

# ラベル付き依存関係に基づく用例文検索

江川 誠二\*, 加藤 芳秀, 松原 茂樹 (名古屋大学)

Searching Example Sentences based on Typed Dependency

Seiji Egawa, Yoshihide Kato, Shigeki Matsubara (Nagoya University)

## 1 はじめに

大規模コーパスの効果的な活用には、コーパス検索が重要である。これまでに依存構造を用いたコーパス検索手法が提案されている [1]。この手法は、キーワード系列をクエリとして受け取り、キーワードをこの順で含むコーパスの各文において、キーワードが形成する依存構造パターンを同定し、文をパターン別に分類する。ユーザは、キーワードを入力するだけで依存構造を活用した検索を行うことができる。

しかしこの手法では、各キーワード間の依存関係の有無のみを考慮し、その種類は利用しておらず、関係の種類に応じた分類はできない。

そこで本稿では、依存関係の種類を考慮した依存構造パターンによってコーパス中の文を分類する方法を提案する。これにより、より細かく分類された検索結果をユーザに提示できる。

## 2 依存構造に基づくコーパス検索

文献 [1] の手法では、入力としてキーワード系列を受け取り、各文にあらかじめ付与された依存構造を参照しながら入力されたキーワード間の依存関係を同定し、同定された依存構造パターンに従って文を分類する。

依存構造パターンは 3 項組  $d = (h, L, R)$  で、 $h$  は単語位置であり、これを  $d$  の主辞と呼ぶ。  $L$  および  $R$  は依存構造パターンのリストである。  $L$  中の依存構造パターンの主辞が左から  $h$  に依存することを意味し、  $R$  の場合は右からの依存を意味する。

依存構造パターンはキーワードからボトムアップに生成され、2 つの依存構造パターンの主辞の間に依存関係が成り立つとき、それらを結合して、より大きなパターンを生成する。

例えば、Fig.1 の文と依存構造、及びクエリ “have mind” が与えられると、mind が have に右から依存するという関係が成り立つのでこれらを結合し、Fig.2 の依存構造パターンが得られる。

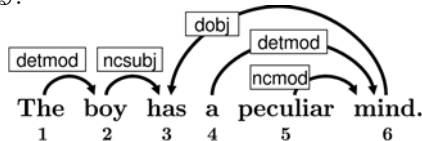


Fig. 1: Typed dependencies in “The boy has a peculiar mind.”

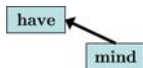


Fig. 2: Dependency structure pattern for “have mind”

同じパターンを持つ文同士をまとめて提示することにより、ユーザは容易にキーワード間の関係を知ることができる。

## 3 依存関係の種類を導入

文献 [1] の手法では、依存関係の種類を考慮した分類ができない。例えば、Fig.1 及び Fig.3 の文と依存構造に対して、クエリ “have mind” を用いてパターンを同定すると、have と mind の間の依存関係に異なるラベルが付与されているにもかかわらず、ともに Fig.2 のパターンが生成されてしまう。

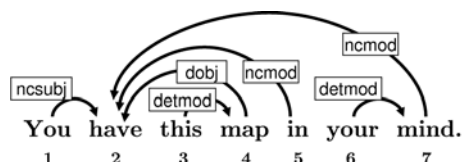


Fig. 3: Typed dependencies in “You have this map in your mind.”

この問題を解決するために、本手法では、依存構造パターンを 5 項組  $d = (h, L, R, D_L, D_R)$  に拡張し、依存関係の種類を区別する。  $D_L, D_R$  は依存関係の種類のリストである。  $D_L$  の  $i$  番目の要素は、  $L$  の  $i$  番目の要素の主辞と  $h$  との依存関係の種類を表す。  $D_R$  についても同様である。

依存構造パターン結合時に、関係名を  $D_L$ 、あるいは  $D_R$  に記録することにより、関係の種類を考慮したパターンを生成できる。

例えば、Fig.1 の依存構造に対しては、have と mind を結合するとき、その間にある依存関係に付与されたラベル dobj を記録し、Fig.4(a) のようなパターンが得られる。 Fig.3 の依存構造に対しては、関係が ncmmod であるようなパターンが得られる。

これにより、[1] の手法で区別できない文も分類できる。

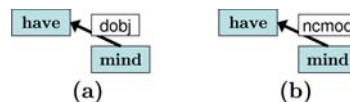


Fig. 4: Typed dependency structure patterns for “have mind”

## 4 実装

本節では提案手法の動作例を示す。検索対象として「The Daily Yomiuri」150,000 文を依存構造解析器のひとつである RASP[2] で解析したものを用いた。

クエリ “have mind” に対して、have (has, had も含む)、mind をともに含む文は 72 文あり、そのうち依存構造パターンが生成されたのは 9 文であった。

文献 [1] の手法では、これらの文に対して Fig.2 のパターンのみが同定され、1 つに分類された。これに対し本手法では、Fig.4 の 2 つのパターンが同定され、2 つに分類された。

分類された文のうち、一方は mind が直接目的語として have に依存する文、もう一方は mind が in を介して have に依存する文であった。依存関係の種類を導入することにより分類がうまくできることが確認できた。

## 5 おわりに

本稿では、依存構造パターンに基づく文の分類において、依存関係の種類を利用して、より細かく分類する手法を提案し、その有効性をシステムを実装して確認した。

文献

- (1) 加藤他: 言語処理学会第 12 回年次大会, pp.668-671, 2006
- (2) Briscoe, E. and Carroll, J: LREC 2002, pp.1499-1504, 2002