

計算機上での文章の意味表現における Bayesian Belief Network による語の共起情報の統合化の試み

中 挟 知延子*

1. はじめに

自然言語処理における、形態素解析・構文（統語）解析は、文の構造を解析するもので、意味的側面にはあまり言及していない。しかし、昨今意味解析がますます重要となっており、文章の意味内容を表現した構造をコンピュータに持たせることは重要な課題である。著者は、コンピュータ上に、自然科学分野の論文文章から意味構造を表現し、これらを参照することにより、ユーザとしての自然科学研究者が、真に得たいデータや文章の適確な要約、意味内容の把握・参照ができる研究支援システムの開発を進めている [Nakabasami 99]。現段階では、金属材料設計分野の論文に焦点をおいている。材料設計研究は未知の物質を設計・開発する分野であり、研究者は研究を行っていく上でさまざまな過去の文献、とりわけ既存の研究論文を参照する。その際、膨大な数にのぼる論文の中から、研究者の要求に応じて論文を検索し、その論文の内容に関して応答できるシステムを用いることで、よりよい研究支援が提供できると考えられる。論文は曖昧性の少ない自然言語であることが要求されるために、自然言語処理の解析対象とされることが多いが、日々生成されその数は増大し、無数に存在する。これに対応していくためには、コンピュータによって自動的に論文の内容を可能な限り厳密に把握し、抽出できるようにすることが必要となる。

文章における語についての意味構造は、語彙意味論 (Lexical Semantics) の分野などにおける研究から提案されているが [Pustejovsky 93, 95]、文に対する意味構造を扱ったものはない。それに対して本研究では、自然言語文章から自動的に生成可能な、統語・意味の両面を取りこんだ文の意味表現として、構文解析された語に、意味役割を考慮した意味タグを付与した文の意味構造データを提案している。論文の意味内容を把握するためには、語のレベルだけではなく、文の意味内容を把握することが重要となる。文の内容把握のためには統語情報 (syntax) と意味情報 (semantics) が不可欠の要素であり、文の記述内容の表現はお互いに補い合って表現できる [Jackendoff 92]。

しかし、文に出現する語に意味タグを付与したものを合わせたデータだけでは情報が十分であるとはいえない。文に共起する語の相互作用から、文の意味内容が変化することも考えられる。そのため本論文では Bayesian Belief Network (BBN) を用いて、文中に共起する語の意味タグの関係を検証、評価する。BBN は、実世界における現象の因果関係を記述し、ある事象の確からしさが変化

*東洋大学国際地域学部；Faculty of Regional Development Studies, Toyo University

した時に、関係する他の現象の確からしさの変化を反映するネットワークである [Jensen 96, Mitchell 97]。BBN を用いて語の共起情報を意味構造データに統合することで、適確な意味表現をシステムに取り込めると期待される。本稿では、そのための BBN に関する検証と評価を行うと共に、従来共起情報の抽出や語の多義性解消に用いられてきた Naive Bayes [Domingos 96] による推定方法との比較を行っている。

以下2章では意味構造データについて述べる。3章では BBN について概観したあと、意味構造データ中の意味タグの推定に関し、BBN に情報エントロピーの考えを取り入れた手法について述べる。4章ではサンプルデータに本手法を適用した実験結果について報告する。5章では実験結果について Naive Bayes の手法を用いた場合と比較し、結果を考察する。最後に6章でむすびとする。

2. 意味構造データ

2.1 概要

意味構造データは (2.1.1) のリスト構造で表現される。

(2.1.1) $((ENTITY)(EVENT)(PROPERTY))$

意味構造データでは、統語情報でのカテゴリだけではとらえられない語の持つ潜在的な文での役割を考え、文を構成する要素を ENTITY, EVENT, PROPERTY という3つの大きなカテゴリに分けている。構文解析された語は付与された意味タグに基いてこれらのカテゴリごとにまとめられる。ENTITY は文において事象 (EVENT) を行う、あるいは事象が起こされる実体である。EVENT は事象そのものである。PROPERTY は ENTITY が持つさまざまな性質、特徴であり、ENTITY をさまざまな側面から記述する。構文解析結果（統語情報）で動詞とされた語は事象を表している限り文の意味表現においては EVENT とし、名詞とされた語でも意味役割により事象を表しているものであれば EVENT に含める。これら3つはさらに (2.1.2) のリスト構造を持つ。

(2.1.2) $(ENTITY) \rightarrow ((tag_1 term_1) \dots (tag_i term_i))$
 $(EVENT) \rightarrow ((tag_1 term_1) \dots (tag_j term_j))$
 $(PROPERTY) \rightarrow ((tag_1 term_1) \dots (tag_k term_k))$

上記のリスト構造において、i, j, k は ENTITY, EVENT, PROPERTY に分類された term の個数を示すインデックスである。各 term は文での意味役割を示す意味タグ tag が付与されている。

2.2 意味タグ

意味タグは自然言語処理研究においては語の多義性解消に用いられている [Yarowsky 92]。本研究での意味タグは、構文解析された文における NP, VP 等の統語情報と、専門家の手による [ASM 88] を加工した材料設計分野のシソーラスならびに基本用語の意味タグデータに基づいて付与される。付与する意味タグの種類については、従来さまざまな提案が語の意味役割とともになされてい

る [Fillmore 75, Saint-Dizier 95]. 本研究では今までに提案されている意味役割を参考にして, 対象文章とする材料設計論文の内容に沿うように設定している. 意味タグとその説明およびカテゴリを表 1 に示す.

表 1 意味タグ一覧

意味タグ	説明	カテゴリ
MAT(material)	実験に用いる材料	ENTITY
ALY(alloy)	// 合金	ENTITY
ATO(atom)	材料を構成する物質名	ENTITY
PART(part)	材料の一部	ENTITY
SOU(source)	実験に用いる補助材料	ENTITY
PRD(product)	実験の結果生じる副産物	ENTITY
ACT(action)	実験者によって行われる作用, 補助作業	EVENT
EVE(event)	実験における事象を記述するデフォルトのタグ	EVENT
MAN(manipulation)	実験者が行う実験に用いる材料の前処理	EVENT
TRAN(transition)	実験中の材料の状態変化	EVENT
TRANS(transformation)	// 形状の変化	EVENT
EVA(evaluation)	実験者の評価	EVENT
MES(measurement)	// 測定	EVENT
PHE(phenomenon)	実験材料の引き起こす現象	EVENT
QAL(quality)	実験結果として得られる質的なもの	PROPERTY
QAN(quantity)	// 量的なもの	PROPERTY
APA(apparatus)	実験に用いる装置	PROPERTY
TOL(tool)	// 小道具	PROPERTY
CON(container)	// 容器	PROPERTY
MET(method)	// 方法, 手法	PROPERTY
LAW(law)	// 法則, 原理	PROPERTY
CDT(condition)	実験条件	PROPERTY
PRO(property)	材料の外見の特徴	PROPERTY
COMP(composition)	// 化学組成	PROPERTY
STRU(structure)	// 物理構造	PROPERTY

表 1 で記した意味タグ付与を行う semantic_tagging 関数のアルゴリズムを示す. アルゴリズム中の thesaurus とは, 材料設計分野のシソーラス, termdata とは基本用語の意味タグデータを指す.

```

semantic_tagging(term)
Input: 構文解析された文の統語タグが付いた各語 (term)
Output: 意味タグを付与した各語
semantic_tag <- "UNKNOWN"
For each entry in termdata
    //entry: (semantic_tag2 term2)
    if (term1 == term2) || (term2 is the last word of term1) then
        semantic_tag <- semantic_tag2
    unless semantic_tag2 == "UNKNOWN" then
        return semantic_tag
For each entry in thesaurus
    //entry: (sub|rel term2 term3)
    //(sub x y): x is sub-concept of y
    //(rel x y): x is synonym of y
    if (term1 == term2) && (term1 is sub-concept of term3) then
        semantic_tag <- semantic_tagging(term3)
    unless semantic_tag == "UNKNOWN" then
        return semantic_tag
    if (term1 == term2) && (term1 is synonym of term3) then
        semantic_tag <- semantic_tagging(term3)
return semantic_tag

```

アルゴリズム1. semantic_tagging 関数

2.3 意味構造データの自動生成

意味構造データを自動生成する関数 `add_semantics` のしくみをアルゴリズム2に示す。`add_semantics` 関数は前節における `semantic_tagging` 関数を呼び出し、得られた意味タグの種類によって (2.1.2) で示した ENTITY, EVENT, PROPERTY の3つの意味ブロックのいずれかに分類する。(2.1.2) で示した文中の語の意味役割は統語情報を保持しつつ意味内容を持たせているが、例えば表層の統語情報では名詞であっても、意味的には事象を表す語もある。`add_semantic` 関数はそのような情報も考慮して分類する。3つの意味ブロックに分類された後、それらは一文単位で結合され、(2.1.2) のリスト構造を形成する。アルゴリズム1,2は、Common Lisp [Steele 90, Norvig 92, Graham 96] を用いて実装されている。

なお、LISP のプログラムでは構文解析を施したサンプル文章の先頭から一文ずつ読み込み、`add_semantics` 関数を適用し、出力としての意味構造データを順にファイルに蓄えていく。また、意味タグが決定できなかった語については“unknown”を示す“unk”タグが付与されている。


```
add_semantics(parsed_sentence)
```

Input: 統語解析を行った一文

Output: 意味構造データ

```
For each block in parsed_sentence
```

```
    //block: (syntactic_tag, term)
```

```
    semantic_tag <- semantic_tagging(term)
```

```
    Case syntactic_tag
```

```
    NP|NPL:
```

```
        Case semantic_tag
```

```
        MAT|ALY|ATO|PART: add to ENTITY block
```

```
        ACT|MAN|TRAN|TRANS|EVA|MES|PHE: add to EVENT block
```

```
        other: add to PROPERTY block
```

```
    VP:
```

```
        Case semantic_tag
```

```
        "UNKNOWN": semantic_tag <- "EVE", add to EVENT block
```

```
        other: add to EVENT block
```

```
return ENTITY + EVENT + PROPERTY
```

アルゴリズム 2. add_semantics 関数

3. Bayesian Belief Network (BBN)

3.1 概要

BBN はベイズの定理 [Bernardo 93, Casella 90] に基づく不確実な情報の表現と確率推論の枠組である。ベイズの定理は (3.1.1) で表される。

(3.1.1) Bayes Theorem:

$$P(h|D) = \frac{P(h|D)P(h)}{P(D)}$$

$P(h|D)$: posterior probability

$P(h)$, $P(D)$: prior probability

$P(D|h)$: conditional probability of D given h

BBN ではネットワークの各ノードを事象とし、ノード間のアークは接続されたノード間に因果関係が存在することを表している。各ノードにはその親ノードの確率変数に対する条件付き確率が与えられる。BBN を用いることで、対象領域で起こる事象の間の因果関係を完全かつコンパクトに記述できる [Heckerman 96, Nishida 99]。本研究では、共起する語にはそれらが持つ意味役割すなわち付与された意味タグ間に相互作用が存在するとして、結果から原因を推定する診断的推論の一つとして BBN を適用する。BBN の各ノードを各語に付与された意味タグとして、統語の階層構造にしたがってノードを結んでいく。BBN における各ノード間に成立する関係は (3.1.2) のように表される。

$$(3.1.2) \quad P(y_1, \dots, y_n) = \prod_{i=1}^n P(y_i | \text{Parents}(y_i))$$

(3.1.2) において、 $y_i (1 \leq i \leq n)$ は BBN の各ノードを示し、 $P(y_1, \dots, y_n)$ はそれらに特定の値を指定した時の共起確率を示している。 $\text{Parents}(y_i)$ はノード y_i に対する直接の親ノードの集合を示しており、 $P(y_i | \text{Parents}(y_i))$ は親ノードに特定の値を定めた時の y_i の起こる条件付き確率を指す。BBN の各ノードにはあらかじめ主観的に生起確率が与えられており、あるノードについて事実が判明した時、BBN 内でその事実を伝播させ、他のノードに対応する事象の確率を更新させる。BBN における各ノードの確率が変化する様子を図 1 に示す。図 1 において、 ev^+ 、 ev^- はノード V の上流、下流にある事象、 $\lambda(V)$ は V が観測されたときの V の下流にある事象 ev^- の生起確率、 $\lambda_v(U)$ はノード V から親ノード U に伝達されるメッセージ、 $\pi(V)$ は V の上流にある事象 ev^+ が観測されたときの V の生起確率、 $\pi_{V_j}(V)$ はノード V から子ノード V_j に伝達されるメッセージを表している。以上の設定によって BBN は上流・下流から流れてきたメッセージに基づいて目的のノード V の条件付き確率 $P(V | ev^+, ev^-)$ を計算し、その結果を上流・下流に伝播する。

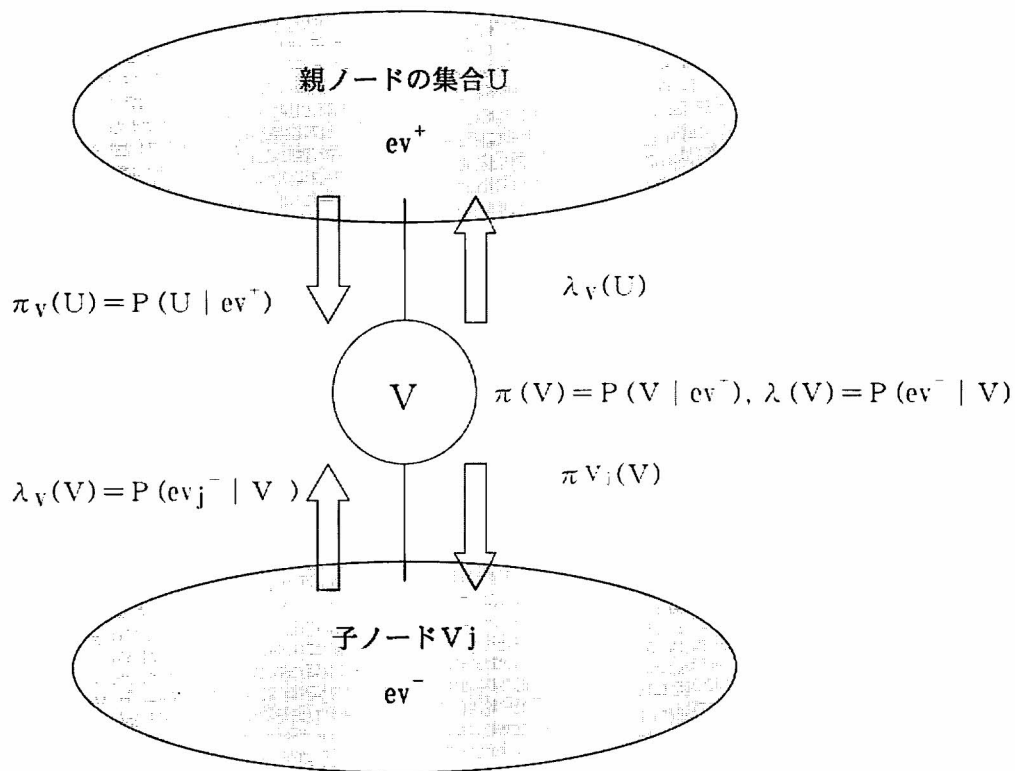


図1 BBN における確率伝播のしくみ

本研究では BBN の各ノードを文中に出現する意味タグとし、ENTITY に属する意味タグが出現したときに、EVENT に属する意味タグ、PROPERTY に属する意味タグが出現することは因果関係が成立すると仮定してノード（タグ）間の BBN を構成する。これにより各カテゴリに属する意味タグの共起関係が表される。つまり、2つの意味タグが共起しやすいほどそれらのノードの伝播後の確

率は高くなる。これ以降 ENTITY, EVENT, PROPERTY カテゴリに属する意味タグをそれぞれ ENTITY タグ, EVENT タグ, PROPERTY タグと記す。例として、文 1 をあげる。

(文 1)

Low density inclusions such as oxides and nitrides are allowed to rise to the surface in the hearth and are then either swept back by the beam and/or physically restrained by a mechanical barrier from reaching the solidifying ingot at the other end.

文 1 を構文解析した結果は以下になる。

(文 1 の構文解析結果)

(S (S (NP (NPL Low density inclusions) (ADJP such (PP as (NPL oxides and nitrides)))) (VP are (VP allowed (TOINF (VP to (VP rise (PP to (NPL the surface)) (PP in (NPL the hearth)))))) and (S (VP are (ADVP then (NP (NPL either swept back) (PP by (NPL the beam)))) (S (NPL and/or physically) (VP restrained (PP by (NPL a mechanical barrier)) (SBAR from (SS (VP reaching (NPL the solidifying ingot) (PP at (NPL the other end))))))))))

文 1 に対応する自動生成された意味構造データは以下のようなになる。

(文 1 の意味構造)

((MAT SOLIDIFYING INGOT) (ATO OXIDES AND NITRIDES)(PART SURFACE)(SOU BEAM))
 ((EVE ALLOWED) (EVE RESTRAINED)(EVE SWEEP BACK))
 ((QAL LOW DENSITY INCLUSIONS)
 (CON HEARTH)
 (TOL MECHANICAL BARRIER) (UNK OTHER END)))

ここで ENTITY, EVENT, PROPERTY に分けて記述すると、以下のようなになる。

ENTITY: ((MAT SOLIDIFYING INGOT) (ATO OXIDES AND NITRIDES) (PART SURFACE)(SOU BEAM))
 EVENT: ((EVE ALLOWED) (EVE RESTRAINED)(EVE SWEEP BACK))
 PROPERTY: ((QAL LOW DENSITY INCLUSIONS)(CON HEARTH) (TOL MECHANICAL BARRIER) (UNK OTHER END)))

文 1 の意味タグから図 2 の BBN を構成する。図 2 において、EVE ノードと CON ノード間には “in” という前置詞によってより強い関係が保たれている、すなわち “in” の出現によって 2 つのノードの共起する確率が高まっていると考える。図 1 と対応させて図 2 を説明する。図 2 では ENTITY, EVENT, PROPERTY の順でノードに親子関係を持たせている。図 2 において太線で囲まれたノードについて説明する。まず、親ノードから子ノードへ伝達される確率 π について考える。 $\pi_{eve}(U)$ は ENTITY カテゴリに属する意味タグの中で MAT タグの出現する確率 $P(MAT | ENTITY)$ であり ENTITY ノードから EVENT ノードへ伝達されるメッセージとなる。 $\pi_{pro}(EVE, “in”)$ は EVE タグと “in” が共起している時の確率で CON ノードへ伝達される。次に子ノードから

親ノードへ伝達されるメッセージを考える。 $\lambda_{eve}(MAT|EVENT)$ は EVENT タグと MAT タグの共起する確率である。 $\lambda_{pro}(CON) = P(CON|EVE, "in")$ は PROPERTY タグの中で CON タグの出現する確率を示す。

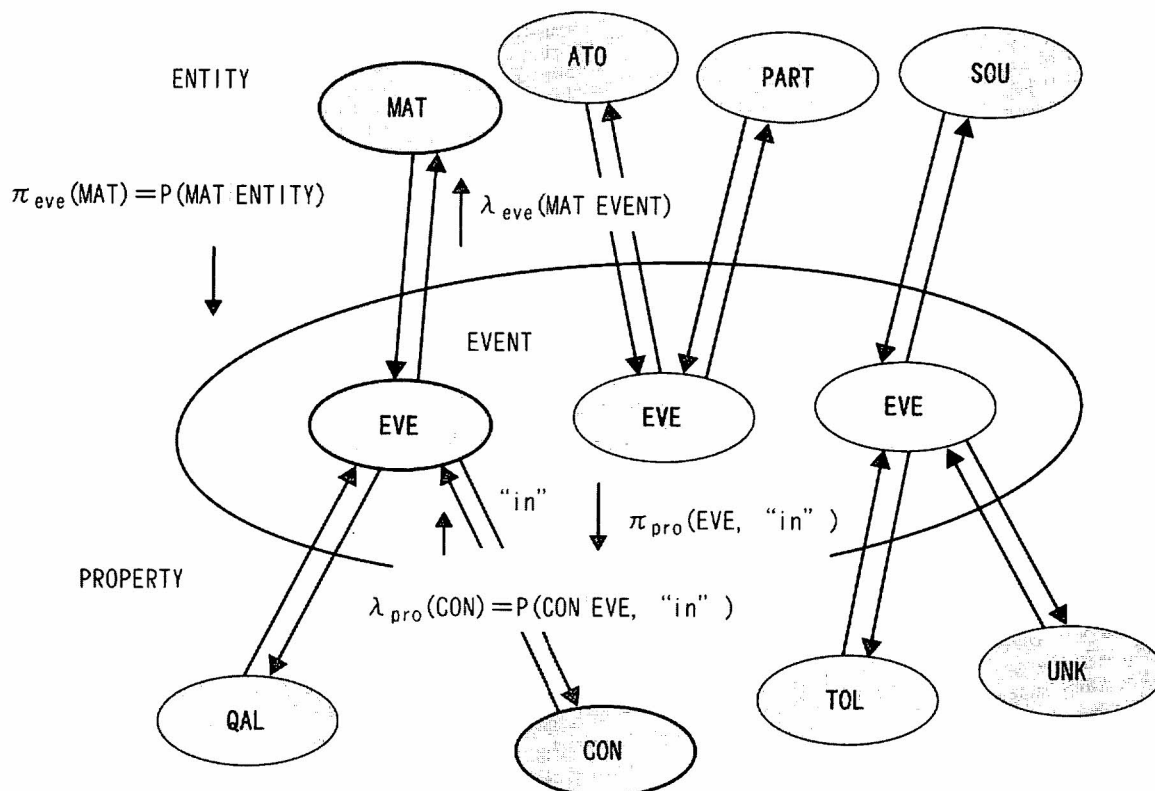


図2 意味タグの BBN における表現

以上の BBN の設定に基づいて、以下に場合についての共起情報が得られると考えられる。

- (1) PROPERTY タグが、EVENT タグと前置詞との共起によって変化する場合。
- (2) PROPERTY カテゴリに属する UNK タグは“unknown”を表しており、シソーラス等の知識から意味タグを決定できなかったことを示している。それらの UNK タグを付与された語の最も確からしい意味タグを ENTITY, EVENT カテゴリの意味タグとの共起から推論する場合。

2 つとも PROPERTY タグの決定に関わるものである。これらは実際の研究者の設計支援環境を考えた場合、材料や実験方法はある程度明確になっているのに対して、材料の持つ性質や、実験条件、実験結果における内容は研究者の抽出したい対象になることが多いので、本研究の目的に沿ったものであるといえる。BBN を用いた時、図2の PROPERTY に属するノードにおける PROPERTY タグの中で、EVE タグが BBN 伝播後に最も高い生起確率をとる意味タグが適切な意味タグになる。

3.2 BBN での実装

本研究では BBN における確率の算出を HUGIN [Hugin 96] のデモ版を用いて行った。最初に

各ノードには、親ノードである意味タグが文中に出現した場合の条件付き確率を与えなければならない。ここで条件付き確率を (3.2.1) で定義する。エントロピー (Entropy) とは、情報理論で頻繁に用いられる概念であり、任意のサンプルの集合における不純度つまりサンプルデータがどのくらい不規則に混じり合っているのかを示すものである。また、サンプルデータを分類するために必要とするビット量を表している。エントロピーが少ないと、サンプルデータを分類しやすく、多いと分類しにくいことを示す。 $n(x, y)$ は x, y が共起する頻度を示す。 a, b は重み付けのための定数である。

(3.2.1) において、 tag の条件付き確率を、 tag が親ノードの意味タグと共起する情報だけから得るのではなく、エントロピーを取り入れて a, b によって重みをつけて算出する。本研究ではエントロピーを、語に対して意味タグを付与する作業における容易さの目安として用いる。ある状況に対してある意味タグが少しの場合にしか付与されないならば、エントロピーは少ないが、逆にそのような場合に別の意味タグを付与することは許されないこともありうる。これは少ない語についての貴重度を無視できないことを表しており、それに対する重みを備える必要がある。一方で、ある状況に対してほとんどの場合一つの意味タグが付与されることが適しているならば、その意味タグは受容度が大きく、付与する作業も容易である。これら二つの場合に対する重みとしてエントロピーを導入し、貴重度と受容度を考慮する。(3.2.1) における $1 - Entropy(tag)$ は、貴重度と受容度が大きい意味タグについて、より条件付き確率に多くの値が加えられるように設定している。

(3.2.1)

$$P(tag|Parent_tag) = a \times \frac{n(tag, Parent_tag)}{n(Parent_tag)} + b \times (1 - Entropy(tag))$$

$$Entropy(tag) = -(p^+ \log_2 p^+) - (p^- \log_2 p^-)$$

$$p^+ = n(tag, Parent_tag), p^- = n(\neg tag, Parent_tag)$$

(3.2.1)で、 $P(tag|Parent_tag)$ は候補となる意味タグが文中の目的の位置に $Parent_tag$ と共起する条件付き確率、 p^+ はサンプルデータ中で、 tag が親ノードにおける意味タグと共起する割合、 p^- は同じカテゴリの tag 以外の意味タグが親ノードと共起する割合を示す。

4. BBN を用いた意味タグの推定

4.1 EVENT タグと前置詞との共起による PROPERTY タグの推定

例として文 2 とその意味構造をあげる。

(文 2)

It was then filled into mild steel cans under vacuum and extruded to 130 mm diameter bars.

(文 2 の意味構造)

((MAT 130 mm diameter bars))

((EVE filled) (EVE extruded))

((CON mild steel cans) (CDT vacuum) (QNT 130 mm)))

構文解析の結果から得られる係り受けを考慮すると文2のBBNは図3になる。文2で前置詞と共に起している語の一つに“vacuum”がある。サンプルデータから図3のBBNを構築した場合の初期確率を算出し、PROPERTYタグがEVEならびに前置詞“under”と共に起した場合、BBNにおけるメッセージの伝播によって、最もEVEタグの確率が高くなるPROPERTYタグを求める。

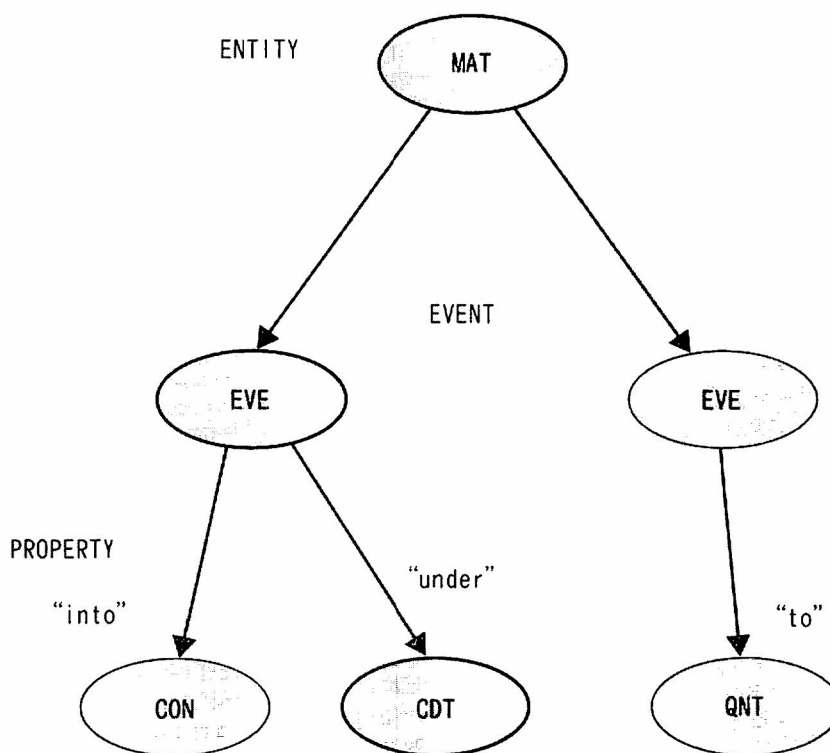


図3 文2のBBN表現

図3のBBNにおいて、ノードに与える初期確率を以下に示す。 $n(x, y)$ は x, y が共起する頻度を示す。

- (1) MAT ノードに与える初期確率

$$P(MAT|ENTITY)$$

$$p^+ = n(MAT|ENTITY), p^- = n(\overline{MAT}, ENTITY)$$

$$ENTITY = \{MAT, ALY, ATO, PART, SOU, PRD\}$$

- (2) EVE ノードに与える初期確率

$$P(EVE|MAT)$$

$$p^+ = n(EVE, MAT), p^- = n(\overline{EVE}, MAT)$$

- (3) CDT ノードに与える初期確率

$$P(CDT|EVE, "under")$$

$$p^+ = n(CDT, EVE, "under"), p^- = n(\overline{CDT}, EVE, "under")$$

サンプルデータとして材料設計に関する論文100件を用いた。一論文あたり英文で平均8ページ、約3,500ワードである。(1)～(3)に示した確率をPROPERTYタグについて計算し、BBNを用いて

生起確率を求めた。サンプルデータにおける PROPERTY タグの出現頻度、エントロピー、伝播前の生起確率、BBN による確率伝播後の EVE タグの生起確率を表 2 に示す。

表 2 PROPERTY タグの出現頻度と EVE タグの BBN 伝播後の生起確率

tag _k	n(tag _k , EVE, “under”)	Entropy	P(tag _k EVE, “under”)	P(EVE MAT, “under”, tag _k)
QAL	18	.569	.194	.222
QAN	35	.828	.243	.264
APA	0	.000	.000	.000
TOL	0	.000	.000	.000
CON	2	.111	.190	.219
MET	9	.355	.183	.212
LAW	0	.000	.000	.000
CDT	66	.999	.394	.367
PRO	4	.629	.185	.214
COMP	0	.000	.000	.000
STRU	0	.000	.000	.000

n(MAT) = 10,238, n(ENTITY) = 28,657, n(EVE.MAT) = 7,147, n(EVENT) = 26,943

a = 0.8, b = 0.2

4.2 PROPERTY タグが UNK である語の意味タグの推定

文 2 で “vacuum” の意味タグが未知であるとして実験を行う。BBN を用いて EVE タグの生起確率が最大になる PROPERTY タグを設定することは、文 2 で適切な意味タグに変更することと同じである。よって、BBN により、“under” と共起する場合、CDT タグを用いることで EVE タグの生起確率が最大になるので、“vacuum” の意味タグは CDT に設定される。

5. 評価

5.1 Naive Bayes Decision Rule

意味タグの推定には、共起情報の抽出でよく用いられる Naive Bayes の手法があげられる [Gale 92, Michie 94, Russell 95, Mitchell 97]。Naive Bayes では、(5.1.1) で示される式により、意味タグが決定される。(5.1.1) から tag を推定した結果を表 3 に示す。

(5.1.1)

$$tag = \underset{tag_k \in PROPERTY}{\operatorname{argmax}} [\log_2 P(tag_k) + \log_2 P(EVE|tag_k) + \log_2 P(\text{“under”}|tag_k)]$$

$$PROPERTY = \{QAL, QAN, APA, TOL, CON, MET, LAW, CDT, PRO, COMP, STRU\}$$

$$P(tag_k) = \frac{n(tag_k)}{n(PROPERTY)}$$

$$P(EVE|tag_k) = \frac{n(EVE, tag_k)}{n(tag_k)}, P(\text{“under”}|tag_k) = \frac{n(\text{“under”, } tag_k)}{n(tag_k)}$$

表3 Naive Bayes による PROPERTY タグの推定

tag _k	n(tag _k)	n(EVE, tag _k)	n("under", tag _k)	Naive Bayes
QAL	9,486	834	20	-14.717
QAN	8,346	932	40	-13.372
APA	2,074	146	0	-8.342
TOL	2,056	352	0	-7.072
CON	355	16	2	-19.003
MET	4,105	354	10	-15.745
LAW	433	12	0	-11.947
CDT	7,208	491	82	-13.050
PRO	6,678	491	6	-16.712
COMP	3,523	221	0	-7.744
STRU	3,105	179	0	-8.048

$n(\text{PROPERTY}) = 47,369$

5.2 本手法と Naive Bayes との比較

本研究で用いた BBN とエントロピーによる手法と、Naive Bayes の決定規則による手法とを比較する。Naive Bayes での結果は表3にあるように、TOL が最大になっているが、TOL の $P(\text{"under"} | \text{tag}_k)$ は0である。これは TOL と “under” がまったく共起していないことを示しており、本論文で目的とする直前の前置詞を手がかりに意味タグの推定を行うことから外れている。Naive Bayes では、 $P(\text{"under"} | \text{tag}_k)$ で表される条件付き確率に対する底が2の対数を計算するために、 $n(\text{tag}_k)$ に比べて $n(\text{"under"}, \text{tag}_k)$ が非常に小さい場合には対数の値が小さくなる。そのため表3の一部の意味タグにあるようなデータが0の場合には、意味タグと “under” が共起していないにも拘わらず、対数の値が Naive Bayes に影響しない。よって $P(\text{"under"} | \text{tag}_k)$ の値が0の意味タグは無視して考えることにする。

そうした場合、CDT が -13.050 と最大値をとる。一方で本手法でも、表2から EVE タグの BBN 伝播後の生起確率を最大にするのは CDT の0.367である。次に大きい値は Naive Bayes では QAN の -13.372であり、本手法では QAN の0.264である。Naive Bayes では他の値と比べると CDT と QAN の値は接近しており、文章の意味構造を作成する以前の構文解析のエラーを考慮すると、値が最大であっても CDT が最適であると決定するためには有意か否かの判定が必要である。それに対して本手法では、CDT と QAN の値は他の意味タグの値も考えた場合でも CDT の方が顕著に大きく、最適な意味タグとして CDT を決定できる。

Naive Bayes の欠点として、文中で近接する語の間に強い依存関係が存在する場合、それらの関係をうまく量的に表せないことがある。Naive Bayes が “Bag of Words Model” [Manning 99] をベースにした推定手法であることから明らかであろう。本稿で推定の目的とする対象が、前置詞とその直後の語の意味タグという、位置に関する強い依存関係があるために、Naive Bayes で適切な

結果が得られなかったとも考えられる。

別の問題として、本手法では $n(\neg tag, Parent_tag)$ を求める際に、tag 以外の PROPERTY タグと共起する Parent_tag (ここでは EVE タグと前置詞 “under”) をカウントしている。文章データの中には PROPERTY タグと共起しない Parent_tag も存在する。本手法では PROPERTY タグの推定という限定したタスクに適用するためにそのように設定し、PROPERTY タグ以外はノイズとして無視した形になっているが、この設定が妥当であるか否かはさらなる検証を要する。

6. むすび

本論文ではまず、文章の意味表現のために提案した意味構造データについて述べた。それらに付与されている意味タグの説明を行い、意味タグを自動付与するアルゴリズムを示した。アルゴリズムは Common Lisp を用いて実装されている。次に付与した意味タグの中で、特定の前置詞と共起する語に対して、BBN を用いて意味タグの推定を行う手法を提案した。BBN に与える条件付き確率は、共起頻度とエントロピーの値から重みを付けて決定した。提案した手法と、Naive Bayes の決定規則で行った意味タグの推定結果を比較し考察した。

今後の課題として、対象データの充実と、BBN による意味タグの推定手法の精緻化と拡張が考えられる。

参考文献

- [ASM 88] *Thesaurus of Metallurgical Terms* 8th edition, ASM International Metals Park Ohio U.S.A. & Institute of Metals, London, England (1988).
- [Bernardo 93] Bernardo, J. M., & Smith, A. F. M.: *Bayesian Theory*, John Wiley & Sons (1993).
- [Casella 90] Casella, G., & Berger, R. L.: *Statistical Inference*, Duxbury Press (1990).
- [Domingos 96] Domingos, P., & Pazzani, M.: Beyond independence: Conditions for the optimality of the simple Bayesian classifier, *Proceedings of the 13th International Conference on Machine Learning*, pp.105–112 (1996).
- [Fillmore 75] Fillmore, C. J.: The Case for Case, In Bach & Harms, T.(eds.), *Universals in Linguistic Theory*, Holt, Rinehart & Winston. 田中春美, 船城道雄(監訳), 格文法の原理: 言語の意味と構造, 三省堂 (1975).
- [Gale 92] Gale, W. A., Church K. W., & Yarowsky, D.: A method for disambiguating word senses in a large corpus, *Computers and Humanities* 26, pp.415–439 (1992).
- [Graham 96] Graham, P.: *ANSI Common Lisp*, Prentice Hall (1996).
- [Heckerman 96] Heckerman, D.: Bayesian Networks for Knowledge Discovery, *Advances in Knowledge Discovery and Data Mining*, pp.273–305 (1996).
- [Hugin 96] *Hugin Gui Version 4.5 Demo*, Hugin Expert A/S, <http://www.hugin.dk> (1996).
- [Jackendoff 92] Jackendoff, R.: *Semantic Structures*, MIT Press (1992).
- [Jensen 96] Jensen, F. V.: *An Introduction to bayesian networks*, New York: Springer Verlag (1996).
- [Manning 99] Manning, C. D.: *Foundations of Statistical Natural Language Processing*, MIT Press (1999).
- [Michie 94] Michie, D. J., & Taylor, C. C.: *Machine Learning, neural and statistical classification*, (edited collection), New York: Ellis Horwood (1994).
- [Mitchell 97] Mitchell, T. M.: *Machine Learning*, WCB McGraw-Hill (1997).

- [Nakabasami 99] Nakabasami, C., & Hoshimoto, K.: Extracting Knowledge from Technical Papers in Metallurgy on the basis of a Generative Lexicon, *Proceedings of the Conference Pacific Association for Computational Linguistics*, pp.137-142 (1999).
- [Nishida 99] 西田 豊明: 人工知能の基礎, 丸善 (1999).
- [Norvig 92] Norvig, P.: *Paradigms of Artificial Intelligence Programming: Case Studies in Common Lisp*, Morgan Kaufmann Publishers (1992).
- [Pustejovsky 93] Pustejovsky, J., Bergler, S. and Anick, P.: Lexical Semantic techniques for corpus analysis, *Computational Linguistics 19*, pp.331-358 (1993).
- [Pustejovsky 95] Pustejovsky, J.: *The Generative Lexicon*, MIT Press (1995).
- [Russell 95] Russell, S., & Norvig, P.: *Artificial Intelligence: A modern approach*, Englewood Cliffs, NJ: Prentice-Hall (1995).
- [Saint-Dizier 95] Saint-Dizier, P.: *Computational Lexical Semantics*, Cambridge University Press (1995).
- [Steele 90] Steele, Guy: *Common Lisp*, Digital Press (1990).
- [Yarowsky 92] Yarowsky, D.: Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora, *In Proceedings, COLING-92*, pp.454-460 (1992).

An Attempt to Incorporate Co-occurring Information of Words into Semantic Representation for Machine-readable Texts using Bayesian Belief Networks

Chieko NAKABASAMI

Abstract

In the processing of natural language, it is critical to grasp whole meanings that are conveyed via sentences. Sophisticated semantic representation is necessary for the automatic construction of sentence structure. For the semantic representation of sentence structure, this paper proposes a “semantic block” in which both the syntactic information and the thematic roles of words appearing in sentences are taken into account. This study focuses on technical papers on material design. The sentence structure is composed of three semantic blocks: “ENTITY,” “EVENT,” and “PROPERTY.” The ENTITY block represents physical objects and agents that cause events. The EVENT block represents the events themselves. The PROPERTY block represents the properties of objects in the ENTITY block. The semantic block is implemented by a list structure. This paper reports the results obtained after applying this method to extract information from papers on material design.

In addition, Bayesian Belief Networks (hereafter BBN) are used in order to incorporate co-occurring information of semantic tags into the semantic block. This paper discusses possible ways to use BBN for the semantic representation of a semantic block. The stochastic quantification estimated by BBN enables the semantic block to integrate co-occurring information into the semantic representation of words appearing in it.

Key Words : semantic representation, semantic tagging, bayesian belief network