

創造的設計支援に用いる仮説的設計知識生成のための 自動文書タグ付け手法の提案

Proposal of Automatic Document Tagging Method for Hypothetical Knowledge Generation
in Creative Design Support System

吉岡 真治*¹ 佐藤 孝彦*¹ 森本 憲悟*² 武田 英明*³ 下村 芳樹*⁴
Masaharu YOSHIOKA Takahiko Satoh Kengo MORIMOTO Hideaki TAKEDA Yoshiki SHIMOMURA

*¹北海道大学大学院情報科学研究科

Graduate School of Information Science and Technology, Hokkaido University

*²東京大学人工物工学研究センター

Research Into Artifacts, Center for Engineering, The University of Tokyo

*³国立情報学研究所

National Institute of Informatics

*⁴首都大学東京

Dept. of System Design, Tokyo Metropolitan University

In order to design new artifacts that are different from existing ones, it is crucial to have new knowledge for the design problem. We have already confirmed that combination of existing knowledge from different domain may be useful for this process by analyzing design cases. In this paper, we briefly review our hypothetical knowledge generation support environment “Universal Abduction Studio” (UAS) and propose an automatic method to organize design document for this environment.

1. はじめに

本研究では、既存の製品の部分的改良などではない、新しい設計解を生成するような設計を「創造的設計」と呼ぶ。このような設計を支援するためには、既存の問題領域固有の知識のみを利用するだけでは困難であり、それまでに、その領域ではあまり考えられていなかった新しい知識を生成し、利用することが不可欠である。

この問題に対し、我々は、異なる領域間の関係に類推に基づく対応関係を見出すことにより、問題領域で利用可能な新たな仮説的知識を生成し、創造的設計を支援する Universal Abduction Studio (UAS) の構想 [下村 06] を提案し、プロトタイプシステムによる基本的動作の検証を行っている [森本 05]。

しかし、本プロトタイプシステムでは、既存の文書に対し、文章の意味的解析 (深層格 [国語研 97] による係り解析、オントロジー情報の付加など) を行った結果を手作業でタグとして付加する必要があった。そこで、本研究では、この UAS のための設計文書の自動タグ付け手法を提案すると共に、対象とする知識を増やした場合に発生した問題点について考察する。

2. Universal Abduction Studio (UAS)

2.1 問題構造の類似性に基づく仮説的知識生成

問題構造の類似性に基づく仮説的知識生成とは、知識を適用する状況が似ているという事から、知識間の関係を導きだし、仮説的知識を生成する方法である。

この方法では、異なる領域において用いられる設計知識と、その設計が対象とする問題構造を表す典型的な設計対象モデルが対になって保存されていることを仮定する。この時、設計者が持っている問題を表現した設計対象モデルを用いて、異なる領域の設計知識を適用するための設計対象モデルから類似した

部分を取り出すことにより、異なる領域知識に基づく概念間の関係を見つけ、知識の変換に役立てる。

例えば、図 1 は問題構造の類似性を示した例である。この例では、クレーンの設計において支柱の軽量化を考える場合に、橋の設計における橋脚の軽量化に利用できる知識を利用できる可能性があり、クレーンにおけるアームの役割を、橋における橋梁が果たす関係にあるということが、問題構造の類似性から得られる。この概念間の対応関係に基づき、知識の変換を行うことにより、異なる領域の知識を対応づける方法である。

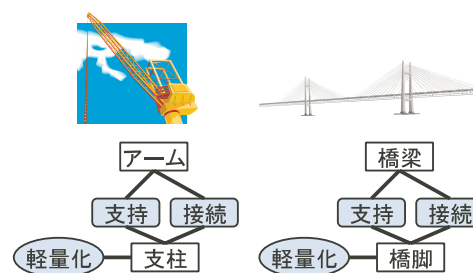


図 1: 異なる問題領域間の類似性の発見

2.2 UAS プロトタイプシステム

この問題構造の類似性を利用して、知識間の関係を導き新たな知識を生成し利用する UAS のプロトタイプシステムを作成した [下村 06, 森本 05]。本システムでは、設計に関する知識を記述した文書に対し、その意味的關係をタグの形で付与した設計文書を対象として文書間の類似性判定を行い、新しい仮説的知識を生成する。

本システムで扱う対象とする知識は、「～ならば～する」といったような、if-then 形式で記述された文書を対象としたタグ付けを行っている。具体的には、文書に現れた概念ならびにその係り受け関係に基づいて、以下の目的に応じた 3 種類の情報が付与している。

A: 吉岡真治, 北海道大学大学院情報科学研究科, 札幌市北区北 14 条西 9 丁目, 011-706-7107, yoshioka@ist.hokudai.ac.jp

オントロジーに基づく情報：異なる領域の概念対の対応関係をとるための基本となる情報として利用

- EDR 電子化辞書の概念体系 [EDR 01] に基づく情報
- 対象が持つ属性やそのデフォルト値に関する情報

概念間の関係を表す情報：複数の概念対間の対応関係をとる際に、意味のある対応関係のみに限定するための制約条件として利用

- 動詞と名詞の係り関係を表す深層格 [国語研 97] の情報
- 名詞と属性の関係を表す対象-属性関係

視点に関する情報：概念間の対応関係をとる際に、その知識と本質的に関係のない属性が利用されたりしないように、注目する属性の組み合わせを視点として定義し利用

また、本システムでは、このタグ付けされた文書と、同じ形式で表現された設計課題の類推による対応関係を求めることにより、仮説的知識を生成する。この時、複数の知識に対して対応関係がつけられるため、対応する概念の数ならびにオントロジーに基づく概念の類似度を利用したスコアを計算し、そのスコアの高い順にユーザに提示する。詳細については、[森本 05] を参照されたい。

2.3 UAS プロトタイプシステムの評価

本プロトタイプシステムの有効性を検証するために、設計ノウハウ集 [渡辺 88] から抜粋した 30 の文書に手動でタグ付け作業を行った知識を利用した実験を行った。この知識に対し、「カメラのフィルムの駆動伝達機構」ならびに「気体を密封する容器」に関する問題に関する問題記述に関しても同様のタグ付けを行い、これらの問題解決に役立つ知識を生成できるかどうかを検証し本システムが有効な仮説的知識を生成可能であることを確認した。

3. UAS のための設計文書の自動タグ付け手法

前節で紹介した UAS のプロトタイプシステムの実験においては、システムの実現可能性を検証するために、少数の既存の設計文書に対し、手作業でタグ付けした情報に基づいて実験を行った。しかし、多種多量の知識を生成し利用するためには、手作業で、全ての文書にタグ付けをしていく手間が大きいため、計算機上に蓄積された文書情報から自動的に本システムで利用可能な形式のタグ付け支援をすることが望まれる。よって、本研究では、UAS のための設計文書の自動タグ付け手法を提案する。

3.1 自動タグ付けシステムの基本方針

プロトタイプシステムで行っている作業の中で「視点に関する情報」の付与という作業は、その現象を支配している物理法則などを理解する必要があり、その自動化が困難であると考えられる。よって、本研究では、それ以外の「オントロジーに基づく情報」と「概念間の関係を表す情報」についての情報を自動的に付加する方法を提案する。ただし、オントロジー情報のうち、「対象が持つ属性やそのデフォルト値に関する情報」については、オントロジー構築が不十分であるため、現時点では、自動化作業の対象としていない。

具体的には、構文解析機として Cabocha [工藤 02] を利用するとともに、日本語語彙大系 [日本語語彙大系 99] を利用して、

動詞と名詞間の係り関係に関する文法情報を付加し、UAS のための設計文書情報を XML 形式で出力する (図 2)。

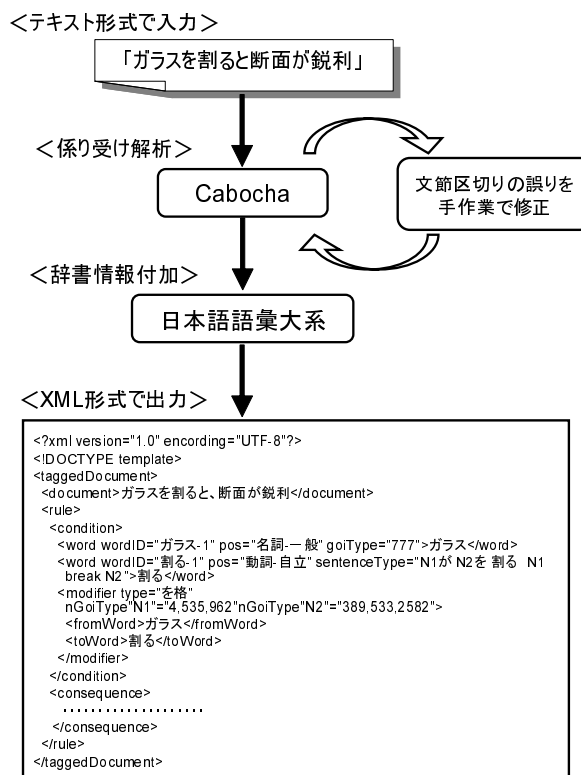


図 2: 自動タグ付けシステムの作業の流れ

各々の手続きの詳細については、以降の節で述べる。

3.2 設計文書からの if-then 型知識の抽出

設計に関連する知識には、様々なタイプが存在する。例えば、物理法則などに基づく因果関係の知識、対象の属性に関する知識、領域固有のノウハウと呼ばれる知識などが存在する。ここで、物理法則などの高度に抽象化された知識については、UAS で提案する方法などを使わなくとも、様々な領域で利用可能であるため、本研究では、領域固有のノウハウと呼ばれるような知識の再利用を主な目標とする。

この、ノウハウと呼ばれるような知識の中にも様々なタイプの知識があり、「～すべき」「～するべからず」といったタイプの良い設計、悪い設計に関する知識や、「～ならば～する」といった因果関係の知識、問題解決の手順を述べたような知識などが存在する。

本研究では、その第一段階として、「ある状況や行為を仮定すると、その結果としてある状況や行為が得られる」という因果関係を表現する知識を「if-then 型知識」と呼び、このタイプの知識のみを取り扱うこととした。この際に、「～すべき」「～するべからず」といった知識に付いては、「～すべき」については、「～するなら良い設計」、「～するべからず」については、「～するなら悪い設計」という因果関係を表現した知識と捉えて同じ枠組みで扱うこととした。

実際の文書中から if 部あるいは then 部に相当する項目を抽出し、タグ付けを行うために、本研究では、[渡辺 88] を参考に、10 種類のパターンを作成し、以降の実験で利用した。

3.3 構文解析結果の修正

Cabocha による構文解析の結果が大きく間違っている場合には、作成した知識が役に立たないものになる可能性が非常に高

い。よって、構文解析の結果を修正することは重要であると考えられる。一方、全ての構文解析の結果をチェックして、正しい構文解析結果を作成することは、タグ付け作業の自動化という観点からは、負荷の大きな作業と考えることができる。よって、以下の基準を満たすものが存在した場合のみ、構文解析結果の修正を人手で行うこととした。

1. 未知語の取り扱い

構文解析の結果として、未知語が存在すると判定された場合には、手作業で品詞情報・形態素区切りを修正する。間違えた分割であっても、未知語を含まない形で分割されたものについては、チェックを行わない。

2. サ変名詞の取り扱い

サ変名詞のあとに「～する」が後続している場合には、動作が主体と考え、サ変名詞として取り扱うが、それ以外の場合は、「普通名詞」として扱う。

3.4 日本語語彙大系に基づく辞書情報の付加

このようにして得られた動詞と名詞の係り関係の情報を元に、UASのプロトタイプシステムで用いているのと同等の、問題構造を表すグラフの作成が可能となる。しかし、単純に動詞と名詞というノードの間に係り関係というラベルなしのリンクを張るグラフとして問題構造を表現すると、「木で橋を作る」と「鉄で建物を作る」という問題記述の間の類似性を考える場合に、「木 鉄」「橋 建物」という対応関係以外に、「木 建物」「橋 鉄」といった対応関係に付いても検討した方が良いことになり、あまり意味のないと考えられる多くの知識候補を生成するという問題が発生する。

この問題を解決するために、プロトタイプシステムでは、主に機械翻訳や自然言語処理などの分野で研究されてきた「深層格」[国語研 97] を利用したタグ付けを行っていた。この「深層格」によるタグ付けは、正確にタグ付けができると、非常に有用な情報となるが、その体系が曖昧なため、深層格を一意に認定することが困難な場合がある。例えば、「フライパンに入れる」という文において、「フライパン」は「入れる」の「場所格」と認定することもできれば「道具格」と認定することも可能である。よって、プロトタイプシステムでは、このタグ付けの曖昧性を複数の格を同時に付けるということで問題解決をはかろうとしていた。

しかし、このような作業を正確に行うためには、動詞と名詞の組み合わせに関する一貫性の高い大規模な辞書が必要であり、係り受け関係の判別の自動化は容易ではない。

これに対し、本研究では、結合価文法に基づく係り受け関係の判別を提案する。結合価文法は、日本語語彙大系[日本語語彙大系 99] で用いられている文法であり、「深層格」の存在は仮定せず、「表層」で述語と任意の格の結合関係を定義する。この文法では、図3に示すような動詞の文系パターンの情報と名詞の意味カテゴリーの情報を組み合わせた辞書情報に基づき動詞と名詞の係りに意味的な制約を加える。

割る		割る	
見出し	: 割る	見出し	: 割る
文型パターン	: N1 が N2 を割る N1 break N2	文型パターン	: N1 が N2 を割る N1 drop below N2
カテゴリ(N1)	: 4(人),535,962	カテゴリ(N1)	: *
カテゴリ(N2)	: 389,533(具体物)	カテゴリ(N2)	: 2585(数量)

図 3: 結合価文法における辞書情報

ここで、例えば、「ガラスを割る」という例文を対象にする場合には、ガラスの意味カテゴリーを調べ(777(ガラス))、そ

の上位概念(539(具体物))を文型パターンの制約に含む図3の左側の割るの意味で用いられていることが判定でき、名詞と動詞の関係としては、389(施設)ないしは539(具体物)と関係付けられる「を格」というラベルを付けることができる。

よって、UASにおけるラベルづけという観点から、この結合価文法を分析すると、格文法を用いていたシステムとほぼ同等なことが実現できると考えられる。

さらに、格文法に基づく辞書に比べ一貫性の高い大規模な辞書が、既に利用可能であることなどを考慮し、本研究では、結合価文法によるラベルづけを行うこととした。ただし、日本語語彙大系は、辞書のエントリーが30万語と豊富ではあるものの、専門用語などにもれも多く、対応する辞書のエントリーが存在しない場合がある。日本語では、複数の形態素から構成される名詞の場合には、その主辞(多くの場合、一番最後の名詞)が意味カテゴリーの分類に役立つことが多いため、この主辞の意味カテゴリーを付与することとした。

また、日本語では主語や目的語の省略などがしばしば行われることを考慮した係り受けの語の補完を行なった。

3.5 XML形式による出力

if部に相当する部分と、then部に相当する部分の各々に付いて、構文解析の結果ならびに辞書情報を付加したタグ付けを行う。UASでは、問題の種類に応じて、if部とマッチングを行うのか、then部とマッチングを行うのかの制御を行う。

4. 自動タグ付け実験と仮説的知識の生成

本自動タグ付けシステムの有用性を検証するために、以下の2つの項目を検証するための実験を行った。

1. 自動文書タグ付けシステムに必要な手作業の量と、生成した情報の精度と再現率の検証
2. 作成されたタグ付け情報を用いた仮説的知識の生成可能性の検証

4.1 実験1 - 自動文書タグ付けシステムの性能評価

「続・機械設計 心得ノート」[渡辺 88] から、if-then知識と考えられる270文・346個の知識を抽出し、入力テキストとし、自動文書タグ付けシステムによって作成された知識が必要な情報を保持できているか検証した。

実際に行った手作業の量、生成した情報の精度と再現率に関する情報は下記の通りである。

- 文節区切りの修正割合
文節数 3594 に対して、265 文節に手作業で修正を加え、 $\frac{265}{3594} \approx 7.37\%$ となった。
- 修飾関係の誤解析割合
上と同様、文節数 3594 に対して、121 文節で誤解析があり、 $\frac{121}{3594} \approx 3.37\%$ となった。
- 述語に直接係る名詞(サ変名詞は述語とする)の辞書情報付加に関する再現率と精度
語彙数 1163 に対して、1053 個の語彙に辞書情報を付加し、再現率は $\frac{1053}{1163} \approx 90.5\%$ となった。
また、その精度は、 $\frac{952}{1053} \approx 90.4\%$ となった。
複合語に対する処理を行うことにより、専門用語を多く含む文書に対しても、ある程度の再現率を確保することができた。誤りは、主に、専門用語が不適切に分割されたために、元の意味が復元できない場合と、専門用語が

一般用語と異なる意味カテゴリーで利用される場合がほとんどであった。

- 述語の辞書情報付加に関する再現率と精度
語彙数 849 に対して、411 個の語彙に辞書情報を付加し、再現率は $\frac{411}{849} \approx 48.4\%$ となった。また、その精度は $\frac{345}{411} \approx 83.9\%$ となった。
動詞にかかっている名詞の意味カテゴリーの分類が間違っている場合やできる限り多くの文型パターンの情報をとろうとするために行った情報の補充が間違った場合などに間違った文型パターンを設定してしまった。

全体の結果を総評すると、述語の辞書情報付加割合があまり高くない。一方で、名詞の意味カテゴリーはよって、このような述語を対象にする場合には、表層の助詞の情報と名詞の意味カテゴリーを利用したマッチングを行う必要があると考えられる。

4.2 実験 2 - 知識統合手法の評価

作成した知識の有用性を検討するために、UAS のプロトタイプシステムを改良し、結合価文法を考慮した知識間の対応関係を生成するシステムを作成し、その結果を検証する。以前のシステムでは、深層格が同じ時に対応関係を生成していたが、改良したシステムでは、助詞によって決まる表層格(「が格」、「を格」など: 文型パターンにより、他の助詞と置き換えが可能)が一致し、その格による動詞と名詞の意味的制約が一致する度合いが高いほど、スコアを高くした。また、動詞・名詞については、以前のプロトタイプと同様に EDR[EDR 01] の概念辞書の概念階層の類似性に基づいたスコアリングを行った。このスコアリングに関して生じた問題は、次節の考察で述べる。

本実験では、本システムと自動タグ付けされた知識を用いて、UAS のプロトタイプシステムで導出された仮説的知識が導出できるかどうかを、「カメラのフィルムに駆動を伝達し、かつ、撮影時には正確な位置にフィルムを停止させることが可能な機構」という問題を例に確認をした。

今回入力した設計課題の出力結果は図 4 のようになった。ここで導出された仮説的設計知識は、「スプロケットでフィルム [チェーン] に大きな駆動 [駆動] を伝達する [伝達する] と、未駆動 [駆動] 時にはスプロケットでフィルム [チェーン] を停止させることができる」([] 内は、元々の知識に存在した語)という以前の UAS のプロトタイプで得られた知識を上位の仮説的設計知識の候補として導出することが可能であった。

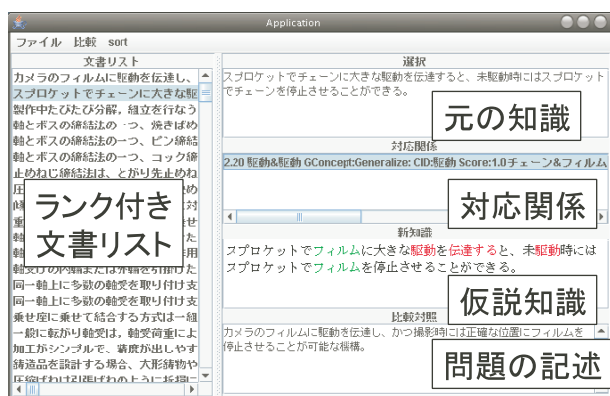


図 4: UAS システムの出力結果

4.3 実験 2 の考察

実験 2 において、実際に多くの知識にタグ付けを行い、プロトタイプシステムと同様のスコアリングを行って、有用な知識を発見しようと試みたところ、あまり適切でないと思われる知識が多く生成されてしまった。これは、プロトタイプシステムにおいて、動詞と名詞について同じように類似度を計算していることに起因していると考えた。

設計問題について考えると、名詞には、その利用目的に応じた多様な概念分類が可能であるが、一般的な辞書である EDR や日本語語彙大系では、このような詳細な概念分類が行われていないことが想定される。また、専門用語のように、辞書に存在しない用語があることを考慮すると、多少、類似度が低くても、仮説知識の候補として生成した方が良い物が存在する。これに対し、動詞は、そもそも数が名詞に比べて少なく、専門用語になった場合でも、主辞は一般的な動詞である複合動詞である場合が多く、辞書に載っていないという場合が少なくない。また、名詞に比べると、中間階層の概念が、「ものを対象とする行為」といった非常に抽象的なものが多く、一定階層以上の抽象化により、あまり適切でない対応関係が増える傾向にある。

このような分析結果に基づき、本システムでは、名詞に関しては、抽象化の段数の制限を設けないが、動詞に関しては、2 階層までの抽象化しか許さないという制約を設けることにより、上記の問題に対応した。

5. まとめ

本論文では、設計者にとって役立つ異なる領域の関連知識を発見し、既存の対象領域知識と組み合わせることにより、新たな仮説的知識を生成する方法である Universal Abduction Studio (UAS) の構想とプロトタイプシステム [下村 06] について概説するとともに、UAS における知識のタグ付けの自動化手法について述べた。また、実験により、本タグ付け手法の有用性を確認した。

今後は、実際に作成された知識の有用性の分析や、その有用性の分析に基づく、生成した仮説的知識のランキング手法などについて検討する必要があると考えている。

参考文献

- [国語研 97] 日本語における表層格と深層格の対応関係 - 国立国語研究所報告 - , 三省堂 (1997).
- [日本語語彙大系 99] NTT コミュニケーション科学基礎研究所 (監修): 日本語語彙大系 CD-ROM 版, 岩波書店 (1999).
- [下村 06] 下村, 吉岡, 武田, 富山: アブダクションに基づく設計者支援環境の基本構想, 日本機械学会論文集 C 編, Vol. 72, No. 713, pp. 274-281 (2006).
- [工藤 02] 工藤, 松本: チャンキングの段階適用による日本語係り受け解析, 情報処理学会論文誌, Vol. 43, No. 6, pp. 1834-1842 (2002).
- [森本 05] 森本, 鬼城, 藤本, 下村, 吉岡, 武田: アナロジーを用いた仮説的設計知識の生成手法-概念ネットワークからのモデルの生成と利用-, 2005 年度人工知能学会全国大会講演論文集, 人工知能学会 (2005), 3F1-04 (CD-ROM).
- [渡辺 88] 渡辺秀則: 続・機械設計心得ノート - ベテラン設計者のノウハウ集 -, 日刊工業新聞社 (1988).
- [EDR 01] 日本電子化辞書研究所: EDR 電子化辞書 (第 2 版) 仕様説明書, TR2-006(改) (2001).
- [武田 04] 武田, 藤本, 吉岡, 下村, 富山: 設計情報の知的処理のためのタグ付け方法の検討, 2004 年度人工知能学会全国大会講演論文集, 人工知能学会 (2004), 3H3-02 (CD-ROM).