

DODDLE: 計算機可読型辞書を利用した 領域オントロジー構築支援環境 (1) — 概念階層構築支援 —

3 A F - 8

小森 聡* 関内 律恵子* 榎松 理樹** 青木 千鶴* 山口 高平*

* 静岡大学

** 岩手県総務部県立大学整備室

1 はじめに

近年、人工知能の各分野で、オントロジー工学の研究への関心が高まっているが、特に、自然言語理解と知識工学において、その関心は大きい。即ち、自然言語理解の分野では、言語処理のタスクの観点から、多数の概念を含む計算機可読型辞書 (MRD: a Machine Readable Dictionary) の開発が進められ、今までに、WordNet (プリンストン大学) や EDR (日本電子化辞書研究所) などの MRD が開発されている。

一方、知識工学の分野では、知識システムの部品ライブラリーの観点から整備が進められ、特に、推論 (問題解決) エンジンを構築するためのライブラリーとしてのタスクオントロジー、知識ベースを構築するためのライブラリーとしての領域オントロジーの研究開発が進みつつある [1]。

筆者らは、独立に進められてきたオントロジー工学の統合を計る事を目的として研究を進めており、本稿では、比較的構築が困難とされている領域オントロジーを構築するために、MRD の活用法について検討する。以下においては、MRD を利用した概念階層構築支援について報告し、文献 [2] においては、MRD を利用した概念定義構築支援について報告する。

2 MRD と領域オントロジー

MRD で提供されている概念階層構造は、多数の概念をカバーしており、領域 (オントロジー) の概念階層構造を構築するための基盤として利用可能である。しかしながら、これは言語処理の観点から規定されたものであり、特定の問題領域における概念階層構造と比較すれば、種々の差異があると推定される。この差異は、タスクやコンテキストに依存して概念の意味が変化する Concept Drift の一種であり、それを管理する機構が課題となる。

以上の事から、MRD を利用して領域概念階層構造を構築するには、MRD から領域概念に関連する情報を抽出し、その関連情報を分析して、Concept Drift が発生している箇所を同定して、ユーザが修正する事を支援する計算機環境が望まれる。以下、本支援方法に基づいて開発された領域オントロジー構築支援環境 DODDLE (a Domain Ontology rapid DeveLopment Environment) について述べる。

DODDLE: A Domain Ontology rapid DeveLopment Environment Using Machine Readable Dictionaries(1)
— Conceptual Hierarchy Support —
Satoshi Komori, Shizuoka University

3 基本設計

図 1 にシステムフローを示す。MRD としては WordNet [3] を用いる。DODDLE への入力、構築したい概念階層のノードとなる概念 (基本的単語) のリストである。DODDLE からの出力は、その問題領域から考えて妥当な概念階層である。

概念階層構築支援は、1. WordNet からの関連情報の抽出と 2. 領域固有性を考慮した修正の 2 つの工程に分かれる。

抽出工程では、まず、入力語彙と WordNet の見出し句のスペルマッチを行い、スペルマッチ階層を作る。複数の概念 (意味) を持つ語句については、ユーザーとインタラクションをとりながら、その問題領域から考えて妥当な意味を持つと思われる概念 (ベストマッチ) を決定する。決定された各ベストマッチに対して WordNet のルートまでの概念階層の抽出を行う。これをベストマッチ階層とよぶ。次に、ベストマッチ階層における、各ベストマッチの位置関係を保持するために必要な中間概念 (SIN: Salient Internal Nodes) 以外の中間概念を取り除く。SIN とは、その下位にあるベストマッチノードと他の SIN を区別するための中間ノードである。以上の操作で求めたベストマッチ、SIN、ルートノードのみから成る概念階層を初期モデルと呼ぶ。ここまでが抽出工程である。

次に、修正工程では、初期モデルにおいて再利用出来る領域と、Concept Drift が発生していると推定されるため修正が必要な領域に分割する必要がある。

ベストマッチは、問題領域から考えてほぼ妥当と考えられた概念であるので、それらが連続するパスは、妥当な概念が集中している為、そのまま再利用できるパスとみなせる。このパスを PAB (Paths including only Bestmatches) と呼ぶ。一方、ベストマッチ以外の概念が含まれる領域は、その概念により概念構造の差異 (Concept Drift) が生じている可能性があるため、移動すべき領域とみなせる。この領域を STM (SubTrees manually Moved) とよぶ。

ユーザー (専門家) は STM を移動する事で初期モデルから領域概念階層を構築する。ここで、STM の移動先についてはユーザが決定し、移動する必要がないと判断した場合は移動しない。最終的に、STM の移動のみでは対処できない箇所についてはノード単位でユーザが修正を行い、領域概念階層を構築する。さらに、スペルマッチに失敗した語句がある場合にはそれらもユーザが付加する。

現在、DODDLE は Perl と Tcl-Tk を用いて、EWS の Unix プラットフォーム上、およびノートパソコンの Linux プラットフォーム上で実装されている。

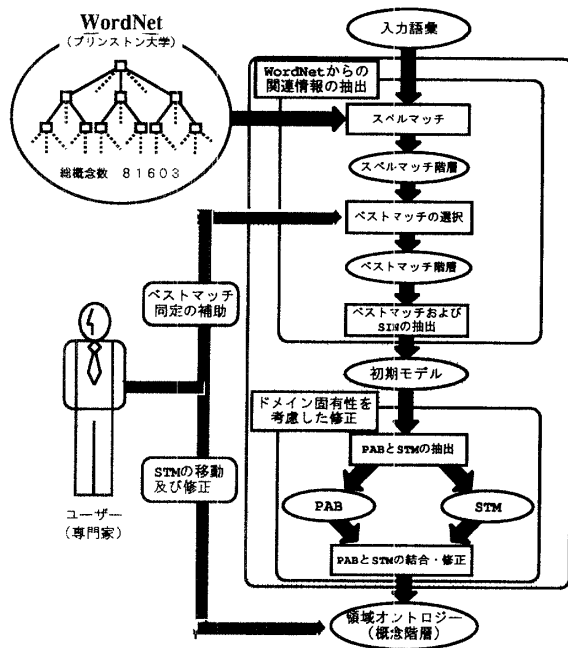


図 1: システムフロー

4 実験と評価

今回は、輸送問題、法律、ソフトウェアプロセス (SP) という3つの領域について DODDLE を適用する実験を行った。入力語彙としては、輸送問題では、輸送問題で利用される12語句、法律では、国際売買法 (以下、CISG と略記) 第2部 [4] に含まれる法律用語群から抽出した46語句、SPは [5] で利用されているリソース体系から抽出した21語句を与えた。

表1に実験の結果を示す。ここで支援率とは、最終的に得られた階層構造のうち、PAB、STMによる階層構築支援がされたと考えられる割合であり、表1で見ると、 $(1 - \frac{X+Y}{A+B}) * 100(\%)$ で表される。

表1より、輸送と法律の問題領域においては、PABの抽出とSTMの移動により、60%以上の支援率を達成する事ができた。一方、領域固有性が他の問題領域よりも高いと考えられるSPにおいては、ルートノードが変更され (リソースがルートノードとなった)、支援率はかなり低くなったが、その変更後の支援率を計算すれば、表1に示すように80%の高い支援率が達成できた。

しかしながら、今回の実験では、Concept Driftによって、WordNetの概念階層では下位部分にある概念が、領域概念階層の上位概念として位置付けられたり、WordNetの概念階層よりもさらに詳細な分類がなされたりする場合があった。このような場合、STM単位の移動だけでなく、ノード単位での移動が必要となり、今後これらの支援方法について検討していきたいと考えている。

5 おわりに

MRDを利用した領域オントロジー概念階層構築支援については、USC/ISIで研究開発されている Onto-

表 1: DODDLE での評価実験結果

領域名	輸送	法律	SP
入力語句数	12	46	21
ベストマッチのノード数	12	44	15
獲得されたPAB数 (A)	1	11	0
獲得されたSTM数 (B)	5	14	10
ノード単位で変更したPAB数 (X)	0	3	0
ノード単位で変更したSTM数 (Y)	2	3	2
支援率 (%)	67	76	80

saurus[6]があるが、Ontosaurusでは、入力単語と SENSUS というMRD間でスペルマッチを実行し、その照合結果に関連する情報を提示するレベルの支援に留まっており、照合結果を分析して、再利用可能領域とそうでない領域に分割して支援する機構までは考慮されていない。

本稿では、MRDとの照合結果を分析して支援する機構を提案して、いくつかの問題領域上で実験・評価した上で、ある程度の支援可能性は示せたと考えている。しかしながら、実験は、資料に基づいて行ったものであり、実際に、専門家に DODDLE を利用してもらった評価とはなっていない。また、問題領域に依存して概念構造が変化する Concept Drift を深く考察して定式化する必要がある、そのような実験と分析を進めながら、DODDLE の支援機構の充実に回りたいと考えている。

謝辞

本研究の一部は、文部省科学研究費補助金重点領域研究「法律エキスパート」(代表:吉野一教授)の補助を受けました。記してここに謝意を表します。

参考文献

- [1] Gertjan van Heijst: "The Role of Ontologies in Knowledge Engineering", Dr.Thesis, University of Amsterdam (1995)
- [2] 関内 律恵子 他: "DODDLE: 計算機可読型辞書を利用した領域オントロジー構築支援環境 (2) - 概念定義構築支援 -", 第55回情報処理学会全国大会, 3AF-7 (1997)
- [3] G.Miller: "WordNet: An on-line lexical database", International Journal of Lexicographer, Vol.3, No.4 (1990)
- [4] 曾野和明, 山手正史: "国際売買法", 青林書院 (1993)
- [5] P.Mi and W.Scacchi: "A Knowledge-Based Environment for Modeling and Simulating Software Engineering Process", IEEE Transaction on Knowledge and Data Engineering, Vol.2, No.3, pp.283-294 (1990)
- [6] B.Swartout, R.Patil, K.Knight and T.Russ: "Toward Distributed Use of Large-Scale Ontologies" KAW96 URL: <http://ksi.cpsc.ucalgary.ca/KAW/KAW96/KAW96Proc.html> (1996)