

文書構造と図表情報の関わりに着目した文書処理について

山口仁司, 愛智宏輔, 永井秀利, 野村浩郷
九州工業大学

文章の要約結果として図表表現を生成する文書処理方式について述べる。文章中の各文を拡張格構造解析し、その解析結果を拡張格構造として表現する。次にそれぞれの拡張格構造表現中で共通な格要素を結合することによって、文章全体を表す意味ネットワークを構成する。この意味ネットワーク上での結合の多さを、格要素の重要度とする。最後に、ある重要度以上の結合を持つ格要素およびその格要素が格関係を持つ述語によって構成される拡張格構造を要約結果として抽出する。このような手順で生成された拡張格構造格関係を、図表表現の部品である図形要素に変換することによって、図表表現による文章の要約を得る。

Document Processing in terms of Cooperative Relationships between Document Structure and Figure/Table

Hitoshi YAMAGUCHI, Kosuke AICHI, Hidetoshi NAGAI and Hirosato NOMURA
Department of artificial Intelligence Kyushu Institute of Technology
Iizuka, 820, Japan
nomura@ai.kyutech.ac.jp

This paper presents a method for generating a figure/table style summarization of a text. First of all, the system analyzes each sentence and then produce an extended case structure representation. Second of all, it combines common case elements in the extended case structures by which it produces a semantic network for the whole text. An evaluation criterion defined by the number of such connections is applied to estimate the contribution of each component in terms of summary. Finally, such selected components and a related predicate figure out an extended case structure which is to be the summary of the text. Such a representation is furtherly converted into a figure/table form by replacing each relationship by a figure/table representative clue.

1 はじめに

本研究は、文章内容を構成する概念の中で主要なものとの相対関係を求め、それから図表表現を生成することにより文章の要約を行う方法について述べる。

図表を使った表現は、論文やマニュアルに対象を限れば多用されている。これは文章表現に対して図および表の使用が、著者から読者への効率良い情報伝達を可能とする特徴があるためである。図表表現もまた情報交換の手段の一つであり言語の表現の一つの形態と考えることができる。従って、図や表を扱った文書処理の研究も必要である。

本研究では、この図表表現の高い情報伝達性に注目し、文章表現では読者に必要とされていた「読む」という作業を省力化するために、「自然言語の文章」から「図表表現」への変換を行い、その結果生成された図表をもって文章の要約とする文書処理方式を考える。また、本研究での対象とする文書は論文あるいはマニュアルに限定するものとする。

計算機による要約処理には従来から様々なものが提案されているが、基本的には、1) 何らかの方法で文章を解析し、2) 解析した結果を評価して重要な単語、文、箇所、意味表現などを決定し、3) その結果を要約として文や単語、あるいは意味表現などから生成された文章や表形式で出力する、という処理が行われる。1) の解析方法は表層的解析によるものと深層的解析によるものに大別される。前者は単語の出現頻度やキーワード、文章中の位置などによる解析であり、後者はスクリプトや物語文法、何らかの意味表現などを文章から得る構文的・意味的解析である。

本稿では、まず実際の図表が文章とどのような意味的関わりを持つかをまとめ、次に文章から要約として抽出される重要な概念を図表表示する要約方式について述べる。

2 図表表現と文章の関わり

まず、図という表現について見てみると、本研究で処理対象としてとりあげる論文等以外にも、様々な文書で様々な形式の図が記述されている。例えば、物語に挿入されるさし絵なども、図の一種とみることもできる。しかし本研究では、図を「ノードとアークによって構成されるグラフ」に限定する。この定義は数学におけるグラフが元になっている。しかし、本研究における図は実際の文書中の図表現を取り扱う為に、ノードやアークの拡張及び数学的なグラフには無い組み合わせによって構成された「図形要素」という単位の組み合わせによって構成されるとする。

図としてとりあげるものをグラフ的表現がなされたものだけに限ることについては、本研究では要約処理の対象が論文、マニュアルに限定されている為に、ノードやアークの拡張や組み合わせで対処できるものが多いと

の判断による。どのような図形要素を図表現の単位とするかについては、実際の論文やマニュアル中に出現する図表現から適当に設定するものとする。

また、表を使った表現は図の場合とは大きく違う特徴を持っている。表は、列と行で構成される。各行はそれを構成する項目によって、一般に“be”を述語として持つ格構造であるとすることができ、列についても各項目は何らかの基準で特定されたクラスを構成している。すなわち、表々現は、項目を列挙するような内容に限定された表現といえることができる。このように、表は、その形式から表現できる内容が図よりも限定されているという特徴がみられるが、本研究では処理の際の簡単化の為に、図の場合と同様の扱いをする。すなわち図形要素として、行列の構造を持つようなものを考えればよい。

文書中に表れる図表表現の性質を考察する際に、その図表を単独に解析するだけでは図表表現が文書中でのどのような働きを持つか、すなわち図表表現と文章表現との関わりは明らかににならない。一般に、文書中に図表表現が使われるときには、図表の前または後、あるいはその両方に、構文的意味的関係を持つ文章が置かれる。

文書中の図や表と、その図表の前後の文章との関連性について、ここでは、

- (1) 図表のもつ概念と文章のもつ概念とに共通、または類似した概念がある
- (2) 前後の文章中に図表表現を指す語がある
- (3) 特に構文的意味的関連性を示すものはない

の三種類を考える。

(1) は、図表とその前後の文章の意味内容を重複させることにより、著者が読者の理解の促進をより努めた、冗長ではあるが親切な文書といえることができる。(2) は、文章の中に「…を図××に示す」や「…は、表××のように分類される」など、その語の前後に記述された図表を指示するという意味で指示詞の範疇に含まれる語が使われているのが特徴である。この指示詞と指示されている図表との間には、明示されたラベルや文書中に記述される順序があり、照応には曖昧性はない。また(3) は、(1)や(2)のように図表と前後の文章との間に共通の語や概念があったり、文中に図表を指し示す語があるのではなく、単に文脈の流れによってのみ図表と前後の文章とが関わっている。この種類の図表表現は、特にマニュアルで用いられる傾向がある。

これらの三種類の図表のうち(1)のタイプは本研究において、文章から要約として出力された図表の自然さを検証するのに用いることができる。文章表現と重複する意味概念を持つ図表表現が、元の文章表現の意味内容のなかでも重要な部分を占めているとすれば、文章の要約処理を行いその結果を図表表現した後に、相当する図表表現との相違を比較することによって要約の精度を確かめることができる。

3 要約方式

3.1 処理の流れ

前節では、実際の図表表現がその前後の文章とどのような関わりを持つか、及び概念の重複のある文書は要約処理に有用なことを述べた。そこで本節では、最終的な処理結果が図表表示される要約の処理方式について述べ、例をあげてこの要約処理がどのように働くかを示す。

ここで述べる要約処理の概要は以下の通りである。

1. 一文単位の格解析
2. 文章の意味ネットワーク化
3. 重要度評価
4. 要約図形の生成

要約対象文章の入力は、代名詞・指示詞などの省略や照応を補った一文単位の拡張格解析による格構造(拡張格構造)とする。拡張格構造はまた、単純な文単位の格構造に加えて句レベルの並列構造、文レベルの接続構造を加えた複合文構造や名詞句修飾構造を含んでいる。名詞句修飾構造には複合名詞句構造、付加詞句構造、埋め込み文構造が含まれる。

この格構造を、同一な概念を指す格要素同士で次節で述べる条件により結合して文章全体を意味ネットワークで表現する。次に、この意味ネットワーク上で重要度評価を行なって意味ネットワークを縮約することで要約処理を行なう。最後に、要約された結果残った意味ネットワークから図形表現を生成して要約結果とする。

3.2 格解析と格構造の結合

拡張格構造で表現された一文単位の概念構造から、文章全体での概念間の関係を明らかにするために、同一な概念を表す格要素又は格要素の一部を結合し、それぞれの格構造を意味ネットワークとして統合することで文章全体の解析を行なう。

格要素の「結合」とは、次のような操作を指す。例えば、ある文 S_1 で概念 C_1 と C_2 の関係を述べており、また別のある文 S_2 で概念 C_2 と C_3 の関係を述べているとき、 S_1 と S_2 に共通した概念である C_2 の結合によって、別個の格構造であった S_1 と S_2 が一つの意味ネットワークとなり、概念 C_1, C_2, C_3 が関係付けられる(図 1)。以後、結合とはこのように異なる格構造中で共通する格要素同士を結びつけ、その結果格構造がネットワークとなる操作を指すものとする。また、「弱い結合」というものも考える。これは名詞句修飾構造などで、同一の格構造中の異なる概念の結合であり、上で述べた「異なる格構造の共通の格要素」と全く違う。弱い結合は概念の重要さの度合いに影響する。

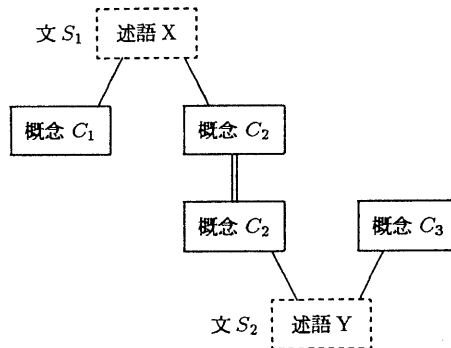


図 1: 格要素の結合

以下に、結合を行うときの条件を述べる。ただし、ここでは格標識については除外して、対象とする格要素だけは体言句のみから構成された格要素だけとする。つまり、副詞句とその他の連用修飾句については考えない。

A. 埋め込み文構造を持たない格要素同士の場合

1. それぞれの全体が全く同一ならば、同一の概念を指すとして結合する。
(例)「翻訳対象」と「翻訳対象」
2. 上記の 1 でなく、両者の関係がパラフレーズである場合、同一の概念を指すとして結合する。
(例)「翻訳対象」と「翻訳の対象」

この場合のパラフレーズとは、一方の付加詞句構造の体言句と体言句を結ぶ「の」を削除し複合名詞句構造に変換したときに他方の複合名詞句に同一であればパラフレーズであると判断し、結合をする。

3. 上記の 1, 2 でなく、一方の格要素だけが付加詞句構造を持つ格要素である場合
(例)「翻訳」と「翻訳の対象」の場合、
「翻訳の対象」→「翻訳」+「対象」として、それぞれの体言句と同一かどうか判断する。

この場合のみ特別に付加詞句構造を構成する体言句をそれぞれ一つの格要素と同等に扱い、名詞句修飾構造を持たないもう一方の格要素と同一かどうか判断し、同一であれば弱い結合をする。

この場合は、付加詞句構造を持つ格要素全体で見ると完全に同一の概念を表すとは言えない場合が多い。しかし、付加詞句構造を構成する体言句自体では同一の概念を指しているのので、何らかの関係があるとして弱い結合をする。

B. 埋め込み文構造を持つ格要素と埋め込み文構造を持たない格要素の場合、埋め込み文は一つの文として格要素の判断を行う。この条件では、埋め込み文構造を持た

ない格要素と埋め込み文が修飾する被修飾句とが同一な場合、

(例)「計算機で翻訳した対象言語」と「対象言語」

と、埋め込み文構造を持たない格要素が埋め込み文の被修飾句でない格要素と同一な場合、

(例)「計算機で翻訳した対象言語」と「計算機」

の二つの場合が考えられるが、いずれも共通した格要素は結合される。

さらに、埋め込み文は被修飾句で表された格要素と埋め込み文の中の格要素との関係を表しており、被修飾句と埋め込み文のそれぞれの格要素とが、弱い結合をする。例では埋め込み文が被修飾句の「対象言語」と埋め込み文「計算機で翻訳した」の中の格要素「計算機」とが弱い結合をする。

C. 埋め込み文構造を持つ格要素どうしの場合

1. 全体が全く同一ならば、同一の概念を指すとして結合する

(例)「計算機で翻訳した対象言語」と「計算機で翻訳した対象言語」

2. 1. でない場合はそれぞれの被修飾句を格要素と考えて以下の条件で分岐する。

(a) それぞれの被修飾句が同一である場合

(例)「計算機で翻訳した対象言語」と「人間が翻訳した対象言語」

(b) (a) でなく、一方の被修飾句ともう一方の埋め込み文の格要素が同一の場合

(例)「計算機で翻訳した対象言語」と「我々が使用した計算機」の場合

一般に名詞句修飾構造は、その被修飾句の意味を限定する。(a)の条件の場合でも、被修飾句が「もの」や「こと」などの形式名詞の場合は、埋め込み文の表す意味内容により「もの」や「こと」など自体が表す意味内容が決められる。従って被修飾句がこのような種類の体言句の時は、埋め込み文が異なればその被修飾句自体の意味も異なるので、結合しない。このような体言句を修飾句依存型の体言句と呼ぶことにする。

もし被修飾句が修飾句依存型の体言句でなければ、埋め込み文によりその意味を限定あるいは説明されるが、被修飾句自体の意味は同一であると考えられるので、被修飾句どうしを結合する。(a)の例ではそれぞれの被修飾句「対象言語」を結合する。

一方(b)の場合は、条件Bと同様に埋め込み文の中の格要素と被修飾句の格要素とが共通しているものは結合する。例では、埋め込み文「計算機で翻訳した」の中の格要素「計算機」と、埋め込み文構造の格要素「我々が使用した計算機」の被修飾句「計算機」とは同一なので結合する。

また条件Bと同様、埋め込み文のそれぞれの格要素とその埋め込み文の被修飾句とは弱い結合をする。例では埋め込み文が被修飾句の「対象言語」と埋め込み文「計算機で翻訳した」の中の格要素「計算機」とが弱い結合をする。(a),(b)のところで示された四つの例ではそれぞれ「計算機」と「対象言語」、「人間」と「対象言語」、「計算機」と「対象言語」、「我々」と「計算機」が弱い結合をする。

D. 句レベルの並列構造を持つ場合

並列構造を構成している体言句ごとに一つの格要素と見て、同一な格要素があれば、並列構造を構成するそれぞれの体言句を一つの格要素として格構造と結合する。

以上のA,B,C,Dで示された四つの条件によって、それぞれの文について結合や弱い結合を行う。その結果、各文の拡張格構造から文章全体の意味ネットワークが形成できる。この意味ネットワークは、結合によって関係付けられたいくつかの概念群を表わしている。ただし弱い結合は、概念間の関係より概念自体の重要さの度合いに影響する。意味ネットワーク上で結合を持つ格要素を、重要概念と定義する。

3.3 重要度評価

前節で述べた結合という操作によって、文章中の各概念が意味ネットワークで表現される。次に、その意味ネットワーク上で要約を行なう方法について述べる。この手順の方針は、重要概念ごとの重要さの度合いを示す「重要度」という尺度を評価していくことで要約を行なう。重要度評価の手順の概要は次のとおりである。

1. 文単位の評価
2. 格関係単位の評価
3. 要約の生成

重要度は基本的に、ある格要素の結合の数である。ただし、弱い結合を行なう名詞句修飾構造を持つ格要素についてはこの限りではない。

3.3.1 文単位の評価

まず文単位での評価をする。この評価を行なう基準は、異なる二つ以上の重要概念間の関係を表している可

能性のある文だけに絞るものとする。つまり、ひとつの文中で二種類以上の重要な概念を持つものだけに絞る。

次のように文単位で重要概念の数を数えていく。

- 文中に他の文の格要素と結合している格要素を一つ持てば重要概念数を1加える
- ただし、弱い結合の場合の、付加詞句構造を持っている格要素は重要概念ではないのでそのような格要素を持つ文には重要概念数は加えない
- 同一文内での結合の場合は、重要概念数は1しか数えない。

このように文単位の重要概念数を数えていき、重要概念数が2未満の文は削除する。そして、ある概念とまた別のある概念との関係を表す可能性のある文だけに絞る。重要概念の数が0の文とは、その文中の格要素が全て他の文の格要素とは異なる、いわば概念的に独立した文であり、重要概念の数が1の場合は、意味ネットワーク上で見るとただ一つの結合によってその文以外の意味ネットワークと接続されているような文である。文単位の評価は、これらの重要概念との関連の少ない文を意味ネットワークから削除する操作である。

3.3.2 格関係単位の評価

次に、ある述語とその述語に直接格関係を持つ格要素(ただし埋め込み文を含んでも良い)を単位として重要度を評価する。すなわち重要な概念間との関係を表す格構造だけを文単位の評価で残った意味ネットワークから抜き出す。

具体的には、文単位の評価と同様の数え方により各重要概念の出現数を調べる。 n 個の文で出現するある重要概念の重要度は、 n になる。このとき、一つの格構造内に重要度2以上の重要概念を含む格構造を抜き出す。

この操作は、一つの述語を介して接続される二つ以上の重要概念をもつ格関係を残し、それ以外の格要素は削除する。

重要概念の中に埋め込み文構造があって、その埋め込み文構造の被修飾句とだけ結合し、かつその埋め込み文が他の重要概念との関係を表していない場合には、埋め込み文を削除する。

3.3.3 要約の生成

最後に縮約された意味ネットワーク上の各格要素を、重要度によって要約として抽出するか削除するかを決定する。ある格要素の重要度は、その格要素の文章中での出現頻度であるから、重要度の高い格要素ほど要約として抽出すべき重要概念であると判断できる。全体の中でもっとも重要である概念、すなわち直接に関連する重要概念が最も多い概念を最重要概念と呼ぶ。

最重要概念の重要度を k としたとき、縮約された意味ネットワーク上には重要度2から k までの格要素が残っている。このうち、どの重要度までを要約として抽出するか決定しなければならない。この判断の基準は主観的なところもあり難しい。ある重要度 n までで要約を行うとしたときに、要約結果として抽出される概念の集合「重要度 n の集合」は次のように定められる。

縮約された意味ネットワークにおいて重要度 n から重要度 k までの重要概念と、それらの重要概念と一つの述語を介して格関係を持つ概念の集合

n の値は、要約処理の対象としている文書や要約として認められる概念の多さなど状況によって変化できる処理が望ましいと考えられる。

以上のような手順で要約を行なう。

3.4 要約例

以上に説明した要約手順で、次のような例について処理を行った経過を示す。例文には説明のために行番号をふってある。

- 1: 文脈がとらえられていないために談話としての意味や発話の意味内容が適切に捉えられないことがある。
- 2: 談話は発話の集合である。
- 3: ただし、それは乱雑に集めた塊でなく、何らかの規約、規則によって統一性を備えたものである。
- 4: 勝手に一文を取り除いたり、付け加えたりするとたちまち談話としての統一性を失ってしまう。
- 5: この統一性を保たせるもっとも大きな要因が文脈であった。
- 6: この統一性は発話列間の言語形式的連鎖に依存する連続性とそれを規約とする意味の首尾一貫性から成る。
- 7: その他修辭法の各種原理、さらに主張の意図などを含む談話の相対的意義を決める目的性要因がある。
- 8: 文脈は大きく分けて 1) 先行発話列、後続発話列に明示的に現れる談話の脈絡に依存する言的要因ないし形式的脈絡、2) 談話文中に明示的に現れなくて隠示的背景知識に大きく依存する外示的要因、3) 社会的習慣、社会的規約に大きく依存する社会的要因、さらに 4) 話者・筆者と、聞き手の意図に大きく依存する心理的要因から成る。
- 9: さらに、5) 談話全体の存在自身を規定する目的性要因などが考えられる。
- 10: 言語的要因は文の談話における言語的連続性に大きくかわり、その他の外示的要因、社会的要因、心理的要因、目的性は文の意味的首尾一貫性に大きく関わっている。
- 11: 後者の群は言語形式外の文脈構成要因という意味で言語外要因と呼べる。

3.4.1 格解析

この例の文章を省略や照応を補って格解析したものを入力とする。

↓概念→文番号	1	2	3	4	5	6	7	8	9	10	11
文脈	1								1	1	
談話	1	1		-	-	-	-	-			
発話の集合		1	1								
統一性			1	-	-	-					
談話としての統一性				1	1	1					
合目的性要因							1		1		
言語的要因								1		1	
外示的要因								1		-	-
社会的要因								1		-	-
心理的要因								1		-	-
その他の外示的要因								-		1	1
その他の社会的要因								-		1	1
その他の心理的要因								-		1	1
重要概念数	2	2	2	1	1	1	1	5	2	4	3

表 1: 重要概念と文単位評価テーブル

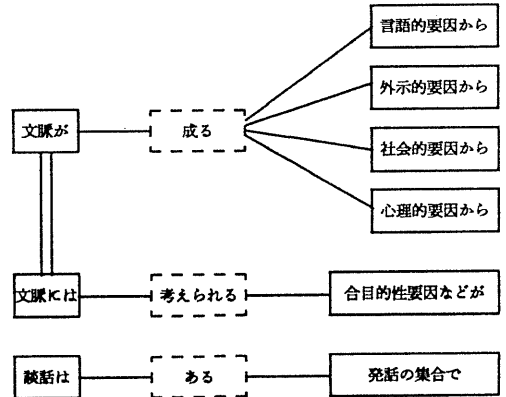


図 2: 格構造単位の選択結果

3.4.2 格構造の結合

第 3.2 節で述べた結合条件を適用して格構造の結合を行ない、文章全体を意味ネットワーク化する。この結果は省略する。

3.4.3 文単位の評価

以下に、結合した概念すなわち重要概念と、その概念が書かれている文を表したものを表 1 に示す。格構造の結合により得られた意味ネットワーク上で他の格構造と結合が行われた概念が重要概念である。重要概念は、表 1 の一番左の列に列挙されている。

表の一番左の列は重要概念であり、最後の行は文単位の重要度である。各行はその重要概念が書かれている文に重要概念数を記している。表中に「1」、「-」の書かれているところは、その文に重要概念が書かれていることを表す。「1」は、重要度を加えるべき格要素を持つ文を示し、「-」は弱い結合をしている場合の付加詞句構造を持つ格要素のある文を示している。

続いて文単位の評価をする。表 1 の重要概念数が 2 未満の文は削除する。すなわち、文番号 4~7 の文をここで削除することにより、ある重要概念とある重要概念の関係を表す可能性のある文だけを選択する。選択された残りの文については、次の述語単位の重要度評価でさらに評価する。

3.4.4 格関係単位の選択

文単位の評価により選択された文番号 1~3, 8~11 の文のうち、重要概念が一つの述語を介して結ばれているものだけに絞る。その結果のネットワーク表示を図 2 に示す。

述語は点線で囲んで、法要素も含めて記述してある。格要素は実線で囲んで、格標識も含めて記述してある。

↓概念→文番号	1	2	3	8	9	10	11	重要度
文脈	1			1	1			3
談話	1	1		-		-		2
発話の集合		1	1					2
統一性			1					1
談話としての統一性			-					0
合目的性要因					1			1
言語的要因				1		1		2
外示的要因				1		-	-	1
社会的要因				1		-	-	1
心理的要因				1		-	-	1
その他の外示的要因				-		1	1	2
その他の社会的要因				-		1	1	2
その他の心理的要因				-		1	1	2

表 2: 重要概念と格関係単位評価テーブル

格要素「文脈が」と「文脈には」は格標識を除けば同一概念をさす。よって、この二つの格要素は結合されている。二本線は結合を表す。

3.4.5 要約の生成

図 2 から概念の選択によって要約の生成を行なう。格要素の概念重要度は、「文脈」の重要度が 3 である。よって最重要概念は「文脈」である。表 reftbl:case に、それぞれの重要概念の重要度を示す。ただし、既に文単位の評価及び格構造単位の評価で、削除された文については表記していないが、削除された重要概念についても重要度を示しており、これらは要約の対象とはならない。

要約として出力する概念の範囲は、

- 重要度 3 (2 でも同様) の集合

述語「成る」と「考えられる」の各々の格構造を「文脈」という概念を表す格要素どうしを介して結合したもの

- 重要度 2 の集合

重要度 3 の集合に加え、その他の述語「ある」の格構造

以上が要約処理によって得られる意味ネットワークである。要約として出力する重要度はこの場合、2 と 3 の 2 段階あるが、本稿では範囲の決定方法については考察しない。

以上、要約の大まかな処理の流れと、その具体的方法について述べた。

4 図形表現の生成

前節の要約処理により得られるのは要約として選択された格関係による意味ネットワークである。本節ではこれを位置関係に置き換えることで図形表現に変換する方法について述べる。図形表現の生成の過程の概要は次のようになる。

1. 意味ネットワークから図形要素への変換
2. 図形要素から図表表現の生成

4.1 意味ネットワークから図形要素への変換

前節の要約処理により得られる意味ネットワークから、図形に対応した文字記述表現である図形要素へ変換するための知識を用意しておく。例えば、次のような変換規則を考えることができる。

構成する (主題格, 構成素格)

→ PART-OF (主題格, 構成素格)

これは、「成る」という述語によって結び付けられた主題格と構成素格という二つの格があったとき、それから「PART-OF」という図形要素を生成するという規則である。

4.2 図形要素の例

上の変換規則のように図表要素が述語に従って変換されるものであることから、図形要素とは、概念間の関係を位置関係で表す述語の分類であるということが出来る。本稿では上で示した例に沿って、次のような図形要素を考える。

PART-OF 全体部分関係を表す

ARROW 矢印で結合され、二つの概念間に時間的な推移や空間的な上下関係があることを表す

LINE 線で結合され、二つの概念間の一般的な、あるいは弱い関連性を表す

EQUAL 二重線で結合され、二つの概念間の並行的な、あるいは強い関係を表す

また、述語から図表への変換規則も次のようなもの考える。

成る (主題格, 構成素格)

→ ARROW (主題格, 構成素格)

考えられる (主題格, 構成素格)

→ ARROW (主題格, 構成素格)

ある (主題格, 陳述対象格)

→ LINE (主題格, 陳述対象格)

これらの図形要素は例示のために非常に簡略化されており、第 2 節で述べた、表をあらわす図形要素も考慮していない。また、変換規則についても、「成る」という述語から矢印を生成することも、あるいはまた集合のベン図のような表現を生成することも任意である。すなわち、変換規則は図表の多様な表現の自由度をなくしてしまい、ある述語からは特定の図表表現の生成しか許さない。

また、図形要素から図形を描くときのレイアウトについても本稿では、左右対称や右、左詰め等、あるいはノードやアークの形態についても基本的な部分以外は細かくは規定しない。従って例えば同一のノードから幾つかの ARROW 図形要素が描かれる際に、

- 矢印をそれぞれ別個に書く
- 矢印を途中までは一本にまとめる
(この場合、木構造の根と葉が明らかになるため、アークは矢印のかわりに単純な線として描かれてもよい)

という区別ができ、さらに ARROW が集まった木構造であることが明らかになった場合に、その根を上下左右どのように配置するかという問題は、本稿ではふれない。

4.3 生成例

第 3.4 節で要約した結果から上であげた変換規則を用いて図形を生成すると図 3 のようになる。

図 3 中、破線で囲まれている部分を含めると、重要度 4 の集合を要約として、図を生成したときの結果となる。含めない場合は重要度 1 の集合から生成された最も簡潔な要約を表すことになる。

5 おわりに

文章中の概念および概念間相互の関係を文章から抽出し、その関係を位置関係に置き換えて図形表現を生成することにより、論文やマニュアル等に限定された文書で要約を行なう方法を示した。要約処理の例として図を生成する場合について述べたが、表の場合は述語によって生成される変換規則だけでなく、列方向の列挙を名詞句や文の並列構造から抽出することで図形要素の形の記述が可能と思われる。

本稿であげた図形要素を、多くの論文やマニュアルで使用されている図表表現の記述にたえられるように拡充することが一つの課題であるが、それと同時にどのよ