概念図と説明テキストの対応付け

1Y－3

藤野 亮之　野久 仁志　黄瀬 浩一　松本 啓之亮
大阪府立大学　工学部

1　はじめに

現在、情報伝達の中心的な役割を担っているのは図とテキストであり、それらを組み合わせて用いている文書は非常に多い。このような文書の情報を全て理解するためには相互の対応関係を考慮した統合的な理解が必要となる。こうした統合的な理解の第一歩として、互いの対応関係の抽出が要求される。

本稿では、図とテキストを用いている文書の一つとして、概念図とその説明テキストを対象とし、対応関係を抽出する方法を提案する。本手法の特徴として、単語に基づいて説明テキストと概念図の対応付けの単位を求めること、説明テキストと概念図の間で単語の抽象度が異なる場合に対処していることがあげられる。

2　概念図と説明テキスト

概念図とは文字列、囲み、矢線の組み合せによって概念と概念の関係を表したものである。概念図中の文字列をラベルと呼ぶ。概念図では概念をラベルと囲みで表現し、概念間の関係を矢線で表現している。概念図において1つの関係とは1本の矢線によって表される関係のことである。

一般に概念図のみで概念間の関係の詳細を述べることは難しいため、多くは説明テキストの助けを借りて詳細な意味表現を行なっている [1]。このことから、概念図と説明テキストには対応関係が存在する。

対応関係についてまず説明テキストを中心に考える。本稿では概念図と説明テキストの対応関係を次の4つに分類する。

I 1つの文が図の1つの関係を説明する場合
II 1つの文が図の複数の関係を説明する場合
III 複数の文が図の1つの関係を説明する場合
IV 複数の文が図の複数の関係を説明する場合

本稿ではこのうちI、IIについて考える。ここで、説明テキストの対応付けの単位を説明の単位と呼ぶ。Iは1文が説明の単位として適当で、そのまま図に対応付く場合である。IIは一文中に説明の単位が複数あるため、一文単位では的確な対応付けが行なえない場合である。この場合は説明の単位ごとに文を分割する必要がある。

次に概念図の立場から考えてみる。概念図は概念間の関係を矢線で繋ぐことで表現しているため、2つの概念間の関係も矢線をたどることで複数の表現ができる。そのため、対応関係を考える際にどの単位で対応付けを行なうかを考える必要がある。また概念図は概略を表すことから、ラベルの単語はその概念を表す単語の総称であることが多い。そのため説明テキストとは抽象度が異なる場合があり、その差を埋めて対応関係をとる方法が必要となる。

以上のようなことをまとめると概念図と説明テキストの対応付けに関する問題は次の3つがある。

1 説明テキストを説明の単位に分割する方法
2 概念図の対応付けの単位を同定する方法
3 ラベルと説明テキストの抽象度の差を埋める方法

これらの対処法を処理の流れに含めて説明を行なう。

3　処理のながれ

本手法では概念図と説明テキストの処理を先に行ない、その結果を用いて対応関係の抽出を行なう。

3.1　概念図の処理

本手法では入力として作図ツール tgif で描画した図を入力とする。概念図からは対応付けの処理に必要となる概念間の関係を抽出する。まず、ラベル c を形態素解析 (juman3.1[2] を利用) により得られる単語 W_k の連接で、

$$c = /W_1/\cdots/W_n/$$

と表す。c の集合を S_f とする。

概念間の関係 R_j はある1本の矢線 j を単位として

$$R_j = \{r_j, s_j, d_j\}$$

と記述する。r_j、s_j、d_j は S_f の要素で、r_j は矢線のラベル、s_j は矢線の元のラベル、d_j は矢線の先のラベルである。この記述で、図1の『認証局』と『消費者』の間の関係を表現すると以下のようになる。

$$R_j = \{/認/証書/, /認証/局/, /消費者/\}$$

3.2　説明テキストの処理

本手法では電子化した説明テキストを入力とする。一般に一文で複数の説明の単位が記述されている場合は説

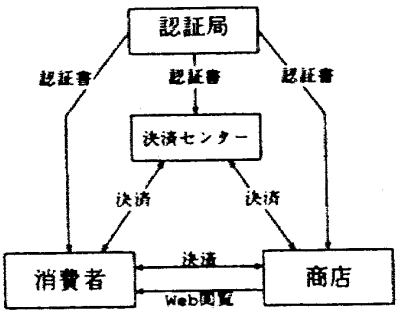

図 1: 概念図の例

Linking Expository Text to Parts of a Diagram
Akinobu Fujino, Hitoshi Nohisa, Koichi Kise
and Keinosuke Matsumoto
College of Engineering, Osaka Prefecture University

明の単位が並列に並べられていることが多い。そこでKNP[2]で構文解析を行ない、一文全体が並列構造でまとめられている場合、その文を分割する。

こうして得られた説明の単位iごとに自立語の連接の集合St_iを得るために形態素解析を行う。この際、対応付けに不必要な単語は不要語としてあらかじめ登録しておき、取り除く。

3.3 対応関係の抽出

対応関係の抽出は概念図と説明テキストの概念のマッチングと対応部分同定の2処理に分けられる。

マッチングとはS_f、St_iを入力とし、以下のようにSt_iに対応する集合M_iを得る処理である。ここでAがBの部分列とは連接B中の1つ以上の連続した単語の組み合わせで連接Aが表現できることである。

$$M_i = \{c | c \in S_f \text{かつ} \exists t \in St_i \ match(c,t)\}$$

$match(c,t)$はtがcの部分列、あるいはcがtの部分列の時に成り立つ。この条件で同じtに複数のcが対応する場合は、最も長く一致したcを選ぶ。

この処理の後、cとtの抽象度が異なる場合について対処するため、M_iに含まれていないcについて意味ネットを用いて処理を行なう。意味ネットとは概念と概念の関係を記述したもので、関係のある概念はリンクと呼ばれる関係でむすばれている。リンクにはその関係によって種類がある。今回使用した意味ネットはEDR(日本電子化辞書研究所)の概念辞書をもとに、約50万概念によって構成されている。この中でcの表す概念から上位下位関係を示すリンクを2本たどり、その過程で求められる概念をすべて取り出す。この概念の名前についてcと同様に連接を求め、マッチングを行なう。得られた概念のうち1つでもマッチングの条件を満たした場合はM_iにcを追加する。

対応部分同定とは図のどの部分に説明の単位iを対応付けるかを求める処理である。この処理で入力とするのはR_jとM_iである。一般に説明テキストは概念の説明もしくは概念間の関係の説明を行なっている。そこで図の対応付けの単位をc、もしくは1つ以上のR_jの組み合わせとする。対応付ける単位は次に示す条件で選択する。

1. $|M_i| = 1$のとき
2. $|M_i| > 1$で次の条件を同時に満たす場合
   - $M_i \subseteq R_{j1} \cup \cdots \cup R_{jm}$
   - mが最小個数になるもの

1の場合はiを$c(\in M_i)$に対応付ける。2の場合はiを取り出された$R_{j1}, \cdots, R_{jm}$に対応付ける。条件を満たす組み合わせが複数ある場合は、$\frac{|M_i|}{|R_{j1} \cup \cdots \cup R_{jm}|}$が最も1に近いものを選ぶ。

4 処理例

情報処理学会の学会誌から得たエレクトリックコマースの解説記事の中から図1、図2に示す部分を対象に処理を行なった。図は簡単化を行なっている。1文目はM_i={/認証/局/,/消費者/,/商店/,/決済/センター/,/認/証書/}となり、対応部分同定の処理で条件2を満たすR_jの組み合わせが選択される。その結果『認証局』から『消費者』『商店』『決済センター』への3つの関係という適切な図の部分に対応付けることができた。2文目は図3に示す構文解析によって一点鎖線の部分で分割され、前半部分は『消費者』－決済－『商店』の部分に、後半部分は『消費者』と『商店』と『決済センター』の間に引かれた決済の関係を表す部分に対応付いた。また意味ネットを利用する場合として、図4に示す説明テキストを用いて処理を行なった。説明テキストでは消費者が客、商店が店と言い換えられているが、図の『消費者』－Web閲覧－『商店』の部分に対応付いた。

認証局はＥＣの参加者（この場合は消費者、商店、決済センター）の社会的、経済的な信用確認をあらかじめ行なって、認証書を発行しておく。
決算処理は決算の種類によって流れが異なるが、例えば、ＭＯＮＤＥＸのような電子現金の場合、消費者と商店間の二者間決済となり、
クレジットカード決済であるＳＥＴプロトコルでは消費者と商店と決済センターの
三者間決済となる。

図 2: 説明テキストの例

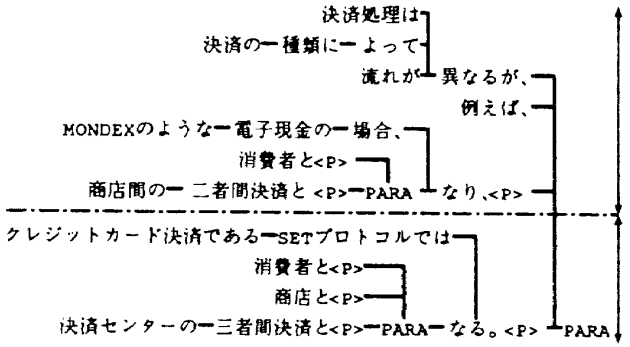

図 3: 構文解析結果

客はインターネット経由で店を訪れ、Ｗｅｂで商品情報の閲覧を行なう。

図 4: 意味ネットの有効性を調べるための例文

5 おわりに

本稿では、概念図とその説明テキストに対して、対応関係を抽出する方法を提案した。今後の課題として、図に対する説明テキストを文書中から自動的に切り出す方法の検討などがある。

謝辞 本研究は立石科学技術振興財団の補助による。

参考文献

[1] 中村裕一、古川亮: "概念図理解を目的としたパターン情報と自然言語情報の統合"、情処学論、Vol.36、No.1、pp.196-205(1995)。

[2] http://www-lab25.kuee.kyoto-u.ac.jp/