

G-3 Content-based network におけるオントロジー自動変換方法の提案 Ontology transformation in content-based network

酒井 隆道[†]
Takamichi Sakai

太田 昌克[†]
Masakatsu Ohta

寺元 光生[†]
Mitsuo Teramoto

1. 序論

近年, Gnutella, JXTA 等の P2P ネットワークが着目を浴びているが, それらの特徴のひとつとして, それがアプリケーション層 (overlay) のネットワークであることが挙げられる. この特徴によって, 下位層である IP の束縛から解放され, 自由で柔軟な名前空間やネットワーク構成を実現可能としている.

このようなアプリケーション層のネットワークの一形態として「Content-based network (以後 CBN と略す)」が提案されている [1][2]. これは, メッセージ内に記述されたメタ情報に基づいた内容指向の配送 (routing) を行うネットワークのことである. 従来型ネットワークにおいては, 各ノードに対して一意に IP 等のネットワークアドレスが付与され, メッセージはそのヘッダにおいて指定されたネットワークアドレスを持つノードに対して配送される. これに対して CBN では, 各ノードは自身が受信したいメッセージに関する条件を記述したフィルタ (filter) を CBN に登録し, メッセージはそのヘッダに記述されたメタ情報と合致するフィルタを登録しているノードに対して配送される.

CBN では広域 (global) な環境に適用できることを主目標の一つに掲げており, その実現に際しては二つの大きな課題がある. 一つ目はルーティングである. 広域環境においてスケーラブルなメッセージ配送を実現できるルーティングアルゴリズムの考案が課題となる. そして二つ目が本稿で扱うオントロジーである. ここで, オントロジーとはメタ情報とフィルタの構造 (データモデル) や用いる語彙を規定する仕様のことであり, CBN におけるアドレス体系となる. CBN では, まさにその内容指向な特性によって「内容の理論 (Content-based AI) [3]」であるとされるオントロジー工学の手法が活用できる.

CBN における具体的なオントロジーの課題として, どのようなオントロジーモデルを用いるかという課題とともに, 特に広域環境を想定した場合, 異なるオントロジーを有するネットワークドメイン間でのメッセージ流通を実現するための, メッセージのオントロジー変換を実現する課題がある. なぜならば, 従来型ネットワークにおいては階層的で一意的な標準アドレス体系をグローバルに規定することができたが, オントロジーに関してグローバルな標準体系を確立することは現実的でない.

以上の状況を踏まえ, 本稿では, 第 2 節において CBN におけるオントロジーモデルに関する考察を行い, 続く第 3 節において, 考察したオントロジーモデルを対象としたオントロジー自動変換方法の提案を行う.

2. オントロジーモデル

CBN において, 採用するオントロジーモデルの決定はその全体設計に対して多大な影響を及ぼす. CBN を

利用するサービスアプリケーションの立場からは, より表現力 (expressiveness) が高く柔軟性のあるオントロジーモデルが望ましい. しかし, 例えばルーティングは本質的に分散アルゴリズムであり, 処理効率や実装の容易性の面からは, 表現力が低くともシンプルなオントロジーモデルが望ましい. 一般に処理効率やスケーラビリティとオントロジーモデルの表現力はトレードオフの関係にあると言える.

具体的なオントロジーモデルの例として, channel, subject といった数値や文字列を基本とした比較的表現力の低いものから, フレーム形式のような構造化されたデータモデルを有するもの, オブジェクト指向モデルのような詳細な関係定義を可能としたもの, あるいは一階述語論理式を用いることによって公理定義を可能としたもの等が挙げられる.

CBN の基本機能は, メッセージの配送先を動的に決定するという検索機能にあり, 検索という利用側面からメタ情報が構造化されているメリットは大きい. 一方で, 前述の通り表現力に富んだオントロジーモデルは, どちらかと言えば集中サーバ向けであり, 処理効率やスケーラビリティに問題がある. よって, CBN のオントロジーモデルとしては, 構造化されつつも適度にシンプルなものに適していると考えられる.

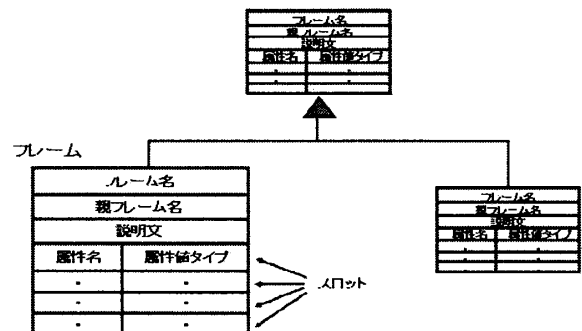


図 1: オントロジーモデル

以上の考察に基づき, 本稿では図 1 に示すフレーム形式を用いたオントロジーモデルを採用する. すなわち, オントロジーはフレームの集合から構成され, フレーム間は継承関係を有している. フレームはフレーム名, 親フレーム名, フレームの説明文 (annotation) と複数のスロット (属性名と属性値タイプのペア) から構成される. そして, フレームのインスタンスがメッセージのメタ情報となる.

なお, オントロジーモデルがこのような階層構造を有していると, 例えば階層的にマルチキャストパス (spanning-tree) を張るといったルーティング方式の実現も容易となる.

[†]NTT 未来ねっと研究所

3. オントロジー自動変換方法

前節で定義したオントロジーモデルに限定しても、なお概念化 (conceptualization) の違いに基づく構造化の仕方や語彙の選択における差異が発生する。本オントロジー自動変換方法は、このような差異を解消することを目的とするものである。以下、その基本となるフレームおよびスロットの対応関係を求めるアルゴリズムをステップ順に説明する。

(1) オントロジーの取得

始めに変換元 (source) オントロジーと、変換先 (target) オントロジーを取得する。

(2) スロットのマッピング

スロットのマッピングは、変換元スロットと変換先スロット (スロット対) の類似度を定義することにより、トータルとしての類似度が最も高くなる組み合わせを求めるといふ「組合せ最適化問題」に帰着できるが、ここではより簡明なアルゴリズムを用いる。

(2.1) スロット対類似度の定義

スロット対類似度を以下の式によって定義する。

$$\text{「スロット対類似度」} = \min(\text{「属性名の類似度」}, \text{「属性値タイプの類似度」}) \quad (1)$$

ここで「属性名の類似度」に関しては既存のシソーラスを用いて算出することも可能だが、本稿では国語辞書と英和辞書から構築した概念ベース [4] を用いて算出する。すなわち、各属性名の概念ベクトルを V_s, V_t とすると、

$$\text{「属性名の類似度」} = \frac{V_s \cdot V_t}{\|V_s\| \|V_t\|} \quad (2)$$

となる。そして「属性値タイプの類似度」に関しては、属性値タイプ (string, integer 等) の階層関係に基づいたシソーラスツリーを構築しておき、以下の式によって算出する。

$$\text{「属性値タイプの類似度」} = \frac{2d_c}{d_s + d_t} \quad (3)$$

ここで、 d_s と d_t は各属性値タイプのツリーにおける深さであり、 d_c は共通上位ノードの深さである。

(2.2) スロット対応の決定

以下 ([1] ~ [2]) のアルゴリズムに従ってスロット対応を決定する。

[1] スロット対の類似度行列 S を作成する。ここで行列要素 S_{ij} は、 i 番目の変換元スロット $slot_i$ と j 番目の変換先スロット $slot_j$ の類似度を式 (1) を用いて算出した値をとる。

[2] スロット対応のペアが作れなくなるまで、以下を繰り返す。

[2.1] S_{ij} が最大値 s_{max} をとるとききの i と j を求める。そして、

```
if (  $s_{max} > \theta(\text{threshold})$  ) {
  スロット対応 ( $slot_i \rightarrow slot_j$ ) を保管する。
} else {
```

ループを終了する。

```
}
[2.2] 上記  $i$  行と  $j$  列を削除する。
```

(3) フレームのマッピング

フレームのマッピングに関しても、基本的にスロットのマッピングと同様の方法を用いる。

(3.1) フレーム対類似度の定義

フレーム対類似度を以下の式によって定義する。

$$\text{「フレーム対類似度」} = \min(\text{「フレーム名の類似度」}, \text{「説明文の類似度」}, \text{「スロット対集合の類似度」}) \quad (4)$$

ここで「フレーム名の類似度」は「属性名の類似度」と同様に概念ベースを用いて算出する。また「説明文の類似度」に関しても、各説明文中の自立語に対応する概念ベクトルの加重平均を計算することにより、式 (2) を用いて算出する。そして「スロット対集合の類似度」は、ステップ (2) において決定されたスロット対集合を用いて、以下の式によって算出する。

$$\text{「スロット対集合の類似度」} = \frac{\sum \text{スロット対類似度}}{\text{スロットの数}} \quad (5)$$

この式は、対応するフレームは同様のスロットを保持しているという考えに基づいている。

(3.2) フレーム対応の決定

フレーム対応の決定はステップ (2.2) のスロット対応の決定と同様である。すなわち式 (4) を用いてフレーム対の類似度行列 T を作成し、同様のアルゴリズムによってフレーム対応を決定する。

4. 結論

本稿では、CBN におけるオントロジーモデルに関する考察を行い、オントロジー自動変換方法の提案を行った。今後、提案したアルゴリズムの評価と改良を行う予定である。

参考文献

- [1] 星合 隆成, 小柳 恵一, ビルゲ スクバートル, 久保田 稔, 柴田 弘, 酒井 隆道: 意味情報ネットワークアーキテクチャ, 電子情報通信学会論文誌 B Vol.J84-B No.3 pp.411-424 (2001 年)。
- [2] Carzaniga, A., Rosenblum, D.S., and Wolf, A.L.: Design and Evaluation of a Wide-Area Event Notification Service, *ACM Transactions on Computer Systems* 19(3):332-383 (2001)。
- [3] 溝口 理一郎: オントロジー研究の基礎と応用, 人工知能学会誌 Vol.14 No.6 (1999 年)
- [4] 笠原 要, 松澤 和光, 石川 勉: 国語辞書を利用した日常語の類似性判別, 情報処理学会論文誌 Vol.38 No.7 (1997 年)
- [5] 伊藤 史朗, 上田 隆也, 池田 裕治: 分散情報源に対する情報エージェントのための事例に基づくフレームマッピング, 電子情報通信学会論文誌 D-I Vol.J81-D-I No.5 pp.433-442 (1998 年)。