

5P-7

事象型要約情報抽出システム*

稲垣 博人、中川 透†

NTTヒューマンインタフェース研究所‡

1 はじめに

我々は、文書や文で伝達される情報の形式化やその形式化に必要な知識(記述量、記述方法)の明確化が、伝達される情報を解析したり、解析した情報を用いたアプリケーションを実現する際に重要であると考えている[1, 2]。本稿では、伝達される情報が現実世界における事柄を表現する場合の伝達情報の形式を規定し、さらに、その形式を用いて実現される事象型要約情報抽出システム(ABEX system)について述べる。このシステムでは、伝達される情報の中で現実世界における事象だけをフィルタリングし、その情報を事象の生起時間と類似性の観点でグラフ表現したり、様相の観点から文を分類・表示する。

2 伝達情報の形式

伝達情報が現実世界における事柄を表現するとすると、伝達情報は、現実世界の変化を表現する事象(Event;E)と状態(State;S)、さらにそれらに対して人間が付与した様相(Modality;M)から成り立っていると仮定できる。ここでは、状態(S)は扱わないと仮定すると、新聞のような情報取得者(IP)が情報を情報伝達者(SP)に伝え、それを情報受信者(HR)に言葉で伝える形態での情報受信者(HR)に伝達される伝達情報は、

$$\sum_i M_{SP_i}(M_{IP_i}(E_i)) + \sum_j \sum_k R_{j,k}(M_{SP_j}(M_{IP_j}(E_j)), M_{SP_k}(M_{IP_k}(E_k)))$$

と表せる。そこで、伝達情報を表すのに必要な、事象(E)、様相(M)、関係(R)の枠組を規定し、伝達情報の細部の形式を規定する。

伝達情報の細部を規定し、処理に必要とされる知識を限定するために、領域の限定を行う。領域の限定は、事象を生起させる動作主とその動作主が生起させる動作の領域の決定を意味する。(ここでは、動作主は企業に限定する。)

2.1 事象(E)の枠組

ある領域における活動を表すのに必要な意味素性(DDF)は、企業活動では、individual(知的個体および知的個体の集合)、element(活動する時に使用される実体)、thing(活動により変化する実体;企業活動の領域では、money, man, productの3つに細分

化。productについて、補助的な表現のみ扱う。)、action(活動)、time(活動が生じた時間)、location(活動が生じた場所)の6つの意味素性である。

また、意味素性の事象における役割(Case)を、agent(事象における動作主)、object(動作主の活動の対象)、action(動作主の活動)、from(object自身の基点)、to(object自身の着点)、time(時間)、location(場所)の7基本格と基本格を補助する7準基本格で表現すると規定する。

2.2 様相(M)や関係(R)の枠組

本システムでは、事象の解析結果を用いて、事象の時間と類似性の関係のグラフ化と、様相の観点からの分類・表示の2つのアプリケーションを実行するため、時間、様相概念の把握、事象間の関係抽出がさらに必要となる。事象における時間の概念は、現実世界の時間の概念と同一であるので、そのまま現実世界の表す時間概念を扱えば良い。

様相(M)は、事象に対する人間が付与した様相表現であるので、ここでは、テンス、アスペクト、モーダルで様相(M)を表現する。テンス、アスペクトについては、通常のカテゴリを用い、モーダルは様相演算子と事象に対する人間(情報取得者または情報伝達者)の心的状態(信念、欲求、意図[3])を用いて表現する。

事象間の関係(R)においては、記述すべき知識を規定するため、以下の3種類の関係のみ解析する。

- スクリプト的関係(事象の全ての役割が一致し、かつ、関係(R)が時間的に必然なもの。)
- 同一関係(事象の全ての役割が一致する。)
- 物理的関係(事象のある役割が一個以上一致する。)

各事象の役割は事象(E)の解析によって決定されるので、時間的關係を規定する知識だけが必要となる。

3 知識データと処理アルゴリズム

3.1 知識データ

知識データの形式は、汎用性を考慮して、各種の素性と素性値からなるような素性構造の集合で一元的に表現する。

素性としては、事象、様相、関係の枠組で設定したもの以外に、意味素性の制限や事象(E)における役割を決定するconstraint素性を用意する。

*Event-oriented Abstraction Extraction System

†INAGAKI, Hirohito and NAKAGAWA, Tohru

‡NTT Human Interface Laboratories

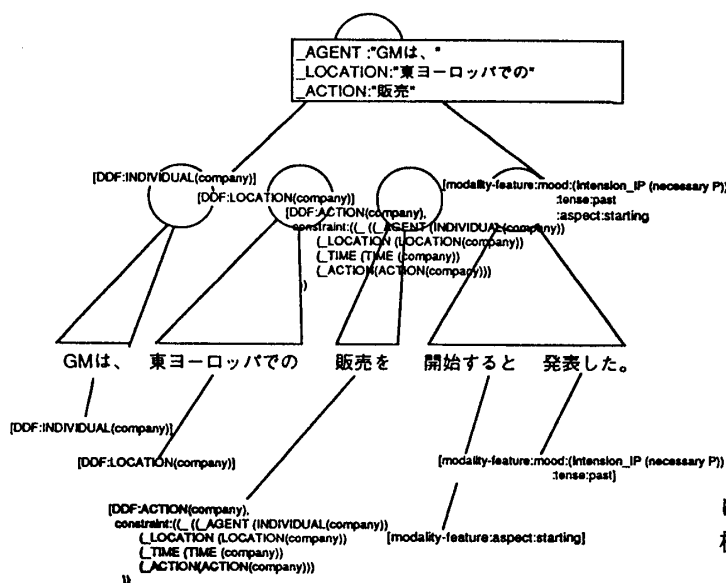


図 1: 最終的な事象解析結果の例
但し、各単語にリンクしているのは各単語の素性構造、丸はノードを表す。また、一番上は、事象の素性構造が完成した時の変数の値を表す。

3.2 処理アルゴリズム

3.2.1 事象一次選択処理

事象一次選択処理では、constraint 素性を持つ単語（通常は企業の活動を表す動詞）に着目した処理で、文章を事象単位に分割すると共に、設定した領域（ここでは企業活動）に関係する事象だけを抽出する。

3.2.2 事象解析処理

選択された事象について、まず、単語認定、構文木の作成を行い、各単語に素性構造を付与する¹。次に、各語に付与される素性構造を構文木に従って unification²を繰り返し、constraint 素性を満足する事象の素性構造を作成する。（図 1）

unification が成功し、事象の素性構造が全て完成したら終了。つまり、事象の素性構造が完成するということは、その素性構造は、各素性を満足する素性構造であり、事象における役割や様相(M)、時間概念などの各素性値を矛盾なく保持することを意味する。unification が失敗したら、別の候補に対して同様の処理を行う。

3.2.3 事象間の解析処理

事象間の解析処理では、関係(R)の枠組で述べた3つの事象間の関係を解析する。スクリプト的關係では、事象(E)間や様相(M)間の時間的に規定される関係をスクリプト知識に記述し、その知識を用いて解析する。それ以外の関係については、各事象(E)にお

¹ここでは、単語の認定及び素性構造付与は形態素解析を用い、構文木は、係り受け解析を用いて自動的に作成する。

²unification においては、親ノードの文字列は、子ノードの各文字列を結合した値を持ち、素性構造は子ノードの素性構造を unification した値を持つ。

客観的事実

高田機工は十六日、橋梁の大型化に対応して和歌山県下津町に進出し、新工場を建設することで建設用地は旧丸善石油精製工場跡地の一部を、丸善下津興産から土地十七万八千五百六十六平方メートルを約百億円で買収、来春から新工場建設に着工する。これに伴い、岸和田工場は閉鎖し売却する。新工場は和歌山県海草郡下津町大字方地内に、一期計画として最新設備の橋梁工場を建設する。

企業の広報活動

用地売買契約に調印したと発表した。

企業の意向

平成五年四月操業の予定。

図 2: 情報取得者の様相の観点から分類・表示した例

ける基本格、準基本格の要素を比較することにより解析される。

3.2.4 事象のグラフ表現処理

グラフ表現では、X 軸として事象の生起する時間、Y 軸として事象間の類似性を用いて、各事象の関係をグラフ的に表現する。X 軸の時間は、事象の素性構造の時間素性から取得する。Y 軸の事象間の類似性は、スクリプト的關係にある事象をもっとも高くし、それ以外については、事象間の基本格、準基本格の一致率から求める。

3.2.5 様相表現による文の分類・表示処理

様相表現による文の分類では、事象の持つ様相表現に基づいて文を分類し、分類された文を接続して、表示する。事象の様相表現は、事象の素性構造の様相素性から取得する。様相素性が全く付与されていない事象については、分類様相名は「客観的事実」とした。それ以外については、素性に付与されている値を直接分類名として用いる。分類後、文は文書での出現順に接続され、それぞれの様相名を付与して表示される。情報取得者の様相(M_{IP})の観点から分類すると図 2 のような分類表示が得られる。

4 まとめ

本稿では、事象型要約情報抽出システムの基本的な枠組とその枠組を用いて、時間と事象間の類似性を軸としてグラフ表現したり、事象の様相の観点から文を自動分類・表示するシステムについて述べた。

今後は、大規模なデータでの評価や企業以外の領域での評価を進める。

参考文献

- [1] 稲垣博人. 事象解析による要約情報の抽出. 情報処理学会自然言語研究会, NL84-3, 1991.
- [2] 稲垣博人, 中川透. 情報、処理の部分性を考慮した文章解析の実現. 情報処理学会第 43 回大会, 5H-5, 1991.
- [3] M.E. Bratman. Plans and resource-bounded practical reasoning. *Computational Intelligence* 4, pp.349-355, 1988.