、Web を用いてカテゴリタイプごとの重要な観点を抽出する手法を提案する。具体的には、まず Web を用いた計算可能な指標となる観点を定義する。次に、カテゴリタイプを定義し、タイプごとにカテゴリを用意して、各観点におけるオブジェクトの値を算出する。さらに、被験者の判断した典型度との相関をとることで重要な観点を抽出し、典型性の観点とカテゴリタイプの関係性を考察する。

キーワード オブジェクトの典型性, 認知心理学

Extracting Important Viewpoints on Category Types for Selecting Typical Objects

Tatsuya FUJISAKA[†] Takayuki YUMOTO^{††} and Kazutoshi SUMIYA^{†††}

[†]Graduate School of Human Science and Environment, University of Hyogo

1-1-12 Shinzaik-honcho, Himeji, Hyogo, 670-0092 Japan

^{††}Graduate School of Engineering, University of Hyogo 2167 Shosha, Himeji, Hyogo, 671-2280 Japan

^{†††}School of Human Science and Environment, University of Hyogo

1-1-12 Shinzaik-honcho, Himeji, Hyogo, 670-0092 Japan

E-mail: [†]nd10u025@stshse.u-hyogo.ac.jp, ^{††}yumoto@eng.u-hyogo.ac.jp, ^{†††}sumiya@shse.u-hyogo.ac.jp

1. はじめに

ユーザにとって未知のオブジェクトが多く含まれるカテゴリから、典型性を基準にオブジェクトを選定した状況が多々存在する。たとえば、Amazon.com で「東野圭吾の書籍」を初めて購入しようと書籍リストを閲覧しているユーザがいるとする。しかし、その著者が書いた書籍は多く出版されているため、知識を持ち合わせていないそのユーザにとって、その中から選択を行うことは困難である。この際に、たとえば「東野圭吾といえば書籍 A」や「東野圭吾の代表作品は書籍 B」のように、典型性を考慮した選定が可能となれば、有用な支援が行なえと考えられる。

これまでの選択支援に関する先行研究として、信憑性や理解容易性を基準とした Web ページの選択支援

が提案されている[13,9]。しかし、本研究のように商品や店舗などのユーザの嗜好に関わるオブジェクトの場合は、典型性という新しい基準が必要であると考えた。そして、本研究では特に Murphy[8] の「人々は非典型的なオブジェクトより典型的なオブジェクトを好みやすい」という考えを参考に、未知のオブジェクト集合の中から典型的なオブジェクトを選定し提示することが、多くの場面で有効であると考えた。そのため、本研究ではあるカテゴリに属するオブジェクト集合が与えられた際に、そのカテゴリにおける典型的なオブジェクトを抽出する手法を考える。

これまで認知心理学の分野では、典型性の概念についての研究が多くされている。そして、そこではカテゴリにおいて典型度を定める観点が複数存在するとい

われている[2,7]. そのため, 本研究ではその分野で考えられてきた典型性の観点を整理し, その観点に基づき Web を用いて典型度を算出することで, あるカテゴリから人間の感覚に合う典型的なオブジェクトを抽出できると考えた.

しかし, カテゴリといっても多様な種類が存在する. そのため, どんなカテゴリにおいても典型的なオブジェクトを導出できる, 汎用的な観点が存在するかもしれない一方で, カテゴリごとに典型的なオブジェクトを導出できる観点が異なる可能性もある. すなわち, 典型性の観点とカテゴリの関係性が曖昧であるため, どんなカテゴリでどの観点を重視して典型度を算出する方が良いか定かではない. また, これを明らかにするためにカテゴリ 1 つ 1 つに対して評価を行い, 規則を導き出すことも考えられるが, 膨大な量のカテゴリが存在することから困難である.

そこで, 本研究ではオブジェクト集合の特徴から導き出されるカテゴリの性質をカテゴリタイプと定義し, Web を用いてタイプごとの重要な観点を抽出する手法を提案する. まず, 認知心理学の分野における典型性の概念を参考に, 本研究で扱う典型性の観点を整理する. 次にその観点を, Web を用いた計算可能な指標であると定義し, カテゴリタイプごとに用意されたカテゴリにおけるオブジェクトの各観点の値を算出する. タイプとしては, 集められたオブジェクトから導き出されるカテゴリの抽象度と, 集められたオブジェクトの要因の 2 つを軸に定義する. そして, タイプごとにどの観点から導き出された値が人間の考える典型度と相関があるか比較し, 重要な観点抽出の評価を行なう.

本論文の構成は以下のとおりである. まず, 2 章で関連研究について紹介する. 3 章で提案手法について説明する. 4 章で実験と考察について述べ, 5 章でまとめを述べる.

2. 関連研究

2.1 認知心理学分野での典型性

認知心理学の分野では, 典型性について古くから考えられてきており, その理論がいくつか存在している. それらにおいて, 典型性はカテゴリという概念上で定義されるという共通した前提条件がまず存在するとされている. そして, カテゴリと典型性の関係を示す主要理論の 1 つにプロトタイプ理論が定義されている.

Rosch によって導き出されたプロトタイプ理論[10, 11]では, カテゴリに属するための必要十分条件は必要なく, 最も典型的なオブジェクト (プロトタイプ) を中心に, それとの類似によって他のオブジェクトの典型性が導出される. また, プロトタイプ選定のための認知科学的なユーザ実験も複数行われている.

そのため, 本研究ではこれまでも典型性について考えてきた[6]が, ここでは認知心理学で研究されてきた典型性の概念を参考にすることで, オブジェクトの典型性を考える. そして, 特に Web で簡便に計算可能な指標となる観点を定義することで, 自動的に典型度を算出する手法を提案する.

2.2 オブジェクトの典型度算出

近年のオブジェクトの典型度を測る先行研究として, 佖らは心理学者の Barsalou が提案した central tendency, ideal, frequency of instantiation の 3 つの典型性の観点[2]に着目し, それらに基づき分析対象の軸と典型性のタイプの軸から典型性の観点を整理し, COOKPAD をコーパスとしたレシピの典型度の算出する手法を提案し, 各観点の有用性を検証している[12]. 一方, Cai らは典型性に基づき人々の考えや感覚を考慮することは, 従来の推薦システムの改良につながると考え, プロトタイプ理論に基づく映画サイトにおける推薦手法の提案している[3, 4].

これらの研究と本研究は, 認知心理学の分野で考えられてきた定義や理論を参考にオブジェクトの典型性について考えている点で類似している. しかし, これらの研究では料理レシピサイトにおいて「カルボナーラ」のような特徴の類似するレシピが存在するカテゴリや, 映画ごとの説明文から類似する映画を分類したカテゴリなど単一のカテゴリでのみしか実験を行っていないため, 手法の汎用性が低いといえる. そのため, 本研究ではカテゴリに応じた典型的なオブジェクトの選定を可能とするために, カテゴリタイプに着目し, タイプごとに重要な観点の抽出を目指す.

3. カテゴリタイプにおける重要な観点の抽出

3.1 典型性の観点の整理

本節では, カテゴリにおいて定義される典型性の観点を, 認知心理学の分野での考え方を参考にすることで整理し, 本研究で扱う 4 つの観点について説明する.

まず, 認知心理学の分野におけるカテゴリの典型性の観点についての先行研究[2,5,7]を調査した. その結果, 以下に示す複数の観点が存在することが分かった. それぞれの観点とその考え方について, 簡単に紹介する.

1) 理想属性

カテゴリが提供する目的として連想される理想的な属性であり, それを持つオブジェクトは典型性が高いと位置付けられる. たとえば, ダイエットフードの場合であれば, 「低カロリー」が理想属性となる.

2) 多属性構造

ユーザがオブジェクトを選択する際に影響する主要

な信念となる項目群のことを指し、それを持ち合わせているオブジェクトほど典型的と位置付けられる。

3) 家族的類似性

カテゴリ内のそれぞれのオブジェクトは異なる存在であるが、いくつかの類似した属性を持ち合わせているという考え方で、この場合類似した属性を多く共有するオブジェクトほど典型性が高く位置付けられる。

4) 代表性

人が考えるカテゴリの中心的存在とオブジェクトとのあてはまる度合いで、それが高いほど典型性が高い。

5) カテゴリの事例としての接触頻度

現実空間において触れる程度で、それが高いほど典型的なオブジェクトとして位置付けられる。

6) 親密性

馴染みの程度で、それが高いほど典型的なオブジェクトとして位置付けられる。

7) 第一想起または想起率

カテゴリのオブジェクトとしての思い浮かべやすさで、それが高いほど典型性が高く位置付けられる。

8) 独自性（弁別性）

他のオブジェクトとの違いや検索の手がかりとなるユニークな属性を表し、その性質を持つほど典型的とみなされる。

このように、典型性を構成すると考えられる観点は様々あるが、本研究では Web で簡便に計算可能な指標となる観点を用いる。すなわち、4) 代表性、5) カテゴリの事例としての接触頻度、6) 親密性、7) 第一想起または想起率を本研究で扱う観点とする。また、7) を「想起されやすさ」とすることで1つの観点とする。上記の4つの観点は、カテゴリ名やオブジェクト名のみを用いて値を容易に算出可能であるが、1), 2), 3), 8) の観点は、オブジェクトに関する属性情報が必要である。しかし、それを客観的に決めることは難しく、さらに自動で抽出する手法を考案しなければならない。そのため、本研究ではそれらは用いない。

3.2 Web を用いたオブジェクトの各観点における指標値

本節では、代表性、カテゴリの事例としての接触頻度、親密性、想起しやすさの4つの観点を、Web を用いた計算可能な指標と定義し、観点ごとに値を算出する手法について説明する。

3.2.1 代表性

代表性は、「人間が考えるカテゴリの中心的存在とオブジェクトとのあてはまる度合い」と定義されている。直感的に考えると、これはカテゴリの基準となるオブジェクトとしてどれくらい考えられているかであ

る。そのため、オブジェクトが比較対象として用いられている Web ページ数を指標にできると考えた。

たとえば、ある人によって「チョコレートパフェに比べ、このパフェは甘さが控えめだった」や「チョコレートパフェよりも、抹茶パフェは見た目が綺麗」と書かれた文書が Web ページに存在しているとする。この場合、書き手が「このパフェ」や「抹茶パフェ」を説明するために、「チョコレートパフェ」を比較対象としている文章と考えることができる。つまり、このような文書が多くの人によって書かれている場合、「チョコレートパフェ」はカテゴリの基準となるような代表性が高いオブジェクトと見なせる。

従って、オブジェクトが比較対象として用いられている Web ページ数が、代表性を測る指標になると仮定し、計算する。まず、比較対象を抽出するためのクエリを用意する。たとえば、「チョコレートパフェに比べ」という文書は、「チョコレートパフェ」を比較対象とした文書と考えることができる。この場合、「 o_i に比べ」をクエリとしたフレーズ検索を用いることで、 o_i を比較対象としている文書を取得できる。そのような考え方から、「 o_i に比べ」、「 o_i と比べ」、「 o_i よりも」のフレーズ検索クエリを用意する。次に、それらを用いて Web 検索した際の検索結果件数を抽出する。ここでの検索結果件数は、 o_i が比較対象として用いられる数を表しており、それが高いほど代表性が高いことを意味する。

ここで、代表性の値 $rep(c, o_i)$ の算出式を式(1)に示す。 FQ はフレーズ検索クエリの集合を表し、 $H_{web}(q(o_i))$ は、 o_i をクエリとするフレーズ検索時の検索結果件数を表す。

$$rep(c, o_i) = \log \left(\sum_{q \in FQ} H_{web}(q(o_i)) \right) \quad (1)$$

3.2.2 カテゴリの事例としての接触頻度

カテゴリの事例としての接触頻度は、「現実空間においてカテゴリの事例として触れる程度」と定義されている。直感的に考えると、これは現実空間でどれくらい目にしやすいかである。現実空間を Web 空間に見立てた際、たとえばカテゴリ名で検索し、その結果の上位の目立つ場所でカテゴリ名と共に現れるオブジェクトは、人が目にする確率が高いと考えることができる。

従って、本研究では Web の検索結果の順位におけるオブジェクトの出現数が、接触頻度を測る指標になると仮定し、計算する。まず、検索エンジンを用いてカテゴリ名 c のクエリから上位 n 件の Web ページのタイトルとスニペットを取得する。上位 n 件とするのは、ランキングの上位に現れるオブジェクトを対象とする

ためである。また、タイトルとスニペットに絞るのは、検索結果のリンクを辿ることなく、人が閲覧できる文書であり、目にしやすいと考えたためである。次に、タイトル、スニペットそれぞれで c に属するオブジェクト o_i の出現した順位とそこでの出現数(tf)を計算する。そして、 n 件中でのタイトルまたはスニペットでの o_i の出現数(df)を計算する。これより、下記の式において接触頻度の値 $con(c, o_i)$ を算出する。

$$con(c, o_i) = w_1 \times \sum_{j=1}^n \left(\frac{tf_{title}(j, o_i)}{j} \right) \times df_{title}(o_i) + (1 - w_1) \times \sum_{j=1}^n \left(\frac{tf_{snippet}(j, o_i)}{j} \right) \times df_{snippet}(o_i) \quad (2)$$

ここでの、 $tf_{title}(j, o_i)$ は c をクエリとして得られた j 位の Web ページのタイトル中での o_i の出現数を表し、 $tf_{snippet}(j, o_i)$ は c をクエリとして得られた j 位の Web ページのスニペット中での o_i の出現数を表す。つまり、上位のタイトルまたはスニペットに o_i が多く出現しているほど、値が大きくなる。また、 $df_{title}(o_i)$ と $df_{snippet}(o_i)$ は n 件中でのタイトルまたはスニペットでの o_i の出現数を表し、最大値は n となる。この際、Web ページのタイトルにオブジェクトが出現する方が、より目立つところに含まれるため、閲覧されやすいと考えた。そのため、タイトル、スニペットそれぞれの結果の値に w_1 の重みをかけることで、最終的に $con(c, o_i)$ の値を算出する。

3.2.3 親密性

親密性は、「馴染みの程度」と定義されている。直感的に考えると、これは人々がどれくらいオブジェクトを経験してきたかに関係がある。つまり、小さい頃からよく食べてきた「チョコレートパフェ」は、他のパフェよりも人にとって馴染みのあるパフェとなる。この際、たとえば「チョコレートパフェ」の方が「チーズパフェ」より、人々にとって馴染みのあるパフェならば、Web においてもそのパフェの方が多くの Web ページで記述されていると推測できる。

そこで、本研究ではオブジェクトについて言及するために作成した Web ページの数が、親密性を測定する指標になると仮定し、計算する。ここでは、式(3)に示すように、 o_i をクエリとして Web 検索した際の検索結果件数 $H_{web}(o_i)$ を、親密性 $fre(c, o_i)$ として算出する。

$$fre(c, o_i) = \log(H_{web}(o_i)) \quad (3)$$

3.2.4 想起しやすさ

想起しやすさは、「カテゴリのオブジェクトとしての思い浮かべやすさ」と定義されている。そのため、

Web の画像検索を用いることができると考えた。

たとえば、カテゴリ名で画像検索した際、全体的に画像としてよく現れるオブジェクトは、人にとって想起されやすいと考えることができる。また、オブジェクト名で画像検索した際の検索結果件数も想起されやすさにつながると考えられる。これは、画像として認識される可能性を意味し、人の思い浮かべやすさに依存していると捉えることができる。

従って、本研究では画像検索における結果のページが、想起しやすさを測る指標になると仮定し、2 つの手法から計算する。まず、カテゴリ名で検索した際に得られた検索結果をサンプリング結果として用いる手法を説明する。始めに、画像検索エンジンを用いてカテゴリ名 c のクエリから m 件の画像を取得する。 m 件は、サンプルとして取得する数を表す。次に、画像に写るオブジェクト o_i を分析するために、ここでは画像のラベルを用いる。つまり、画像ラベル内の o_i の頻出数を計算することで、サンプルの中でのオブジェクトの現れやすさを測定する。そして、サンプル中で多くの画像で頻出するオブジェクトは、想起しやすさが高いと判断する。

しかし、画像検索の場合、ラベルという短い文字列中に o_i が必ずしも出現するとは限らず、正確な算出ができない可能性がある。そこで、もう 1 つの手法として画像検索におけるオブジェクト単位での検索結果件数を用いる。この際、 o_i をクエリとして得られた件数は、画像としての人の思い描きやすさと考え、その値が高いほど想起されやすいことを意味する。

従って、本研究では下記の式で示したように、2 つの手法から得られた値を加算した値を、想起されやすさの値 $ima(c, o_i)$ とする。ここでの $tf_{label_j}(o_i)$ は、画像 j のラベル中における o_i の頻出数を、 $H_{image}(o_i)$ は画像検索における o_i の検索結果件数を表す。

$$ima(c, o_i) = \log \left(\sum_{j=1}^m tf_{label_j}(o_i) \right) + \log(H_{image}(o_i)) \quad (4)$$

3.3 カテゴリタイプ

カテゴリには、様々な種類が存在するが、本研究では 2 つの軸からカテゴリタイプを定義する。まず 1 つ目の軸として、カテゴリの抽象度に着目する。たとえば、「家具」のような抽象度の高いカテゴリ、「椅子」のような中間的な抽象度のカテゴリ、「キッチンチェア」やのような抽象度の低いカテゴリが存在する。これらは、オンラインショッピングにおける階層型のカテゴリ構造と考えることができる。

また、この際に中間的な抽象度を決定する方法が、

表 1 実験に用いるカテゴリタイプにおけるカテゴリ

		抽象度		
		高		低
要因	内的	<u>タイプ1 (4カテゴリ)</u> ・乗り物 (18個) -自動車, バイク, 飛行機 etc. ・スポーツ (18個) -サッカー, 野球, テニス, 卓球, バトミントン, カーリング etc. ・楽器 (18個) -ピアノ, ギター, サックス, フルート, リコーダー etc. ・飲み物 (10個) -水, ジュース, お茶, 紅茶 etc.	<u>タイプ2 (4カテゴリ)</u> ・花 (18個) -バラ, あじさい, ひまわり, 桜, etc. ・ジュース (14個) -オレンジジュース, パナナジュース, りんごジュース, ぶどうジュース etc. ・パスタ (15個) -ミートソース, カルボナーラ, ボロネーゼ, ペスカトーレ etc. ・鍋 (18個) -キムチ鍋, ちゃんこ鍋, チゲ鍋 etc.	<u>タイプ3 (4カテゴリ)</u> ・トヨタ (17個) -プリウス, マークX, カローラ etc. ・チョコレートケーキ (9個) -ガトーショコラ, オペラ, エクレア, ブラウンー etc. ・腕時計 (14個) -ロレックス, カシオ, オメガ etc. ・パフェ (19個) -イチゴパフェ, 抹茶パフェ, チーズパフェ, とんかつパフェ etc.
	外的	<u>タイプ4 (3カテゴリ)</u> ・ペット (13個) -犬, 猫, うさぎ, ハムスター, 熱帯魚 etc. ・元気を出す手段 (6個) -名言, 音楽, 栄養ドリンク, サプリメント etc. ・感動できるコンテンツ (6個) -本, 映画, テレビ, 音楽 etc.	<u>タイプ5 (3カテゴリ)</u> ・室内用ペット (11個) -チワワ, ロシアンブルー, トイプードル, コーギー etc. ・元気が出るアイテム (6個) -ニンニク, レッドブル, 相田みつを, アロマキャンドル etc. ・ダイエット方法 (7個) -バナナ, 筋トレ, コアリズム, プロテイン etc.	

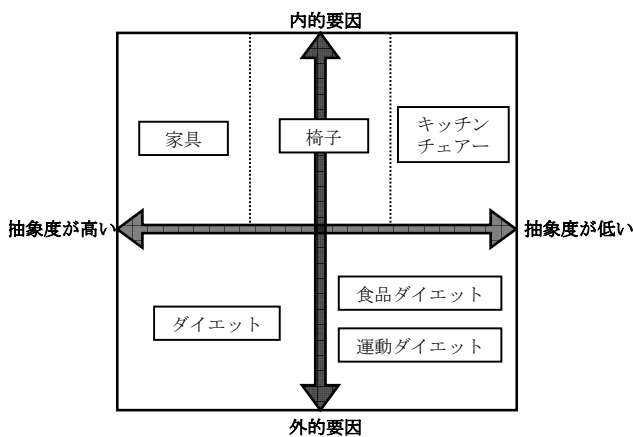


図 1 カテゴリタイプ

先行研究[11]で考えられている。ここでは、そのオブジェクトを被験者に視覚的に見せた際の回答から判断できるとしている。たとえば、家具としてキッチンで使う椅子が写った写真を提示した際、人間はそれを「家具」や「キッチンチェア」と答えるよりも、「椅子」と回答する傾向にある。その際、「椅子」が中間的なカテゴリと見なされる。

次に、2 つ目の軸として、そのカテゴリにおけるオブジェクトがどのような意味合いで集められたかという要因に着目する。ここでは、オブジェクトの内的要因と外的要因の2 つに大きく分類できると考えた。

たとえば、「鳥」カテゴリでは「飛べる」、「羽毛」、「くちばし」といった特徴が類似するオブジェクトが集まるため、オブジェクト自身の性質（内的要因）から構成されるカテゴリと捉えることができる。一方、「ダイエットに効くアイテム」では、カテゴリのオブ

ジェクトを決定するための目的が存在しており、それに基づき集められる。従って、オブジェクト自身というより、それが目的に合致するかといった外的要因から構成されるカテゴリと捉えることができる。

そのため、本研究では図 1 に示す 2 つの軸からカテゴリのタイプを決定し、そのタイプにおいて 3.2 で定義した典型性の観点から、重要な観点を導出する。

4. 実験と考察

4.1 実験方法

本節では、代表性、接触頻度、親密性、想起しやすさの 4 つ観点に基づき、カテゴリタイプごとに重要な観点が存在するか、実験より考察する。そのため、表 1 に示したように抽象度という軸と要因という軸からカテゴリのタイプを 5 つに分類し、各タイプにおいて数個のカテゴリを用意する。表中の括弧内の数はそのカテゴリにおけるオブジェクト数を表し、その下にオブジェクトの例も示す。また、目的に合致するかといった外的要因から構成されるカテゴリは、中間的な抽象度の判定が難しいため、ここでは相対的に抽象度の高いカテゴリ、低いカテゴリの 2 つのタイプで実験を行う。図 2 に沿って実験方法を説明する。

まず、人手によりカテゴリにおける典型度を評価し、正解データを作成する。この際、プロトタイプ理論に則り、まずカテゴリの中で最も典型的なオブジェクトを選定し、それとそれぞれのオブジェクトを比較することで、典型度を 7 段階 (7: 最も典型度が高い) で評価する。今回は、著者が評価を行なう。

次に、3.2 で示した手法により、各観点におけるカテゴリごとのオブジェクトの値を算出する。この際、4

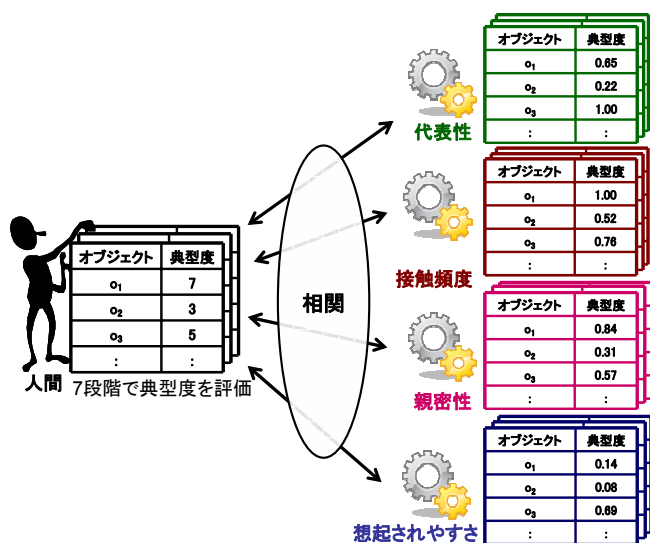


図2 実験方法の概要

つの観点の値を算出するために Google の検索結果を用いる。また、接触頻度の測定では上位 1000 件 ($n=1000$) の Web ページを取得し、それを本実験の対象データとする。また、算出の際の重み(w_1)は 0.7 とする。想起されやすさの測定では、Google の画像検索エンジンを用いて、画像のラベル 400 件 ($m=400$) をサンプリングデータとし、検索結果件数もその検索エンジンから得られた結果を用いる。これらの条件より、4 つの観点で値を算出する。そして、観点ごとの結果で値が最大のオブジェクトを選定し、その値で正規化する。それより算出された値を、それぞれのオブジェクトの典型度とする。

最後に、人手による典型度と各観点から導出された典型度を比較する。この際、人手による典型度との近さを 4 つの観点で評価するために、相関をとる。本研究では、様々な相関分析の中から式(5)に示すピアソンの積率相関係数[1]を求める。

$$\text{ピアソンの積率相関係数} = \frac{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (5)$$

ここでは、2 組の数値からなるデータ列 X, Y が与えられたとき、相関係数が求められる。分子は変数 X と変数 Y の共分散、分母はそれぞれの変数の標準偏差である。また、 X_i と Y_i はそれぞれのデータ列の要素、 $\bar{X}$ と $\bar{Y}$ は相加平均、 n はデータ列の長さを表す。この際、ランキングの相関を求める、スピアマンの順位相関係数[1]も方法として考えられた。しかし、たとえば 0.01 の差があるオブジェクト同士に対して、明確な順位付

表2 タイプ1での人手による典型度と各観点の相関係数

	乗り物	スポーツ	楽器	飲み物	平均
代表	0.65	0.71	0.57	0.61	0.64
接触	0.94	0.84	0.83	0.75	0.84
親密	0.90	0.82	0.81	0.70	0.81
想起	0.94	0.90	0.76	0.59	0.80

表3 タイプ2での人手による典型度と各観点の相関係数

	花	ジュース	パスタ	鍋	平均
代表	0.55	0.69	0.92	0.62	0.70
接触	0.76	0.88	0.81	0.75	0.80
親密	0.67	0.60	0.82	0.35	0.61
想起	0.64	0.59	0.85	0.30	0.59

表4 タイプ3での人手による典型度と各観点の相関係数

	トヨタ	チョコレートケーキ	腕時計	パフェ	平均
代表	0.61	0.82	0.68	0.70	0.70
接触	0.33	0.62	0.66	0.76	0.59
親密	0.54	0.83	-0.23	0.32	0.37
想起	0.39	0.73	0.07	0.45	0.41

表5 タイプ4での人手による典型度と各観点の相関係数

	ペット	元気を出す手段	感動できるコンテンツ	平均
代表	0.80	0.63	0.22	0.55
接触	0.93	0.69	0.82	0.81
親密	0.64	0.05	-0.49	0.07
想起	0.78	0.61	-0.02	0.46

表6 タイプ5での人手による典型度と各観点の相関係数

	室内用ペット	元気が出るアイテム	ダイエット方法	平均
代表	0.78	0.82	0.79	0.80
接触	0.75	0.15	0.43	0.44
親密	0.67	0.66	0.25	0.53
想起	0.79	0.78	0.24	0.60

けはできないと考え、人手による典型度と各観点の値の相関を求める方が正確な評価になると考え、ここではピアソンの積率相関係数を採用する。

これより、カテゴリタイプごとに重要な観点を決定

できるか実験し、典型性の観点とカテゴリタイプの関係性を考察する。

4.2 実験結果と考察

実験結果として、表 2～6 に人手による典型度と各観点における典型度の相関係数を示す。ここでは、タイプごとに用意したそれぞれのカテゴリにおける相関係数を示し、グラフの右端にはそのタイプでの平均を示す。各観点について考察を行なう。

まず、代表性のそれぞれの表の結果を見てみると、内的要因/外的要因に関わらず、抽象度が高い場合（表 2, 5）は相対的に相関係数が低く、抽象度が低くなる（表 3, 4, 6）につれて相対的に相関係数が高くなる傾向が見られた。従って、代表性は特に抽象度が低いカテゴリにおいて重要な観点になり得ると推測される。

次に、接触頻度のそれぞれの表の結果を見てみると、内的要因/外的要因ともに、抽象度が高い場合（表 2, 5）は相対的に相関係数が高く、抽象度が低くなるにつれて相対的に相関係数が低くなる（表 3, 4, 6）傾向が見られた。すなわち、代表性とは逆の結果を示した。従って、接触頻度は特に抽象度が高いカテゴリにおいて重要な観点になり得ると推測される。

また、親密性と想起されやすさのそれぞれのカテゴリタイプの結果を見てみると、内的要因で集められたカテゴリタイプ（表 2, 3, 4）では抽象度が高い場合は相対的に相関係数が高く、抽象度が低くなるにつれて相対的に相関係数が低くなる傾向となり、接触頻度の結果と同様であった。一方、外的要因で集められたカテゴリタイプ（表 5, 6）では、抽象度が高い場合に相対的に相関係数が低く、抽象度が低くなるにつれて相対的に相関係数が高くなる傾向となり、代表性の結果と同様であった。

しかし、この 2 つの観点における相関値は、接触頻度または代表性に比べ低いことが各表より明確である。つまり、人手による典型度とずれがあると考えられ、今回定義したカテゴリタイプにおいては重要な観点にならないと考えられる。従って、この実験から以下の 2 つの知見が得られた。

- 抽象度が低いカテゴリにおいて、代表性が重要な観点になり得る
- 抽象度が高いカテゴリにおいて、接触頻度が重要な観点になり得る

5. おわりに

本研究では、Web を用いてカテゴリタイプごとの典型性の重要な観点を抽出する手法を提案した。ここでは、まず Web を用いた計算可能な指標として代表性、接触頻度、親密性、想起されやすさの 4 つの観点を定

義した。そして、実験でカテゴリタイプごとのカテゴリにおけるオブジェクトの値を算出し、人手による典型度と比較した。その結果、抽象度が低いカテゴリにおいて代表性、抽象度が高いカテゴリにおいて接触頻度が重要な観点になり得ることが知見として得られた。

今後は、各観点における値の算出アルゴリズムを改良し、観点を組み合わせた際の結果の考察を行うことがあげられる。また、カテゴリの数を増やして実験を行うことで、得られた知見の妥当性を示す必要がある。さらに、今回定義したカテゴリタイプだけでなく、新しいカテゴリタイプについて考え、典型性の観点とカテゴリタイプの関係性を考察することがあげられる。

参 考 文 献

- [1] 青木繁伸: R による統計解析, オーム社 (2009)
- [2] L.W. Barsalou. “Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories”, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, pp. 629–654, 1985.
- [3] Y. Cai and H. Leung, “Multi-prototype concept and object typicality in ontology”, *Proc. of the 21st International Florida Artificial Intelligence Research Society Conference*, pp. 470–475, 2008.
- [4] Y. Cai, H. Leung, Q. Li, J. Tang and J. Li. “Recommendation based on object typicality”, the 19th ACM international conference on Information and knowledge management (CIKM2010), pp. 1529–1532, 2010.
- [5] J. Cohen and K. Basu, “Alternative Models of Categorization: Toward a Contingent Processing Framework”, *Journal of Consumer Research*, volume 13, pp. 455–472, 1987
- [6] 藤坂 達也, 湯本 高行, 角谷 和俊, “典型性を考慮した同位オブジェクトの検索方式” Web とデータベースに関するフォーラム, 2G-1-1, 2011.
- [7] B. Loken and J. Word, “Alternative Approaches to Understanding the Determinants of Typicality”, *Journal of Consumer Research*, pp. 111–126, 1990
- [8] G. L. Murphy, “The big book of concepts”. MIT Press, 2002.
- [9] M. Nakatani, A. Jatowt, and K. Tanaka, “Adaptive Ranking of Search Results by Considering User's Comprehension”, *Proc. of the 4th International Conference on Ubiquitous Information Management and Communication (ICUIMC 2010)*, pp. 182–192, 2010.
- [10] E Rosch, “Natural categories”, *Cognitive Psychology* 4, pp. 328–350, 1973.
- [11] E Rosch, “Principles of Categorization”, *Cognition and Categorization*, Lawrence Erlbaum Associates, pp. 27–48, 1978.
- [12] 佃 洸撰, 中村 聡史, 山本 岳洋, 田中 克己 “オブジェクトの典型度分析とその検索への応用”, Web とデータベースに関するフォーラム, 2G-1-2, 2011.
- [13] Y. Yamamoto and K. Tanaka “Enhancing Credibility Judgment of Web Search Results”, *Proc. of the 29th international conference on Human factors in computing systems (CHI2011)*, pp. 1235–1244, 2011.