

概念体系間の比較と重ね合わせによる関連発見

Comparison and Superposition Methods for Discovery of Relationship between Concept Systems

伊藤 英毅 武田 英明

Hidetake Itoh Hideaki Takeda

奈良先端科学技術大学院大学 情報科学研究科

Graduate School of Information Science, Nara Institute of Science and Technology

Abstract: Many concept systems are used in our daily life, and even there exist some different concept systems for single area. We need methods to understand relationship among such various concept systems. In this paper, we propose comparison and superposition methods to discover relationship between concept systems such as completion of one system by the other and discover of new aspect. In order to realize then, we first propose comparison and matching methods, then show a superposition method. Then we evaluate our methods with classification systems used in industry. Matching between categories are measured by correspondence ratio and coverage ratio. The former is a ratio the number of matched items in a category to the total of the category, and it can indicate how much a category is matched to the other category. The latter is a ratio the number of matched items in a category to the total of the corresponded category, and it can indicate relationship between the original and corresponded categories in size. Then we can impose categories in a concept system to the other by using these two ratios. We examined these methods with two classification systems used in industry, and can identify characteristics of these system by these methods

1 はじめに

1.1 既存の概念体系

近年のコンピュータとネットワークを中心とした情報基盤に多くの活動や対象物の概念が取り込まれている。結果として多くの概念体系が存在している。それらの概念体系は以下のように分類整理することができる。

1. 記述体系

概念の振る舞いや関係が概念の間の相互関係として記述されたもの。概念間の関係の種類は様々な関係がある。例としてマニュアル、教本、料理本のレシピ等。

2. 分類体系

対象とする分野に含まれる概念を、ある分類基準で分類したもの。分類体系はさらに次のように分類できる。

(a) 網羅的でない分類: 対象分野を網羅的に分類されているでもなく、複数のカテゴリに属する等の重複があり得る分類。例として、料理本の料理一覧、yahoo等のディレクトリサービスのカテゴリ等。

(b) 網羅的な分類: 対象分野の全ての概念がいずれかのカテゴリの属するように、網羅的に分類したもの。例として、職業別電話帳の職業分類、商品分類等。また、情報処理や標準化を目的としたコード体系等。

3. 語彙、用語体系

対象とする概念の語彙や用語を定義したもの。例としてシソーラスや専門分野の専門用語辞書等。

1.2 複数の分類体系の利用

本研究では、前節で述べた概念体系の一つである分類体系に注目する。分類体系は同一分野においても多数存在している。それぞれの分類体系は対象となる世界をそれぞれの視点によって体系化されている。例えば、学校にある机はメーカー等の商品カタログにおける

商品分類から、学校が管理している備品としての分類もある。これらの複数の分類体系は異なる目的や用語で作成されたり、対象の異なる属性に注目している等のため、関連性が明確ではない。しかし、体系間の相互の関連性を発見し体系相互の情報を利用することによって、情報の補完（ないものが見つかる）や発想の転換（ない視点が見つかる）を期待できる。そこで、本研究では分類体系間の比較及び重ね合わせにより複数体系間の関連性発見及び関連付けの手法を提案して実施評価を行う。尚、本研究では概念間の関連性として類似関連と上位下位関連に焦点を当てる。

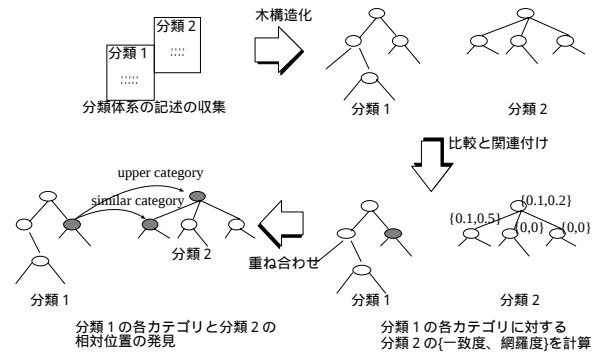


図 1: 分類体系間の比較、重ね合わせの流れ

2 分類体系間の比較と重ね合わせ

2.1 概要

図 1 は本研究においての分類体系間の比較から重ね合わせの流れの概要を示したものである。分類体系の木構造化では対象とする分類体系を木構造で表現し、分類体系の階層やカテゴリと要素の関係を明らかにする。続いて、比較と関連付けではカテゴリ間の要素の一致を調べカテゴリ間の関連を決定する。この際に、一致度と網羅度というパラメータを導入し、カテゴリ間の類似関係、階層関係の指標を与える。重ね合わせは関連付け結果を利用した分類体系間の相対位置の発見を行い、人の判断や基準において、関連付け結果の妥当性を確認する。

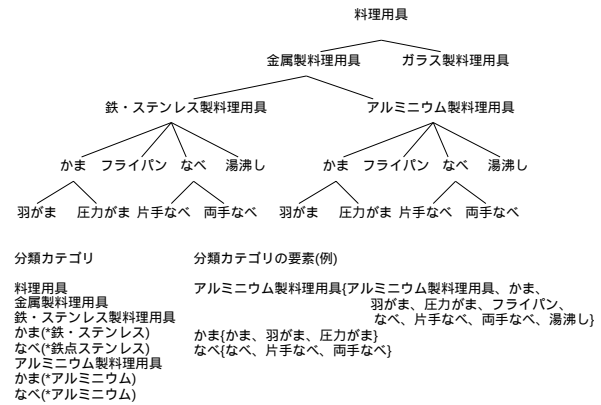


図 2: 分類体系の木構造

2.2 分類体系の構造

分類体系の中で最も多い階層型分類は、ある分類基準で枝分かれする木構造で表現することができる。各節は概念の共通の属性を集めた分類のカテゴリとなる。これを分類カテゴリと呼ぶ。各カテゴリにはその分類カテゴリに属する要素が存在する。カテゴリに含まれる要素を調べる場合には、その対象となるカテゴリ以下の階層の要素すべてが対象となる。その時、カテゴリの要素にさらにカテゴリが含まれる場合はそのカテゴリも要素の一つとして含める。例えば、図 2 の「アルミニウム製料理用具」の要素は、アルミニウム製料理用具、かま、羽がま、圧力がま、フライパン、なべ、片手なべ、両手なべ、湯沸しとなる。

2.3 分類体系間の比較と関連付け

2.3.1 カテゴリ間の要素の比較

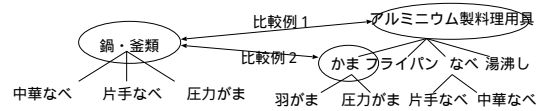
分類体系間の比較は、比較するカテゴリ間の要素の名前（ラベル）の一致を調べる。図 3 の比較例 1 では比較元カテゴリとして「鍋・釜類」を、比較先カテゴリと

して「アルミニウム製料理用具」を比較した場合の例であり、この場合 2 つのカテゴリ間の一致要素は「中華なべ」「片手なべ」「圧力がま」の 3 つになる。また、比較例 2 では比較先カテゴリを「かま」とした場合で、一致要素が「圧力がま」となる。この比較を分類体系間の全てのカテゴリに実施する。また、この比較の際に比較元の要素の名前が比較先の要素の名前に含まれている場合は、その比較先の概念の意味を含めることが可能であることから一致要素の一つとする。

2.3.2 一致度と網羅度

分類カテゴリ間の関連付けには、比較した分類カテゴリ間の一致要素数とそれぞれの分類カテゴリの要素数を元とした一致度と網羅度と呼ぶ基準パラメータを導入する。一致度はカテゴリ間の類似関係を表す指標となり、網羅度はカテゴリ間の上位下位関係を表す指標となる。それぞれのパラメータにはしきい値を設定することで、比較した分類カテゴリ間の関係毎に上位下位関係の決定の基準を設定可能とする。

$$\text{一致度} = \frac{\text{比較カテゴリ間の一致要素数}}{\text{比較元のカテゴリの要素数}} \quad (1)$$



比較例1
「鍋・釜類」 vs 「アルミニウム製料理用具」
一致要素
「中華なべ」「片手なべ」「圧力がま」

比較例2
「鍋・釜類」 vs 「かま」
一致要素
「圧力がま」

図 3: 分類体系の比較

$$\text{網羅度} = \frac{\text{比較カテゴリ間の一致要素数}}{\text{比較先のカテゴリの要素数}} \quad (2)$$

2.3.3 分類カテゴリ間の関連の種類

比較する分類カテゴリ間の要素の関係に注目した場合、図 4で示す 3 つのタイプに分けることができる。第 1 は比較元と比較先のカテゴリの要素の数や概念の範囲がほぼ同じくらいの場合、第 2 は比較元に含まれる要素が比較先の複数のカテゴリに含まれるような場合で、この場合比較元カテゴリの要素の数が比較先カテゴリの数に比べて多くなる、第 3 は比較先のカテゴリに含まれる要素が比較先のカテゴリにほぼ含まれ、さらに比較先のカテゴリには他の要素も多く含まれる場合である。これらの比較カテゴリ間の要素の関係と一致度、網羅度の関係から比較元と比較先のカテゴリ間の関連性を類似、上位下位、無関連として関連付けることができる（図 5 参照）。

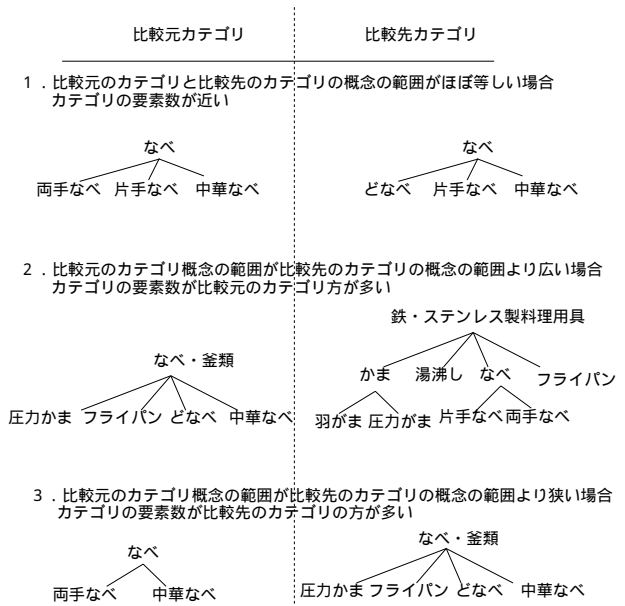


図 4: 比較分類カテゴリ間の要素の関係

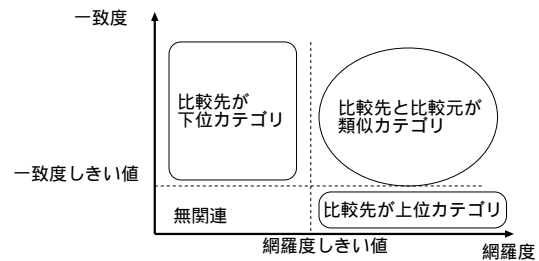


図 5: 一致度、網羅度と比較カテゴリ間の関連の種類

2.4 分類体系間の重ね合わせ

分類体系間の重ね合わせは比較と関連付けで得たカテゴリ間の関連（類似、上位、下位）に基づいて、あるカテゴリの比較先の分類体系での相対位置を決定する（図 6 参照）。重ね合わせの実施により、比較及び関連付けで得た関連性の妥当性を評価し、利用者が目的とする情報の獲得や発見が可能かを判断する為の支援となる。分類体系間の重ね合わせは以下の手順にて行う。

1. 重ね合わせの対象となる類似カテゴリを決定
2. 対象とした類似カテゴリを基準に上位下位カテゴリの発見

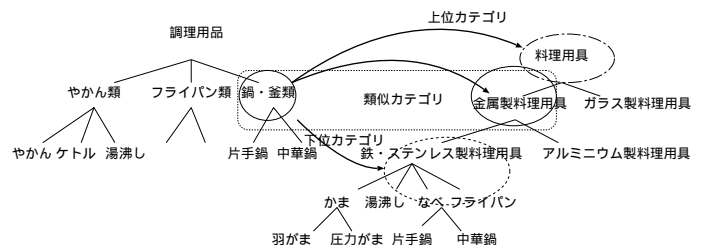


図 6: 分類体系間の重ね合わせ

3 商品分類体系間による比較と重ね合わせ手法の適用評価

分類体系間の比較及び重ね合わせ手法の適用評価に以下の2つの既存の商品分類体系を利用する．比較はそれぞれの中分類を最上位カテゴリ（木構造の根ノード）とした分類体系毎の比較を実施し，それぞれの比較毎の一致度と網羅度を計算する．対象とした中分類とその下位階層に含まれるカテゴリと要素の数を表1に示す．表中の要素数は下位にノードを含まない最終端要素数を示している．

1. JICFS 分類基準

JICFS(Jan Item Code File Service) 分類基準 [1] は JICFS データベースに収録された JAN コード商品情報を効率良く使用できるように設定された JICFS 用の「商品分類コード」で，主として POS データ分析サービスや個々の企業間で行われる電子データ交換，卸売業や小売業への棚割提案等において JICFS 分類コードを「共通の情報交換キー」として利用される．

2. 日本標準商品分類

日本標準商品分類（JSCC）[2] は統計調査の結果を商品別に表示する場合の統計基準として設定されたもので，工業統計調査，生産動態調査等商品別の把握を必要とする諸統計の作成時に利用されるほか，医薬品コード等各企業，業界における商品コードの基準としても利用される．

表 1: 商品分類比較の対象データ

JICFS 中分類	要素数	カテゴリ数
日用雑貨	260	110
医薬品	92	65
化粧品	155	67
家庭用品	449	89

JSCC 中分類	要素数	カテゴリ数
台所用品及び食卓用品	83	22
その他の住居生活用品	94	34
医療品及び関連製品	284	85
医薬品及び関連製品	770	174
化粧品等	163	50
その他の生活用品・文化用品	82	27

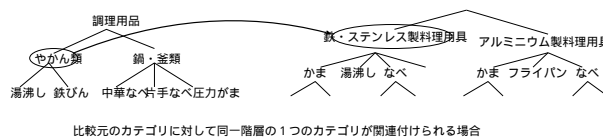


図 7: 比較カテゴリ間の一致要素発見のパターン

4 結果及び考察

商品分類比較結果から，比較カテゴリ間の一致度，網羅度に基づく重ね合わせを実施する．この際に比較元のカテゴリと比較先のカテゴリの間で一致要素が発見されるカテゴリの間に以下のような2つの場合が存在する．

1. 比較先の同一階層の1つのカテゴリと要素の一致が発見される場合（図7上図）
2. 比較先の同一階層の複数のカテゴリと要素の一致が発見される場合（図7下図）

この結果を分類体系の分類体系の木構造の特徴を利用すると以降に述べる木構造上の構造パターンとして示すことができる．重ね合わせではこの構造パターンによって，重ね合わせの基準を決定することが可能となる．そこで，これらの構造パターンと商品分類比較の結果の事例から構造パターンの特徴や要因について考察する．

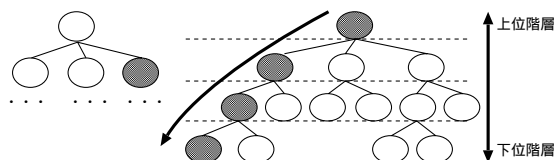


図 8: 「関連付けられた経路」

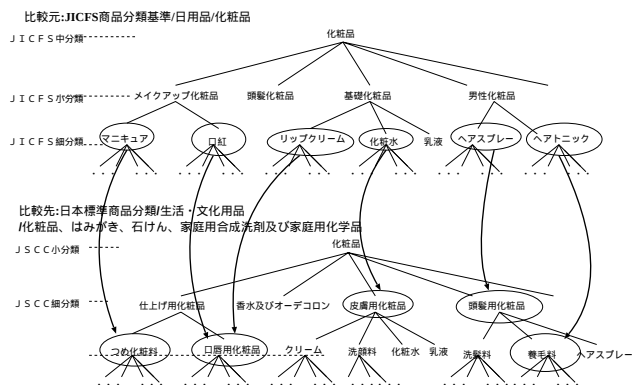


図 9: 化粧品 vs 化粧品他

関連付けられた経路 提案した手法による分類体系間の比較と関連付けでは分類体系に含まれる分類カテゴリ全てを総あたりに関係を調べる．そのため，あるカテゴリの比較結果はその上位カテゴリに反映され，木構造の頂点カテゴリから要素の一致が発見されたカテゴリまでの経路全体に類似，上位，下位等のいずれかの関連付けがされる（図 8 参照）．これを関連付けされた経路と呼ぶ．

単一経路 関連付けられた経路が分類体系の木構造上で分岐せず単一の経路となる場合がある．図 9 の評価サンプルの化粧品 vs 化粧品他を示したものでは，比較元となる JICFS のカテゴリが，例えば「マニキュア」や「リップクリーム」ではそれぞれ JSCC の「つめ化粧品」と「口唇用化粧品」という唯一のカテゴリとの間に要素の一致が発見される，他の JICFS の細分類レベルの場合も多くが比較先の JSCC の唯一のカテゴリとの間で要素の一致が見られる．この場合，関連付けられた経路は「化粧品」からそれぞれのカテゴリまでの単一経路となる．

この構造パターンの要因としては，比較した分類体系間の分類階層や分類の基準が類似していることが考えられる．その結果として，概念の具体的なレベル（例えば JICFS の細分類レベル等の）と同様なカテゴリが比較先の分類体系にも発見される．本構造パターンにて重ね合わせを行う場合は，関連付けられた経路の最下位層のカテゴリを重ね合わせ対象の類似カテゴリとすることができる．

複数経路 関連付けられた経路が複数経路の場合，経路が合流するカテゴリの数に注目してさらにパターンを詳細化する．

合流カテゴリが一つ 複数の経路が合流するカテゴリが一つの場合は，最大一致度となるカテゴリと複数経

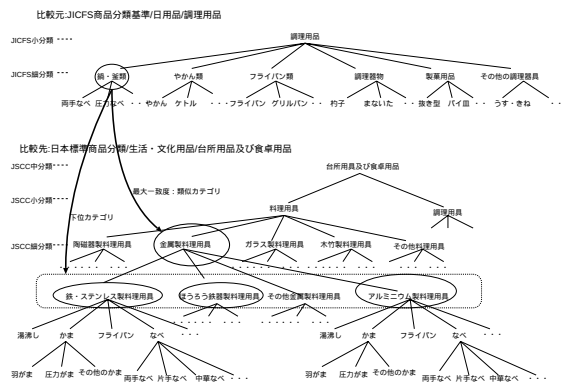


図 10: 調理用具 vs 台所用品及び食卓用品

路が合流するカテゴリが同一となる．この構造パターンの事例として図 10 に評価サンプル調理用具 vs 台所用品及び食卓用品を示す．

ここでの JSCC の分類体系では小分類から細分類，細分類から細々分類と原料を分類基準とした分類を繰り返している．結果として「なべ」等の同一の対象概念をより細分類し，異なる原料のカテゴリに「なべ」等が存在する．このような場合は，比較元の 1 つの要素である「なべ」と複数のカテゴリに細分類された「なべ」が一致要素となる．

この構造パターンの要因としては，分類階層差の大きい体系間の比較をした場合で，特に同一の分類基準を繰り返し対象概念を細分類したカテゴリを比較した場合等に多くあらわれる．

この構造パターンにおける重ね合わせを実施する場合は，合流カテゴリにおいて一致度の値が最大となるため，合流カテゴリを重ね合わせの対象となる類似カテゴリとすることができる．図 10 では比較元として JICFS 分類の「鍋・釜類」を JSCC の「金属製調理用具」に重ね合わせて，その結果として「鉄・ステンレス製調理用具」や「アルミニウム製調理用具」が下位カテゴリとなることを示している．この結果，JICFS 分類基準における「鍋・釜類」に含まれる「両手なべ」は，JSCC の「金属製調理用具」「アルミニウム製調理用具」等に含まれるそれぞれの「両手なべ」を含むことが推察される．

合流カテゴリが複数 複数の経路が合流するカテゴリが複数の場合がある．このパターンは比較元のカテゴリが，分類体系の上位階層レベル（JICFS 小分類レベル等）の場合に多く現れる．この場合いずれの経路も重ね合わせの候補となる．そこで，それぞれの経路が分岐する部分で分割し，分割した経路毎の重ね合わせを実施する（図 11 参照）．このような経路毎に重ね合わせを実施することで，それぞれの分類体系の特徴を

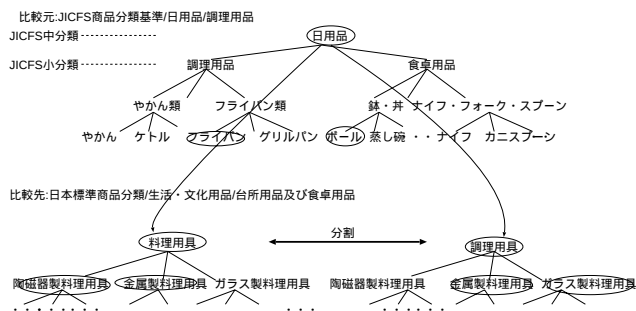


図 11: 分割された経路

考慮することが可能になる。この複数の重ね合わせの可能性の提示により、利用者の目的に応じた情報の選択が可能になると考えられる。

結果、考察のまとめ 以上で示した構造パターンとその要因の分析に基づき、分類体系間の重ね合わせを実施すると、重ね合わせ結果は人の判断において妥当であることが確認できた。この結果から、同一分野の異なる分類体系間の関連性を分類体系のカテゴリと要素に注目した比較及び重ね合わせの方法にて推測可能であることを示した。また、複数の重ね合わせの候補が提示されることで、利用者が自らの判断において選択的な情報の獲得が可能になることが確認された。提案した比較及び関連付け手法と構造パターンの発見に基づく重ね合わせのルール化により、概念体系間の関連性の発見的獲得が可能になると考えられる。

また、これらの分類体系の構造的特徴の結果と合わせて、それぞれの分類体系固有の作成意図や利用目的、分類の対象分野の特性といった要因がこれらの構造に大きく影響を与えていると考えられる。このことは、その対象分野における情報化やモデル化の標準化等への貢献が期待できる。概念体系を扱う上で構造的な特性とともに、意味的特性も重要な要素となると考えられる。

5 関連研究

本研究で示した概念体系はオントロジーの一つとして位置付けることができる。オントロジーは「概念化の明示的記述」等と定義され、対象とする世界の概念とその関係を体系化したもので、その理論や応用が研究されている [3]。オントロジーの問題において、複数のオントロジーを扱う研究の基盤やオントロジーリポジトリの構築を目指して、既存のオントロジーの収集、整理、提供の環境を整備することが提案されている [4]。提案では以下のような方針が挙げられている。

1. オントロジーの収集

形式化の度合や詳細度によって排除したり変形し

たりせず、形式化や粒度の異なるオントロジーを広く収集する。

2. オントロジーの整理

標準化のような一意的な構造ではなく、多様な視点に対して、その視点に則した関係を提示できるような機構の構築。

3. オントロジーの供給

収集されたオントロジーと整理統合した機構によるオントロジーの供給。

本研究では概念体系の一つである分類体系の比較と重ね合わせの手法の提案により、収集されたオントロジーの関連付けや計算可能性の一つを示すことができると考えられる。

6 おわりに

複数概念体系間の関連性発見のための比較及び重ね合わせの手法とその適用評価について報告した。本稿のまとめを述べる。

複数分類体系間の関連性発見のための比較と重ね合わせの手法を提案した。提案手法では比較する分類カテゴリ間の要素の一致数と要素数に着目した一致度と網羅度という指標を導入し、計算的に分類カテゴリ間の関連付けを行う手法を提案した。

また、既存の商品分類体系を利用して、提案した比較と重ね合わせの手法の評価及び分類体系間の比較を実践した。結果から提案した手法が、ほぼ妥当なものであることを確認した。また、分類体系間の比較において、いくつかの構造パターンを発見しその構造パターンに基づく体系間の関連性や要因について検証した。この構造パターンの発見によって重ね合わせの基準等を設定することが可能となり有効な情報となった。

謝辞 本研究において、データの使用を認めていただいた（財）流通システム開発センターに感謝致します。

参考文献

- [1] （財）流通システム開発センター. JICFS 商品分類基準, 1999.
- [2] 総務庁統計局統計基準部（編）. 日本標準商品分類. 財団法人全国統計協会連合, 1990.
- [3] JIPDEC. 大規模知識ベースに関する調査研究 - オントロジー工学に関する調査研究 -. 報告書 財団法人 日本情報処理開発協会, 1998.
- [4] 武田英明. オントロジーを集めよう-world modeling project-. 人工知能学会全国大会（第 13 回）論文集, pp. 175-178, 1999.