(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2011-215898
(P2011-215898A)
(43)公開日　平成23年10月27日(2011.10.27)

(51)Int.Cl.		F I			テーマコード（参考）
G06F 19/00	(2011.01)	G06F	19/00	130	5B075
G06F 17/30	(2006.01)	G06F	17/30	180C	5B091
G06F 17/27	(2006.01)	G06F	17/27	Z	

審査請求 未請求　請求項の数 9　ＯＬ　（全 14 頁）

(21)出願番号　特願2010-83666(P2010-83666)
(22)出願日　平成22年3月31日(2010.3.31)

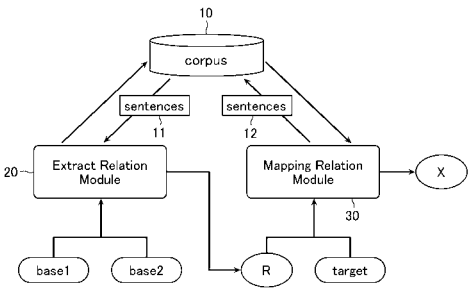
特許法第３０条第１項適用申請有り　平成２２年３月１９日　日本知能情報ファジィ学会発行の「第３５回ファジィ・ワークショップ講演論文集」において発表

(71)出願人　801000027
学校法人明治大学
東京都千代田区神田駿河台１－１
(74)代理人　100092820
弁理士　伊丹　勝
(72)発明者　高木　友博
神奈川県川崎市多摩区東三田１－１－１
明治大学生田校舎内

Ｆターム(参考)　5B075 ND03　QT05
5B091 AA15　AB17　CA05　CA12

(54)【発明の名称】類推方法、類推システム及び類推プログラム

(57)【要約】
【課題】構造写像理論に基づく類推方式によって精度良くある程度正しい解を得る。
【解決手段】類推システムは、類推に用いられる知識情報が蓄積されたコーパス１０と、写像対象となるベース１，２との関係Ｒを抽出する関係抽出モジュール２０と、抽出された関係Ｒをターゲットに写像する関係写像モジュール３０とを備える。そして、コーパス１０からベース１，２が同時に出現する文を抽出し、抽出された文から関係Ｒを表す単語r_iを抽出する。また、ターゲットに関係Ｒを写像して、ターゲットと単語r_iとが同時に出現する文をコーパス１０から抽出し、抽出された文から関係Ｒに基づく解Ｘの候補となる単語x_jを抽出することを、全ての単語r_iについて行う。そして、算出された所属度ｇｒａｄｅＸ（x_j）の値が高い所定数の単語x_jを解Ｘに含まれるターゲットに関係する候補語として抽出する。
【選択図】図１



【特許請求の範囲】
【請求項１】
複数の基底語（ｂａｓｅ１，ｂａｓｅ２）の間の関係Ｒから、目標語（ｔａｒｇｅｔ）との間で関係Ｒにある解Ｘを類推する構造写像理論に基づく類推方法であって、
類推に用いられる知識情報として形態素解析された複数の文が蓄積されたコーパスから、前記複数の基底語が同時に出現する文を抽出する第１ステップと、
前記抽出された文に含まれる単語から前記複数の基底語の間の関係Ｒを表す単語r_iを抽出する第２ステップと、
前記抽出された単語r_iについて、前記関係Ｒへの所属度ｇｒａｄｅＲ（r_i）を算出する第３ステップと、
前記目標語と前記単語r_iとが同時に出現する文を前記コーパスから抽出する第４ステップと、
前記抽出された文に含まれる単語の中から、前記目標語との間で関係Ｒにある単語x_jを抽出する第５ステップと、
前記第４ステップ及び前記第５ステップを前記抽出された全ての単語r_iについて行い、これにより抽出された全ての単語x_jに対して前記解Ｘへの所属度ｇｒａｄｅＸ（x_j）を算出する第６ステップと、
前記算出された所属度ｇｒａｄｅＸ（x_j）の値が高い所定数の単語x_jを前記解Ｘに含まれる前記目標語に関係する候補語として抽出する第７ステップと
を備えた
ことを特徴とする類推方法。
【請求項２】
前記第４ステップでは、前記所属度ｇｒａｄｅＲ（r_i）の値が所定値よりも高い単語r_iと前記目標語とが同時に出現する文のみを抽出する
ことを特徴とする請求項１記載の類推方法。
【請求項３】
前記第５ステップでは、前記単語x_jは前記複数の基底語及び前記単語r_iの前記文における記載順序に基づき抽出される
ことを特徴とする請求項１又は２記載の類推方法。
【請求項４】
前記関係Ｒは、この関係Ｒを構成する単語をr_iとし、メンバーシップ関数の値をｇｒａｄｅ（r_i）とした場合、

$$R=\{grade(r_i)/r_i\}\quad(i=1,2,\cdots,n)$$

として表される概念ファジィ集合であり、
前記解Ｘは、この解Ｘを構成する単語をx_jとし、メンバーシップ関数の値をｇｒａｄｅ（x_j）とした場合、

$$X=\{grade(x_j)/x_j\}\quad(j=1,2,\cdots,n)$$

として表される概念ファジィ集合である
ことを特徴とする請求項１～３のいずれか１項記載の類推方法。
【請求項５】
前記所属度ｇｒａｄｅＲ（r_i）の値は、
ｃｏｕｎｔ（r_i）を関係Ｒとして抽出された単語r_iに対する頻度、Ｎ（ｂａｓｅ１∩ｂａｓｅ２）を複数の基底語が同時に出現した文の数、Ｎ（r_i）をコーパスに含まれる文中に単語r_iが現れる文の数、Ｎｄｏｃをコーパスに含まれる全ての文の数とした場合、

$$gradeR(r_i)=\alpha\cdot\log\left(1+\frac{\beta\cdot Ndoc}{N(base1\cap base2)\cdot N(r_i)}\right)$$

として表される（ただしα、βは調整可能なパラメータ）
ことを特徴とする請求項１～４のいずれか１項記載の類推方法。
【請求項６】
前記所属度ｇｒａｄｅＸ（ｘ$_{j}$）の値は、
ｇｒａｄｅ（ｘ$_{j}$｜ｒ$_{i}$）を単語ｒ$_{i}$が目標語に写像されたときに単語ｘ$_{j}$が解Ｘとして正しいかを示す指数、Ｎｒを目標語に代表される単語ｒ$_{i}$の数とした場合、

$$gradeX\ (x_j)=\sum_{i=1}^{Nr} grade\ (x_j \mid r_i)$$

として表される
ことを特徴とする請求項１～５のいずれか１項記載の類推方法。
【請求項７】
前記ｇｒａｄｅＸ（ｘ$_{j}$｜ｒ$_{i}$）の値は、
ｃｏｕｎｔ（ｘ$_{j}$｜ｒ$_{i}$）を単語ｒ$_{i}$が目標語に写像されたときの解Ｘの候補として抽出される頻度、Ｎ（ｒ$_{i}$∩ｔａｒｇｅｔ）をコーパスに含まれる単語ｒ$_{i}$と目標語とが同時に出現した文の数、Ｎ（ｘ$_{j}$）をコーパスに含まれる文中に単語ｘ$_{j}$が現れる文の数、Ｎｄｏｃをコーパスに含まれる全ての文の数とした場合、

$$gradeX(x_j \mid r_i)=\gamma\cdot\log\left(1+\frac{\delta\cdot Ndoc}{N(r_i \cap \mathrm{target})\cdot N(x_j)}\right)\cdot gradeR(r_i)$$

として表される（ただしγ、δは調整可能なパラメータ）
ことを特徴とする請求項６記載の類推方法。
【請求項８】
複数の基底語（ｂａｓｅ１，ｂａｓｅ２）の間の関係Ｒから、目標語（ｔａｒｇｅｔ）との間で関係Ｒにある解Ｘを類推する構造写像理論に基づく類推方式を用いた類推システムであって、
類推に用いられる知識情報として形態素解析された複数の文をコーパスとして蓄積する蓄積手段と、
前記コーパスから前記複数の基底語が同時に出現する文を抽出すると共に、抽出された文に含まれる単語から前記複数の基底語の間の関係Ｒを表す単語ｒ$_{i}$を抽出する第１抽出手段と、
前記抽出された単語ｒ$_{i}$について、前記関係Ｒへの所属度ｇｒａｄｅＲ（ｒ$_{i}$）を算出する第１算出手段と、
前記目標語と前記単語ｒ$_{i}$とが同時に出現する文を前記コーパスから抽出すると共に、抽出された文に含まれる単語の中から、前記目標語との間で関係Ｒにある単語ｘ$_{j}$を抽出することを、前記抽出された全ての単語ｒ$_{i}$について行う第２抽出手段と、
前記抽出された全ての単語ｘ$_{j}$に対して前記解Ｘへの所属度ｇｒａｄｅＸ（ｘ$_{j}$）を算出する第２算出手段と、
前記算出された所属度ｇｒａｄｅＸ（ｘ$_{j}$）の値が高い所定数の単語ｘ$_{j}$を前記解Ｘに含まれる前記目標語に関係する候補語として抽出する第３抽出手段と
を備えたことを特徴とする類推システム。
【請求項９】
複数の基底語（ｂａｓｅ１，ｂａｓｅ２）の間の関係Ｒから、目標語（ｔａｒｇｅｔ）との間で関係Ｒにある解Ｘを類推する構造写像理論に基づく類推方式をコンピュータに実行させる類推プログラムであって、
コンピュータに、
類推に用いられる知識情報として形態素解析された複数の文が蓄積されたコーパスから、前記複数の基底語が同時に出現する文を抽出させる第１処理と、
前記抽出された文に含まれる単語から前記複数の基底語の間の関係Ｒを表す単語ｒ$_{i}$を抽出させる第２処理と、

前記抽出された単語r_iについて、前記関係Ｒへの所属度ｇｒａｄｅＲ（r_i）を算出させる第３処理と、
前記目標語と前記単語r_iとが同時に出現する文を前記コーパスから抽出させる第４処理と、
前記抽出された文に含まれる単語の中から、前記目標語との間で関係Ｒにある単語x_jを抽出させる第５処理と、
前記第４処理及び前記第５処理を前記抽出された全ての単語r_iについて行い、これにより抽出された全ての単語x_jに対して前記解Ｘへの所属度ｇｒａｄｅＸ（x_j）を算出させる第６処理と、
前記算出された所属度ｇｒａｄｅＸ（x_j）の値が高い所定数の単語x_jを前記解Ｘに含まれる前記目標語に関係する候補語として抽出させる第７処理と
を実行させる
ことを特徴とする類推プログラム。

【発明の詳細な説明】
【技術分野】
【０００１】
この発明は、構造写像理論に基づく類推方式を用いた類推方法、類推システム及び類推プログラムに関する。
【背景技術】
【０００２】
従来より、類推の最も基本的な形として、Ａ：Ｂ＝Ｃ：Ｘという４項類推が知られている。４項類推は、例えばＡとＢ間の関係を推理し、基底領域（ベース）から目標領域（ターゲット）へ関係を写像し、推理された関係をＣとＸに適用するというようなアプローチで行われ、その処理過程はベースからターゲットへと関係を写像することが中心となる（非特許文献１及び２）。なお、ここでベースとは、類推する際に用いる既存の知識のことであり、ターゲットとは、解決しなければならない未知の問題を指す。
【０００３】
また、関係の写像においては、構造写像理論が知られている（非特許文献３）。この構造写像理論では、写像の際に膨大に生じてしまう無意味な知識を、構造を利用することで排除するアプローチが取られている。この理論によれば、妥当な類推写像は、属性の非写像、構造の一貫性、システム性原理の３つの基準を満たすとされている。
【先行技術文献】
【非特許文献】
【０００４】
【非特許文献１】R.J.Sternberg, “Intelligence, Information Processing and Analogical Reasoning”, Lawrence Erlbaum Associates, 1977
【非特許文献２】R.J.Sternberg, “Component Processes in Analogical Reasoning”, Psychological Review, 1977
【非特許文献３】D.Gentner, “Structure-Mapping: A Theoretical Framework for Analogy”, Cognitive Science, 1983
【発明の概要】
【発明が解決しようとする課題】
【０００５】
この発明は、構造写像理論に基づく類推方式によってベースとターゲットが異なる概念に属している場合でも精度良くある程度正しい解を得ることができる類推方法、類推システム及び類推プログラムを提供することを目的とする。
【課題を解決するための手段】
【０００６】
本発明に係る類推方法は、複数の基底語（ｂａｓｅ１，ｂａｓｅ２）の間の関係Ｒから、目標語（ｔａｒｇｅｔ）との間で関係Ｒにある解Ｘを類推する構造写像理論に基づく類

推方法であって、類推に用いられる知識情報として形態素解析された複数の文が蓄積されたコーパスから、前記複数の基底語が同時に出現する文を抽出する第１ステップと、前記抽出された文に含まれる単語から前記複数の基底語の間の関係Ｒを表す単語ｒ$_i$を抽出する第２ステップと、前記抽出された単語ｒ$_i$について、前記関係Ｒへの所属度ｇｒａｄｅＲ（ｒ$_i$）を算出する第３ステップと、前記目標語と前記単語ｒ$_i$とが同時に出現する文を前記コーパスから抽出する第４ステップと、前記抽出された文に含まれる単語の中から、前記目標語との間で関係Ｒにある単語ｘ$_j$を抽出する第５ステップと、前記第４ステップ及び前記第５ステップを前記抽出された全ての単語ｒ$_i$について行い、これにより抽出された全ての単語ｘ$_j$に対して前記解Ｘへの所属度ｇｒａｄｅＸ（ｘ$_j$）を算出する第６ステップと、前記算出された所属度ｇｒａｄｅＸ（ｘ$_j$）の値が高い所定数の単語ｘ$_j$を前記解Ｘに含まれる前記目標語に関係する候補語として抽出する第７ステップとを備えたことを特徴とする。

【０００７】

好ましい実施形態においては、例えば前記第４ステップでは、前記所属度ｇｒａｄｅＲ（ｒ$_i$）の値が所定値よりも高い単語ｒ$_i$と前記目標語とが同時に出現する文のみを抽出し、例えば前記第５ステップでは、前記単語ｘ$_j$は前記複数の基底語及び前記単語ｒ$_i$の前記文における記載順序に基づき抽出される。

【０００８】

また、前記関係Ｒは、例えばこの関係Ｒを構成する単語をｒ$_i$とし、メンバーシップ関数の値をｇｒａｄｅ（ｒ$_i$）とした場合、

$$R=\{grade(r_i)/r_i\}\quad(i=1,2,\cdots,n)$$

として表される概念ファジィ集合であり、前記解Ｘは、例えばこの解Ｘを構成する単語をｘ$_j$とし、メンバーシップ関数の値をｇｒａｄｅ（ｘ$_j$）とした場合、

$$X=\{grade(x_j)/x_j\}\quad(j=1,2,\cdots,n)$$

として表される概念ファジィ集合である。

【０００９】

また、前記所属度ｇｒａｄｅＲ（ｒ$_i$）の値は、例えばｃｏｕｎｔ（ｒ$_i$）を関係Ｒとして抽出された単語ｒ$_i$に対する頻度、Ｎ（ｂａｓｅ１∩ｂａｓｅ２）を複数の基底語が同時に出現した文の数、Ｎ（ｒ$_i$）をコーパスに含まれる文中に単語ｒ$_i$が現れる文の数、Ｎｄｏｃをコーパスに含まれる全ての文の数とした場合、

$$gradeR(r_i)=\alpha\cdot\log\left(1+\frac{\beta\cdot Ndoc}{N(base1\cap base2)\cdot N(r_i)}\right)$$

として表される。ただしα、βは調整可能なパラメータである。

【００１０】

更に、前記所属度ｇｒａｄｅＸ（ｘ$_j$）の値は、例えばｇｒａｄｅ（ｘ$_j$｜ｒ$_i$）を単語ｒ$_i$が目標語に写像されたときに単語ｘ$_j$が解Ｘとして正しいかを示す指数、Ｎｒを目標語に代表される単語ｒ$_i$の数とした場合、

$$gradeX(x_j)=\sum_{i=1}^{Nr}grade(x_j\mid r_i)$$

として表される。

【００１１】

なお、前記ｇｒａｄｅＸ（ｘ$_j$｜ｒ$_i$）の値は、例えばｃｏｕｎｔ（ｘ$_j$｜ｒ$_i$）を単語ｒ$_i$が目標語に写像されたときの解Ｘの候補として抽出される頻度、Ｎ（ｒ$_i$∩ｔａｒｇｅｔ）をコーパスに含まれる単語ｒ$_i$と目標語とが同時に出現した文の数、Ｎ（ｘ$_j$）をコーパスに含まれる文中に単語ｘ$_j$が現れる文の数、Ｎｄｏｃをコーパスに含まれる全ての文の数とした場合、

$$gradeX(x_j \mid r_i) = \gamma \cdot \log\left(1 + \frac{\delta \cdot Ndoc}{N(r_i \cap \mathrm{target}) \cdot N(x_j)}\right) \cdot gradeR(r_i)$$

として表される。ただしγ、δは調整可能なパラメータである。

【００１２】

本発明に係る類推システムは、複数の基底語（ｂａｓｅ１，ｂａｓｅ２）の間の関係Ｒから、目標語（ｔａｒｇｅｔ）との間の関係Ｒにある解Ｘを類推する構造写像理論に基づく類推方式を用いた類推システムであって、類推に用いられる知識情報として形態素解析された複数の文をコーパスとして蓄積する蓄積手段と、前記コーパスから前記複数の基底語が同時に出現する文を抽出すると共に、抽出された文に含まれる単語から前記複数の基底語の間の関係Ｒを表す単語r_iを抽出する第１抽出手段と、前記抽出された単語r_iについて、前記関係Ｒへの所属度ｇｒａｄｅＲ（r_i）を算出する第１算出手段と、前記目標語と前記単語r_iとが同時に出現する文を前記コーパスから抽出すると共に、抽出された文に含まれる単語の中から、前記目標語との間で関係Ｒにある単語x_jを抽出することを、前記抽出された全ての単語r_iについて行う第２抽出手段と、前記抽出された全ての単語x_jに対して前記解Ｘへの所属度ｇｒａｄｅＸ（x_j）を算出する第２算出手段と、前記算出された所属度ｇｒａｄｅＸ（x_j）の値が高い所定数の単語x_jを前記解Ｘに含まれる前記目標語に関係する候補語として抽出する第３抽出手段とを備えたことを特徴とする。

【００１３】

本発明に係る類推プログラムは、複数の基底語（ｂａｓｅ１，ｂａｓｅ２）の間の関係Ｒから、目標語（ｔａｒｇｅｔ）との間で関係Ｒにある解Ｘを類推する構造写像理論に基づく類推方式をコンピュータに実行させる類推プログラムであって、コンピュータに、類推に用いられる知識情報として形態素解析された複数の文が蓄積されたコーパスから、前記複数の基底語が同時に出現する文を抽出させる第１処理と、前記抽出された文に含まれる単語から前記複数の基底語の間の関係Ｒを表す単語r_iを抽出させる第２処理と、前記抽出された単語r_iについて、前記関係Ｒへの所属度ｇｒａｄｅＲ（r_i）を算出させる第３処理と、前記目標語と前記単語r_iとが同時に出現する文を前記コーパスから抽出させる第４処理と、前記抽出された文に含まれる単語の中から、前記目標語との間で関係Ｒにある単語x_jを抽出させる第５処理と、前記第４処理及び前記第５処理を前記抽出された全ての単語r_iについて行い、これにより抽出された全ての単語x_jに対して前記解Ｘへの所属度ｇｒａｄｅＸ（x_j）を算出させる第６処理と、前記算出された所属度ｇｒａｄｅＸ（x_j）の値が高い所定数の単語x_jを前記解Ｘに含まれる前記目標語に関係する候補語として抽出させる第７処理とを実行させることを特徴とする。

【発明の効果】

【００１４】

本発明によれば、構造写像理論に基づく類推方式によってベースとターゲットが異なる概念に属している場合でも精度良くある程度正しい解を得ることができる類推方法、類推システム及び類推プログラムを提供することができる。

【図面の簡単な説明】

【００１５】

【図１】本発明の一実施形態に係る類推システムの全体概要を説明するための図である。

【図２】同類推システムにおける類推方法による類推処理手順を示すフローチャートである。

【図３】同類推システムにおける類推方法による類推処理手順の一例を説明するための図である。

【発明を実施するための形態】

【００１６】

以下に、添付の図面を参照して、この発明に係る類推方法、類推システム及び類推プログラムの実施の形態を詳細に説明する。図１は、本発明の一実施形態に係る類推システム

の全体概要を説明するための図である。本実施形態に係る類推システムでは、構造写像理論に基づく類推方式として類推のベースとターゲットとが明確で、最も簡潔なモデルである４項類推を例に挙げて説明する。

【００１７】

ここで、４項類推とは、Ａ：Ｂ＝Ｃ：Ｄという形式の問題を指す。この４項類推の左辺第一項Ａを複数の基底語のうちの１つであるベース１（ｂａｓｅ１）とし、左辺第二項Ｂを複数の基底語のうちの他の１つであるベース２（ｂａｓｅ２）とする。また、右辺第一項Ｃを目標語であるターゲット（ｔａｒｇｅｔ）とし、右辺第二項Ｄを類推により求めるべき解Ｘとする。そして、写像対象となるベース１とベース２との関係をＲとする。

【００１８】

図１に示すように、コーパス１０は、類推に用いられる知識情報として形態素解析された複数の文が蓄積されたデータベース（ＤＢ）である。関係抽出モジュール（Ｅｘｔｒａｃｔ　Ｒｅｌａｔｉｏｎ　Ｍｏｄｕｌｅ：ＥＲＭ）２０は、ベース１及びベース２に基づいて、関係Ｒを抽出するモジュールである。関係写像モジュール（Ｍａｐｐｉｎｇ　Ｒｅｌａｔｉｏｎ　Ｍｏｄｕｌｅ：ＭＲＭ）３０は、ＥＲＭ２０により抽出された関係Ｒをターゲットに写像するモジュールである。

【００１９】

なお、このように構成された類推システムは、例えばパーソナルコンピュータやワークステーション等のハードウェア上で本発明に係る類推プログラムを実行することにより実現され、類推システムに対して入力されたベース１，２及びターゲットから解Ｘを類推して出力するように機能する。パーソナルコンピュータやワークステーション等のハードウェア構成については公知であるため、ここでは説明を省略する。

【００２０】

具体的には、図２に示すように、まず、コーパス１０から入力されたベース１，２が同時に出現する文１１（図１参照）を全て抽出する（ステップＳ１００）。例えば、図３に示すように、“ｆｉｓｈ：ｓｃａｌｅ＝ｂｉｒｄ：Ｘ”という問題が与えられた場合、図３中矢印（１）で示すように、ベース１，２である「ｆｉｓｈ」、「ｓｃａｌｅ」が同時に出現する文「Ｆｉｓｈ　ｉｓ　ｃｏｖｅｒｅｄ　ｗｉｔｈ　ｓｃａｌｅ．」や「Ｆｉｓｈ　ｈａｓ　ｓｃａｌｅｓ．」等をコーパス１０から全て抽出する。

【００２１】

次に、抽出された文に含まれる単語からベース１，２により写像される概念ファジィ集合の要素となる関係Ｒを表す単語$ｒ_{i}$を抽出する（ステップＳ１０２）。例えば、図３中矢印（２）で示すように、「ｆｉｓｈ」、「ｓｃａｌｅ」の関係Ｒを表す単語「ｉｓ」、「ｈａｓ（ｈａｖｅ）」、「ｗｉｔｈ」、「ｃｏｖｅｒ」等の単語$ｒ_{i}$を抽出された文から抽出する。

【００２２】

関係Ｒは、この関係Ｒを構成する単語を$ｒ_{i}$とし、メンバーシップ関数の値をｇｒａｄｅ（$ｒ_{i}$）とした場合、次式（１）として表される。

【００２３】

【数１】

$$R=\{grade(r_i)/r_i\}\quad(i=1,2,\cdots,n)$$

・・・（１）

【００２４】

そして、抽出された全ての単語$ｒ_{i}$について、関係Ｒへの所属度ｇｒａｄｅＲ（$ｒ_{i}$）を算出する（ステップＳ１０４）。この所属度ｇｒａｄｅＲ（$ｒ_{i}$）の値は、ｃｏｕｎｔ（$ｒ_{i}$）を関係Ｒとして抽出された単語$ｒ_{i}$に対する頻度、Ｎ（ｂａｓｅ１∩ｂａｓｅ２）を複数のベース１，２が同時に出現した文の数、Ｎ（$ｒ_{i}$）をコーパス１０に含まれる文中に単語$ｒ_{i}$が現れる文の数、Ｎｄｏｃをコーパス１０に含まれる全ての文の数とした場合、次式（２）として表される。

【００２５】

【数２】

$$gradeR(r_i) = \alpha \cdot \log\left(1 + \frac{\beta \cdot Ndoc}{N(base1 \cap base2) \cdot N(r_i)}\right)$$

・・・（２）

ここでα、βについて調整し、書き換えると次のように表すことができる。

$$gradeR(r_i) = \log\left(1 + \frac{count(r_i) \cdot Ndoc}{N(base1 \cap base2) \cdot N(r_i)}\right) \cdot \log_{10}(1 + count(r_i))$$

【００２６】

次に、入力されたターゲットに関係Ｒを写像して、ターゲットと単語r_iとが同時に出現する文１２（図１参照）をコーパス１０から全て抽出する（ステップＳ１０６）。例えば、図３中矢印（３）で示すように、ターゲットである「ｂｉｒｄ」と単語r_iである「ｉｓ」、「ｈａｓ（ｈａｖｅ）」、「ｗｉｔｈ」、「ｃｏｖｅｒ」とが同時に出現する文「Ｂｉｒｄ ｉｓ ｃｏｖｅｒｅｄ ｗｉｔｈ ｆｅａｔｈｅｒ.」や「Ｂｉｒｄ ｈａｓ ｗｉｎｇ.」等をコーパス１０から全て抽出する。

【００２７】

そして、抽出された文に含まれる単語の中から、関係Ｒに基づく概念ファジィ集合を構成する解Ｘの候補となる単語x_jを抽出する（ステップＳ１０８）。例えば、図３中矢印（４）で示すように、「ｂｉｒｄ」と関係Ｒのような概念ファジィ集合を構成する解Ｘの候補となるような単語「ｗｉｎｇ」、「ｆｅａｔｈｅｒ」等の単語x_jを抽出する。

【００２８】

解Ｘは、この解Ｘを構成する単語をx_jとし、メンバーシップ関数の値をｇｒａｄｅ（x_j）とした場合、次式（３）として表される。

【００２９】

【数３】

$$X = \{grade(x_j)/x_j\} \quad (j = 1,2,\cdots,n)$$

・・・（３）

【００３０】

その後、抽出された全ての単語x_jについてステップＳ１０６及びステップＳ１０８の処理が行われたか否かを判断し（ステップＳ１１０）、行われていない場合（ステップＳ１１０のＮ）は上記ステップＳ１０６に移行して処理を繰り返すと共に、行われた場合（ステップＳ１１０のＹ）は、抽出された全ての単語x_jに対して解Ｘへの所属度ｇｒａｄｅＸ（x_j）を算出する（ステップＳ１１２）。

【００３１】

この所属度ｇｒａｄｅＸ（x_j）の値は、ｇｒａｄｅ（x_j｜r_i）を単語r_iが目標語に写像されたときに単語x_jが解Ｘとして正しいかを示す指数、Ｎｒを目標語に代表される単語r_iの数とした場合、次式（４）として表される。

【００３２】

【数４】

$$gradeX(x_j) = \sum_{i=1}^{Nr} grade(x_j \mid r_i)$$

・・・（４）

【００３３】

最後に、例えば算出された所属度ｇｒａｄｅＸ（x_j）の値が高い所定数の単語x_jを解Ｘに含まれるターゲットに関係する候補語として抽出し（ステップＳ１１４）、本フローチャートによる処理が終了される。なお、ｇｒａｄｅＸ（x_j｜r_i）の値は、ｃｏｕｎｔ（x_j｜r_i）を単語r_iがターゲットに写像されたときの解Ｘの候補として抽出される頻度、Ｎ（r_i∩ｔａｒｇｅｔ）をコーパス１０に含まれる単語r_iとターゲットとが同時に出現した文の数、Ｎ（x_j）をコーパス１０に含まれる文中に単語x_jが現れる文の数、Ｎｄｏｃをコーパス１０に含まれる全ての文の数とした場合、次式（５）として

表される。

【００３４】

【数５】

$$gradeX(x_j \mid r_i) = \gamma \cdot \log\left(1 + \frac{\delta \cdot Ndoc}{N(r_i \cap \mathrm{target}) \cdot N(x_j)}\right) \cdot gradeR(r_i)$$

・・・（５）

ここでγ、δについて調整し、書き換えると次のように表すことができる。

$$gradeX(x_j \mid r_i) = \log\left(1 + \frac{count(x_j \mid r_i) \cdot Ndoc}{N(r_i \cap \mathrm{target}) \cdot N(x_j)}\right) \cdot \log\left(1 + count(x_j \mid r_i)\right) \cdot gradeR(r_i)$$

【００３５】

ここで、上記ステップＳ１０６では、所属度ｇｒａｄｅＲ（ｒ$_i$）の値が高い単語ｒ$_i$についての関係Ｒのみをターゲットに写像するようにしても良い。このようにすれば、最終的に解Ｘの候補語として不適当な単語が抽出されることを避けることが可能となる。

【００３６】

また、解Ｘの正しい答えの候補語を得るために、Ａ：Ｂ＝Ｃ：ＸとＢ：Ａ＝Ｃ：Ｘの違いは考慮される必要があるので、上記ステップＳ１０８では、例えば単語ｘ$_j$はベース１，２及び単語ｒ$_i$の抽出された文における記載順序に基づき抽出するようにしても良い。この場合、単語ｘ$_j$は、次の規則の下に抽出される。

【００３７】

すなわち、（１）単語ｒ$_i$が記載順序「ベース１→単語ｒ$_i$→ベース２」の関係下で頻繁に抽出された場合は、単語ｒ$_i$の後に記載された単語が単語ｘ$_j$として抽出される。また、（２）単語ｒ$_i$が記載順序「ベース２→単語ｒ$_i$→ベース１」の関係下で頻繁に抽出された場合は、単語ｒ$_i$の前に記載された単語が単語ｘ$_j$として抽出される。更に、（３）単語ｒ$_i$が上記（１）及び（２）の関係下でそれぞれ等しく抽出された場合は、単語ｘ$_j$は単語ｒ$_i$の前後いずれかに記載された単語として抽出される。

【００３８】

本実施形態に係る類推システムでは、このような処理により、ベース１，２とターゲットとが異なる概念に属している場合でも、ある程度正しい解Ｘの候補を精度良く得ることが可能となる。また、得られた解Ｘを類推による候補として類推システムに備えられた図示しない表示手段に表示したり、印刷手段で印刷出力したり、音声出力手段で報知したりして利用者に提示するようにすれば、類推による推薦を行うことも可能となる。次に、上述した類推システムを用いた本出願人による試験について説明する。

【００３９】

この試験においては、フリー百科辞典の英語版Ｗｉｋｉｐｅｄｉａに書かれたある時点の全ての文（約３，６９１，０００項目、約４３，６７０，０００文）を蓄積したコーパス１０を用いた。このＷｉｋｉｐｅｄｉａに書かれた文をコーパス１０に記憶する際には、フリーソフトウェアのＴｒｅｅ　Ｔａｇｇｅｒを用いて形態素解析を行った。また、コーパス１０の作成と検索にはフリーソフトウェアのＬｕｃｅｎｅを使用した。

【００４０】

そして、試験では、単語で表記され、且つベース１、ベース２、ターゲット、解Ｘが全て１単語の名詞である問題のみを対象にした。更に、２つの名詞の関係としては動詞が適切である場合が多いため、関係Ｒは動詞のみとした。なお、４項類推では絶対的な正解を規定することが困難である。これは４項類推が、回答者の持つ知識や主観に少なからず依存するためである。そこで、次のように類推システムの評価を行うこととした。

【００４１】

まず、参加した評価者はそれぞれ４項類推の解Ｘを求め、次に、同じ４項類推について類推システムが解Ｘを求めてその候補に当たる上位１０個を評価者に提示した。評価者は提示された解Ｘを基に、類推システムが行った類推がどの程度正しいかを２，１，０の３段階で評価した。このような評価を１０人の評価者が、合計２０題の問題に対して行うこ

とで、類推システムの評価を行った。

【００４２】

以下の表１に「ｅａｒｔｈ：ｓｕｎ＝ｍｏｏｎ：Ｘ」という４項類推の問題に対して類推システムが求めた関係Ｒと解Ｘの候補上位１０個を示す。

【００４３】

【表１】

“earth：sun＝moon：X” の類推結果

関係	gradeR(r_i)
revolve	1.0000
orbit	0.8511
rotate	0.4370
eclipse	0.2800
illuminate	0.2620
shine	0.2441
absorb	0.2376
envelop	0.2339
radiate	0.2249
obscure	0.2062

解	gradeX(x_j)
earth	1.0000
sun	0.8339
planet	0.6362
orbit	0.5570
jupiter	0.4738
saturn	0.3829
ecliptic	0.3145
spacecraft	0.2929
inclination	0.2860
eclipse	0.2464

【００４４】

問題「ｅａｒｔｈ：ｓｕｎ＝ｍｏｏｎ：Ｘ」は、評価者全員が、類推システムが求めた解Ｘを正しいと評価した問題である。上記表１を見ると、「ｅａｒｔｈ」と「ｓｕｎ」の関係として「ｒｅｖｏｌｖｅ」や「ｏｒｂｉｔ」等が挙げられている。また、解Ｘの候補１位に「ｅａｒｔｈ」という「ｅａｒｔｈ：ｓｕｎ＝ｍｏｏｎ：Ｘ」の解として納得できる単語が挙げられている。

【００４５】

この結果から、類推システムは、「ｅａｒｔｈ：ｓｕｎ＝ｍｏｏｎ：Ｘ」という問題に関して、「ｅａｒｔｈ」と「ｓｕｎ」の関係Ｒを正しく理解し、その関係Ｒを基に正しい解Ｘを導くことに成功したと言える。

【００４６】

また、以下の表２は、「ｆｉｓｈ：ｓｃａｌｅ＝ｂｉｒｄ：Ｘ」という問題と、「ｆｉｓｈ：ｆｉｎ＝ｂｉｒｄ：Ｘ」という問題とについて類推システムが求めた解を比較したものである。この２つの問題は、与えられた３項のうち、ベース２のみが異なる。

【００４７】

【表２】

base2 の変化による類推結果の相違

fish：scale = bird：X	
X の要素	gradeX(x_j)
insect	1.0000
feather	0.8708
prey	0.8032
nectar	0.7349
flycatcher	0.6799
nest	0.6704
buttonquails	0.6445
ground-nesting	0.6165
specie	0.6116
wing	0.6039

fish：fin = bird：X	
X の要素	gradeX(x_j)
feather	1.0000
wing	0.7189
beak	0.6494
tail	0.5421
plumage	0.4391
warbler	0.4323
specie	0.4182
insect	0.4151
nest	0.4143
bill	0.4140

【００４８】

上記表２から、類推システムは、２つの問題に対して適切に異なる解Ｘを求めたことが分かる。実際に、「ｆｉｓｈ：ｆｉｎ＝ｂｉｒｄ：Ｘ」の解Ｘでは、「ｆｉｓｈ：ｓｃａｌｅ＝ｂｉｒｄ：Ｘ」の解Ｘよりも、「ｗｉｎｇ」や「ｔａｉｌ」の順位が上がっている。これは、類推システムが、それぞれの問題に適した関係Ｒを求め、より解Ｘとして適した単語を導き出した結果であると言える。

【００４９】

なお、表２に示した「ｆｉｓｈ：ｓｃａｌｅ＝ｂｉｒｄ：Ｘ」という問題で、類推システムが求めた解Ｘの候補１位に「ｉｎｓｅｃｔ」という単語が選ばれているが、これは、ＥＲＭ２０によって、「ｆｉｓｈ」と「ｓｃａｌｅ」の関係として「ｆｅｅｄ」という単語が抽出されたためである。これは、正しい関係Ｒを得られないために、解Ｘに不適切な単語が含まれた例として挙げられる。

【００５０】

最後に、１０人の評価者による２０問の問題に対する類推システムの評価を表３に示す。

【００５１】

【表３】

20問の問題に対する類推結果とシステム評価

番号	問題	解の１位	解の２位	解の３位	評価の合計
1	fish : scale = bird : ?	insect	feather	Prey	10
2	bird : feather = fish : ?	fin	water	fishing	6
3	fish : fin = bird : ?	feather	wing	beak	17
4	bird : sky = fish : ?	fishing	water	trout	7
5	earth : sun = moon : ?	earth	sun	planet	20
6	earth : sun = electron : ?	nucleus	atom	energy	19
7	river : water = circuit : ?	voltage	capacitor	device	10
8	teacher : student = doctor : ?	patient	tardis	hospital	19
9	sardine : fish = sparrow : ?	bird	seed	insect	15
10	frog : snake = mouse : ?	cat	lizard	snake	20
11	beer : barley = wine : ?	grape	beverage	winery	20
12	france : paris = spain : ?	madrid	france	toledo	13
13	spain : madrid = france : ?	french	paris	swizerland	10
14	america: washington = france : ?	french	switzerland	finland	4
15	animal : food = automobile : ?	vehicle	car	motor	6
16	automobile : gasoline = yacht : ?	boat	sailing	vessel	4
17	automobile : gasoline = train : ?	locomotive	passenger	rail	2
18	winter : snow = summer : ?	winter	olympic	snow	1
19	surplus : deficit = black : ?	hawk	hole	color	2
20	up : down = left : ?	hand	right	arm	7

【００５２】

上記表３では、２０問の問題と、類推結果として類推システムが提示した解Ｘのうち所属度ｇｒａｄｅＲ（x_j）の値が高い上位３つの要素、類推システムが求めた解Ｘについて１０人の評価者が付けた評価点の合計を示している。また、この試験での評価の満点は２０点となっている。

【００５３】

表３に挙げられた結果を見ると、１０点以上の評価を得た問題が１１問あった。そのうち３問が満点の２０点の評価を得ていた。また、５点以下の評価を得た問題が５問あった。表３に示された評価の平均は１０．６となった。

【００５４】

そして、評価者全員が類推システムが求めた解Ｘが正しいとした３問の問題は、「ｅａｒｔｈ：ｓｕｎ＝ｍｏｏｎ：Ｘ」、「ｓｎａｋｅ：ｆｒｏｇ＝ｃａｔ：Ｘ」、「ｂｅｅｒ：ｂａｒｌｅｙ＝ｗｉｎｅ：Ｘ」であり、それぞれ解Ｘの１位に「ｅａｒｔｈ」、「ｍｏｕｓｅ」、「ｇｒａｐｅ」を挙げていた。

【００５５】

逆に結果が悪かった問題は、「ｗｉｎｔｅｒ：ｓｎｏｗ＝ｓｕｍｍｅｒ：Ｘ」という問題で、解Ｘの候補１位には「ｗｉｎｔｅｒ」が挙げられていた。

【００５６】

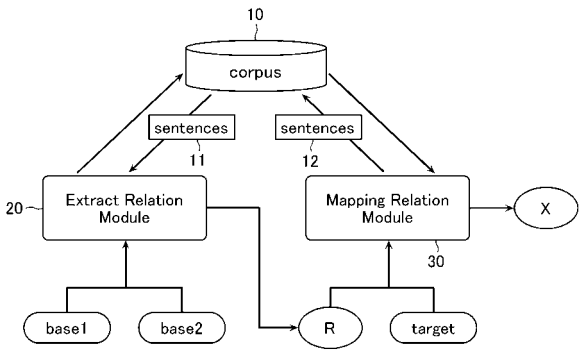
このように、結果が良かった問題では、関係Ｒとして正しいと考えられるものが抽出されていた。逆に、結果が悪かった問題では、関係Ｒとして正しくないと考えられるものが抽出されていた。従って、類推システムでは、関係Ｒの抽出が類推の可否に大きく影響することが分かったが、本実施形態に係る類推システムによれば、ベースとターゲットが異なる概念に属している場合でも、概ね精度良くある程度正しい解Ｘを得ることが可能なことが判明した。

【符号の説明】

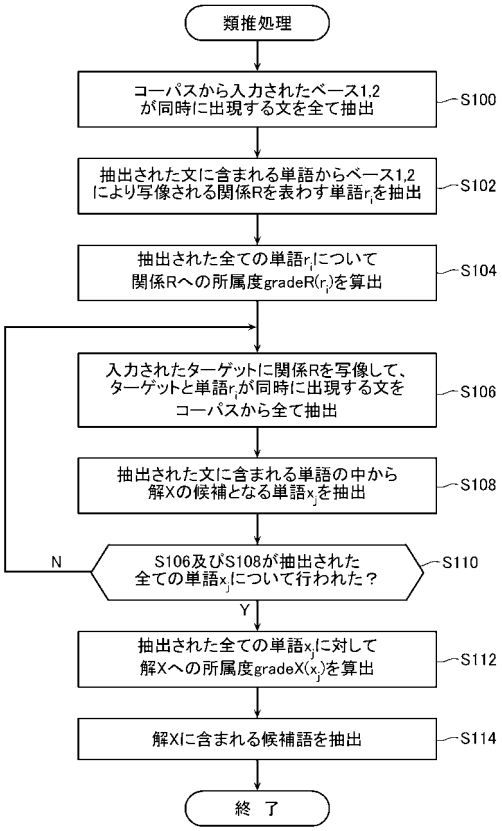
【００５７】

１０　　コーパス
２０　　関係抽出モジュール（ＥＲＭ）
３０　　関係写像モジュール（ＭＲＭ）

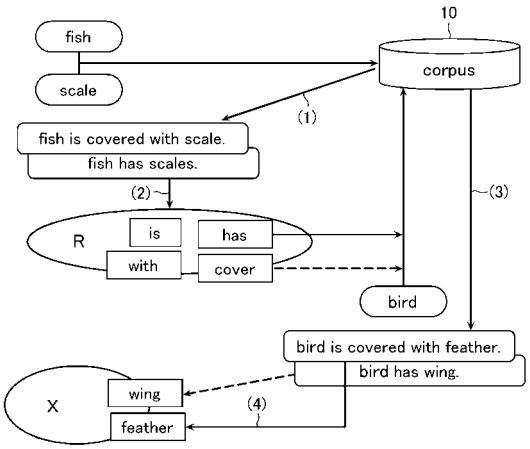
【図１】



【図２】



【図３】


10
fish
scale
corpus
(1)
fish is covered with scale.
fish has scales.
(2)
(3)
R
is
has
with
cover
bird
bird is covered with feather.
bird has wing.
X
wing
feather
(4)