情報構造に基づく文意味計算方式の提案

2N－6

稲垣 博人　早川 和宏　田中 一男

NTTサイバーソリューション研究所

1 はじめに

インターネットの普及により，多くの人がインターネットに触れ，多くの人がインターネット上にデジタル化された情報を流通するようになってきた．このようなデジタル情報流通社会において，限られた時間内に自分に必要な情報を効率良く見つけるための支援技術として，情報の要約，分類，整理などが重要視されている．これらの情報処理のためには，我々は，以前から，情報の構造化が必須であることを提案した[1]．もちろん，情報の構造としては種々の構造が考えられる．我々は，事象という捉え方により，ある行為をいつ，誰が，どこで発生させたかというようなダイナミックな物事の変化に基づく情報構造と，その情報構造の自動抽出方法を提案した[2]．本稿では，さらに，情報構造として，事象という概念だけでなく，静的な状態という概念を加えた情報構造を文の意味構造として規定する．また，意味構造の計算手法を単純化するため，構文解析により生成された2分木と，その2分木のノード間の演算により意味構造を計算する手法を提案する．

2 文の意味構造

文の意味構造として，ここでは，大きく分けると以下のような2種類に分けられるとした．

- ダイナミックな物事の変化
- スタティックな物事の状態

ダイナミックな物事の変化に関しては，以前提案している事象という概念を適用する．これは，物事の変化は，変化をさせた主体と変化した対象が，いつ，どこで，どのように変化したかを定義する構造である．そのため情報構造としては，主体（属性名はAGENT．）と，対象（OBJECT）および，主体が対象に対して行なった行為（ACTION），その他，TIME，LOCATIONなどの基本的な事象の要素とそれらを修飾する修飾要素から構成されている．例えば，“A社は5億円を増資した．”というような文では，AGENT=“A社は”となり，OBJECT=“5億円を”，ACTION=“増資”となる．

A document semantics calculus based on an information structure of sentences.
Hirohito INAGAKI, Kazuhiro HAYAKAWA, and Kazuo TANAKA.
NTT Cyber Solutions Laboratories

表1: 数値計算と文意味構造計算との比較

	数値計算	文意味構造計算
計算単位	数式単位	属性ごとの2項演算単位
条件	-	演算条件
数値	数値	属性値（数値，文字列等）
数式	数式	2分木（構文木）
演算子	四則演算等	約10種類の演算子

一方，スタティックな物事の状態とは，世界に対して何ら変化を起こしていない（起こっていない）が，ある主体の状態，属性などを表現した文である．また，否定表現で，ある主体の状態を表現する場合もありうる．ここでは，肯定的状態をISという属性名で表し，否定的状態をIS_NOTと表す．例えば，“A社の資本金は5億円である．”という例文では，OBJECT=“A社の資本金は”となり，肯定的状態として，IS=“5億円”が与えられる．否定的状態の例としては，“A社はB社の子会社でない．”などの表現例が当てはまる．

3 文意味構造の計算方式

文の意味構造の計算は，2で述べた，文の意味構造として規定した各属性に形態素を当てはめる処理であると単純化した．

表1に数値計算と本稿で提案する文意味構造計算との違いを示すが，計算単位や条件が付く以外は，比較的類似している．数値計算では，数値と数式と演算子の演算方法を規定することにより，計算が行なわれる．

一方，文意味構造の計算では，数式に当たる部分が，構文解析により生成された2分木である．2分木を構成する各ノードは，複数の属性に対する属性値を持ち，各属性ごとに設定された演算子に基づき適切な演算を行う．ただし，属性はいつも計算されるわけではなく，属性のもつ演算条件を満たした場合のみ計算される．

属性値としては，例えば，単純な数値や，表記などの文字列，意味素性などの属性が記述されるだけでなく，文の意味構造で示したような，AGENT，OBJECT，ACTION，ISなどの属性も登録されている．

単純な数値計算の場合，左ノードの値が3で，右ノードの値が5で，演算子が“+”であれば，3＋5＝8と等価となる．また，文字列であれば，左ノードの文字列が“東京”で，右ノードの文字列が“都庁”で，演算子が左ノードと右ノードの連結（例えば連結の演算子を⊕で

表す.)であれば,“東京”⊕“都庁”=“東京都庁”を計算していることになる.

属性の演算子としては,文字列の結合を行なう関数や四則演算以外に,上書き演算子,最大,最少演算子,ポインタの設定演算子などがある.

文意味構造の計算方式の実際の処理の手順を以下に示す.

Step1 形態素・係り受け解析処理

Step2 2分木生成処理

Step3 文意味計算用データ付与処理

Step4 ノード計算処理

Step4(1) 計算対象ノードを2分木の一番上のノードとする.

Step4(2) 計算対象ノードの属性値が計算されていない場合,2分木の左右のノードを計算対象ノードとする.

Step4(3) 左右のノードの属性値が計算されていたら,計算対象ノードの属性値を計算する.

Step4(4) 各属性の計算結果を計算対象ノードに書き込む.

Step4(5) すべてのノードの属性値を計算したら終了.
計算されていない場合,*Step4(2)*にもどる.

Step5 意味計算結果の出力処理

形態素・係り受け解析処理では,入力対象の文章を形態素解析し,係り受け解析を行なう.形態素解析,係り受け解析については,InfoBeeを利用した.*Step2*では*Step1*の情報をもとに2分木を作成する.

文意味計算用データ付与処理では.文意味計算用辞書に基づき,各文字列(意味素)に対し,文意味計算に必要な,種々の属性に対する属性値を付与する.形態素解析処理における形態素単位に必ずしも付与されるわけではなく,1個以上の形態素から構成される文字列(意味素)に付与される.文意味計算用辞書は,辞書見出しに対して,属性名と属性値と演算条件が付与されている.

図1に形態素,係り受け解析の後に意味素に対して属性値が付与された2分木の例を示す.“SEM”は,意味素性を表す.“AGENT=(SEM=company)”は,括弧書きが演算条件を表している.つまり,AGENT属性は,SEM(意味素性)として,“company”をもつノードがなることを表す.

*Step4*では,*Step3*の2分木をもとに,*Step4(1)*～*Step4(5)*ステップに従い,2分木の各ノードの属性の計算が終了するよう再帰的にノード計算処理が行なわれる.最終的に,文の意味が2分木の一番上位のノードに集約される.(図2に文意味計算結果例を示す.)

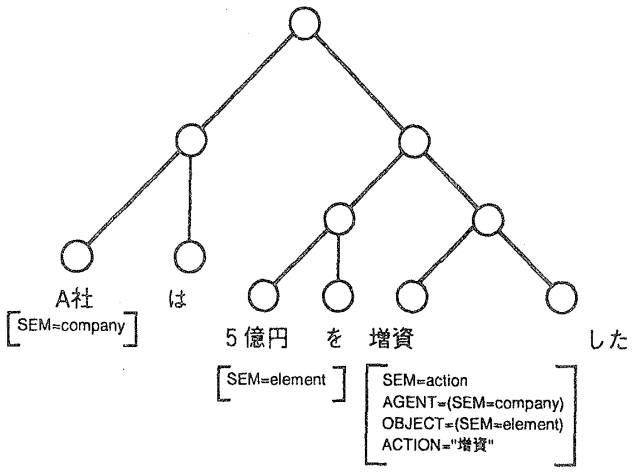

図 1: 2分木構造例

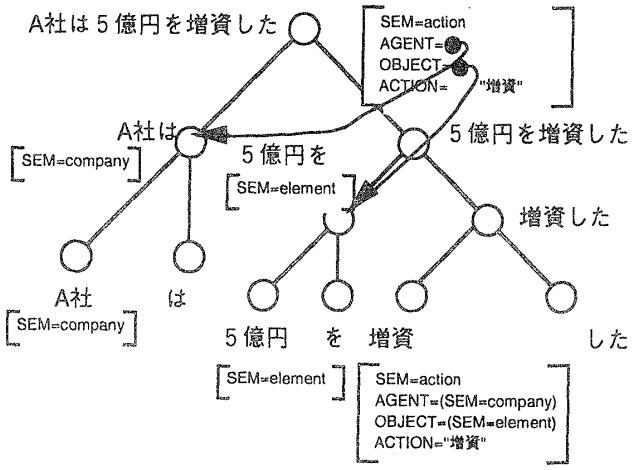

図 2: 文意味計算結果例

4 まとめ

本稿では,文の意味構造として,ダイナミックな事象とスタティックな状態に基づく意味構造を定義し,さらに,その情報構造を計算する手法として,構文解析により生成された2分木を数式と捉え,2分木の各ノードの持つ属性値間の演算により算出する手法を提案した.

今後は,本計算手法により,実際の文章に対する情報構造化を進め,本情報構造を用いた要約・分類・整理などのアプリケーションを実装する.

参考文献

1) 稲垣博人, 早川和宏, 田中一男. 情報流通向けテキストコンテンツ要約手法について. 情報処理学会デジタルドキュメント研究会,DD15-3, 1998.

2) 稲垣博人, 中川透. 出来事型情報の構造化. 情報処理学会第46回全国大会,4A-7, 1993.