

(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2004-362306

(P2004-362306A)

(43) 公開日 平成16年12月24日(2004.12.24)

(51) Int. Cl.⁷
G06F 17/18

F I
G06F 17/18

テーマコード(参考)
5B056

審査請求 有 請求項の数 6 O L (全 26 頁)

<p>(21) 出願番号 特願2003-160465 (P2003-160465)</p> <p>(22) 出願日 平成15年6月5日(2003.6.5)</p> <p>特許法第30条第1項適用申請有り 2003年3月18日 言語処理学会発行の「言語処理学会第9回年次大会 発表論文集」に発表</p>	<p>(71) 出願人 301022471 独立行政法人情報通信研究機構 東京都小金井市貫井北町4-2-1</p> <p>(74) 代理人 100103827 弁理士 平岡 憲一</p> <p>(74) 代理人 100097836 弁理士 福井 國敞</p> <p>(72) 発明者 村田 真樹 東京都小金井市貫井北町4-2-1 独立行政法人通信総合研究所内</p> <p>(72) 発明者 栴山 享子 東京都小金井市貫井北町4-2-1 独立行政法人通信総合研究所内</p>
--	---

最終頁に続く

(54) 【発明の名称】 可視化装置及びプログラム

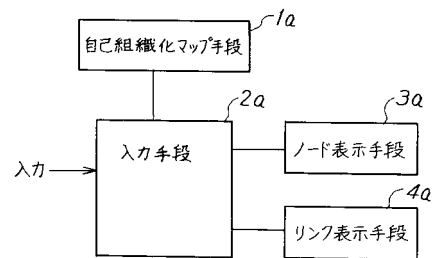
(57) 【要約】

【課題】一つのノードが他の複数のノードとリンクする場合等の数量的なリンク結果を可視化して明確に示すこと。

【解決手段】複数のノード間のリンクデータを入力する入力手段2aと、前記リンクデータにより各ノードの配置座標位置を求める自己組織化マップ手段1aと、前記リンクデータのリンクがもつ値の大きさによる線の太さで二つのノードの配置座標位置を線で結んで表示するリンク表示手段4aと、前記求めた各ノードの配置座標位置にノードを表示するノード表示手段3aとを備える。

【選択図】 図1

本発明の原理説明図



【特許請求の範囲】**【請求項 1】**

複数のノード間のリンクデータを入力する入力手段と、
前記リンクデータにより各ノードの配置座標位置を求める自己組織化マップ手段と、
前記リンクデータのリンクがもつ値の大きさによる線の太さで二つのノードの配置座標位置を線で結んで表示するリンク表示手段と、
前記求めた各ノードの配置座標位置にノードを表示するノード表示手段とを備えることを特徴とした可視化装置。

【請求項 2】

前記リンク表示手段は、前記リンクデータのリンクがもつ値が一定値以上のもののみ、前記二つのノードの配置座標位置を線で結んで表示することを特徴とした請求項 1 記載の可視化装置。 10

【請求項 3】

複数のノード間のリンクデータと前記各ノードの配置座標を入力する入力手段と、
リンクをもつ二つのノードの配置座標位置を線で結んでリンクを表示するリンク表示手段と、
前記入力された各ノードの配置座標位置にノードを表示するノード表示手段とを備え、
前記リンク表示手段は、前記二つのノードを線で結ぶ際、一方のノード側の線の太さを一方のノードから他方のノードへのリンクがもつ値に基づいた線の太さにし、他方のノード側の線の太さを他方のノードから一方のノードへのリンクがもつ値に基づいた線の太さにすることを特徴とした可視化装置。 20

【請求項 4】

複数のノード間のリンクデータを入力する入力手段と、
前記リンクデータにより前記各ノードの配置座標位置を求める自己組織化マップ手段と、
リンクをもつ二つの前記求めたノードの配置座標位置を線で結んでリンクを表示するリンク表示手段と、
前記求めた各ノードの配置座標位置にノードを表示するノード表示手段とを備え、
前記リンク表示手段は、前記二つのノードを線で結ぶ際、一方のノード側の線の太さを一方のノードから他方のノードへのリンクがもつ値に基づいた線の太さにし、他方のノード側の線の太さを他方のノードから一方のノードへのリンクがもつ値に基づいた線の太さにすることを特徴とした可視化装置。 30

【請求項 5】

複数のノード間のリンクデータを入力する入力手段と、
前記リンクデータにより各ノードの配置座標位置を求める自己組織化マップ手段と、
前記リンクデータのリンクがもつ値の大きさによる線の太さで二つのノードの配置座標位置を線で結んで表示するリンク表示手段と、
前記求めた各ノードの配置座標位置にノードを表示するノード表示手段として、
コンピュータを機能させるためのプログラム。

【請求項 6】

複数のノード間のリンクデータと前記各ノードの配置座標を入力する入力手段と、
前記入力された各ノードの配置座標位置にノードを表示するノード表示手段とを備え、
二つのノードを線で結ぶ際、一方のノード側の線の太さを一方のノードから他方のノードへのリンクがもつ値に基づいた線の太さにし、他方のノード側の線の太さを他方のノードから一方のノードへのリンクがもつ値に基づいた線の太さにする、リンクをもつ二つのノードの配置座標位置を線で結んでリンクを表示するリンク表示手段として、
コンピュータを機能させるためのプログラム。 40

【発明の詳細な説明】**【0001】****【発明の属する技術分野】**

本発明は、品詞の共有性の調査結果等の数量的な調査結果を可視化して、調査結果を明確 50

に示すことができる可視化装置及びプログラムに関する。

【0002】

【従来の技術】

従来、数量的な調査結果は、表やグラフにして、可視化を行うのが一般的であった。

【0003】

【発明が解決しようとする課題】

上記従来の表やグラフにして、可視化を行うものは、品詞の共有性の調査結果等のように一つのノードが他の複数のノードとリンクする場合、表やグラフが複雑になり、一度に可視化することは困難なものであった。

【0004】

本発明は、このような従来の問題点の解決を図り、一つのノードが他の複数のノードとリンクする場合等の数量的なリンク結果を可視化して明確に示すことを目的とする。

【0005】

【課題を解決するための手段】

図1は本発明の原理説明図である。図1中、1aは自己組織化マップ手段、2aは入力手段、3aはノード表示手段、4aはリンク表示手段である。

【0006】

本発明は、前記従来の課題を解決するため次のような手段を有する。

【0007】

(1)：複数のノード間のリンクデータを入力する入力手段2aと、前記リンクデータにより各ノードの配置座標位置を求める自己組織化マップ手段1aと、前記リンクデータのリンクがもつ値の大きさによる線の太さで二つのノードの配置座標位置を線で結んで表示するリンク表示手段4aと、前記求めた各ノードの配置座標位置にノードを表示するノード表示手段3aとを備える。このため、複数のノード間のリンク関係とリンクの大きさを明確に示すことができる。

【0008】

(2)：前記(1)の可視化装置において、前記リンク表示手段4aは、前記リンクデータのリンクがもつ値が一定値以上のもののみ、前記二つのノードの配置座標位置を線で結んで表示する。このため、表示するノードが多い場合のノード間のリンク関係とリンクの大きさを明確に示すことができる。

【0009】

(3)：複数のノード間のリンクデータと前記各ノードの配置座標を入力する入力手段1aと、リンクをもつ二つのノードの配置座標位置を線で結んでリンクを表示するリンク表示手段4aと、前記入力された各ノードの配置座標位置にノードを表示するノード表示手段3aとを備え、前記リンク表示手段4aは、前記二つのノードを線で結ぶ際、一方のノード側の線の太さを一方のノードから他方のノードへのリンクがもつ値に基づいた線の太さにし、他方のノード側の線の太さを他方のノードから一方のノードへのリンクがもつ値に基づいた線の太さにする。このため、ノード間で方向性のあるリンクデータを一つの線で表示することができる。

【0010】

(4)：複数のノード間のリンクデータを入力する入力手段2aと、前記リンクデータにより前記各ノードの配置座標位置を求める自己組織化マップ手段1aと、リンクをもつ二つの前記求めたノードの配置座標位置を線で結んでリンクを表示するリンク表示手段4aと、前記求めた各ノードの配置座標位置にノードを表示するノード表示手段3aとを備え、前記リンク表示手段4aは、前記二つのノードを線で結ぶ際、一方のノード側の線の太さを一方のノードから他方のノードへのリンクがもつ値に基づいた線の太さにし、他方のノード側の線の太さを他方のノードから一方のノードへのリンクがもつ値に基づいた線の太さにする。このため、入力手段にリンクデータを入力するだけで、ノードの表示とノード間で方向性のあるデータを一つの線で表示することができる。

【0011】

10

20

30

40

50

【発明の実施の形態】

(1) : 可視化装置の説明

図2は可視化装置の説明図である。図2において、可視化装置には、データ入力部2、ノード表示部3、リンク表示部4が設けてある。データ入力部2は、ノードデータ及びリンクデータが入力される入力手段である。ノード表示部3は、ノードデータによりノードの記号を表示するノード表示手段である。リンク表示部4は、リンクデータに基づき、リンクデータの線の太さで、そのリンクの二つのノードの座標位置を線で結ぶリンク表示手段である。

【0012】

図3は可視化処理フローチャートである。以下、図3の処理S1～S3に従って説明する 10

【0013】

S1 : データ入力部2に、次のような書式で、ノードデータ及びリンクデータを入力する。

【0014】

ノードデータ(ノード情報) :

横軸の座標、縦軸の座標、ノードの記号

横軸の座標、縦軸の座標、ノードの記号

...

横軸の座標、縦軸の座標、ノードの記号 20

リンクデータ(リンク情報) :

ノードの記号、ノードの記号、線の太さ

ノードの記号、ノードの記号、線の太さ

...

ノードの記号、ノードの記号、線の太さ

S2 : リンク表示部4は、入力されたリンク情報に基づき、リンク情報の線の太さで、そのリンクの二つのノードの座標位置を線で結ぶ。

【0015】

なお、ノード座標位置は、ノード情報から取得する。これをすべてのリンクに対して行なう。 30

【0016】

S3 : ノード表示部3は、入力されたノード情報に基づき、そのノードの座標位置に、ノードの記号を表示する。

【0017】

(2) : 自己組織化マップ部を備える可視化装置の説明

図4は自己組織化マップ部を備える可視化装置の説明図である。図4において、可視化装置には、自己組織化マップ部1、データ入力部2、ノード表示部3、リンク表示部4、リンクデータ作成部5が設けてある。自己組織化マップ部1は、入力されたリンクデータから各ノードの適切な二次元での配置の座標を決定する自己組織化マップ手段である。データ入力部2は、ノードデータ及びリンクデータが入力される入力手段である。ノード表示部3は、ノードデータによりノードの記号を表示するノード表示手段である。リンク表示部4は、リンクデータに基づき、リンクデータの線の太さで、そのリンクの二つのノードの座標位置を線で結ぶリンク表示手段である。リンクデータ作成部5は、基データからリンクデータを作成するリンクデータ作成手段である(例として、式5参照)。 40

【0018】

図5は自己組織化マップ部を備える可視化処理フローチャートである。以下、図5の処理S11～S14に従って説明する。

【0019】

S11 : データ入力部2に、リンクデータ作成部5で作成したリンクデータを入力する。

【0020】

S 1 2 : 自己組織化マップ部 1 は、入力されたリンクデータだけを用い、各ノードの適切な二次元での配置の座標を決定する。このノードの適切な二次元での配置の座標を決定する自己組織化の技術は、例えば、次の馬らの文献がある。

【 0 0 2 1 】

(馬青, 神崎享子, 村田真樹, 内元清貴, 井佐原均 “日本語名詞の意味マップの自己組織化” 情報処理学会論文誌, Vol. 42, No. 10, 2001, p. 2379 - 2391. 参照)

自己組織化マップ部の入力の行列は、上記の馬らの文献を参考にして各ノード間の非類似度(距離)を意味する行列を与える。行列の次元はノードの個数とし、行列の縦横の各要素は、各ノードに対応する。行列の非対角要素の各値はリンクデータの線の太さとし、行列の対角要素は 0 とし、2次元配列の自己組織化マップを動かす。自己組織化マップの起動の際の詳細なパラメータは適宜実行のときに定める(なお、この自己組織化マップ部の入力の行列は、馬らの文献以外の方法をとることもできる)。

10

【 0 0 2 2 】

この自己組織化マップにより、ノードの適切な二次元での配置の座標が求まる。これがノード情報(ノードデータ)となる。

【 0 0 2 3 】

S 1 3 : リンク表示部 4 は、入力されたリンク情報に基づき、リンク情報の線の太さで、そのリンクの二つのノードの座標位置を線で結ぶ。

【 0 0 2 4 】

なお、ノード座標位置は、ノード情報から取得する。これをすべてのリンクに対して行なう。

20

【 0 0 2 5 】

S 1 4 : ノード表示部 3 は、入力されたノード情報に基づく、そのノードの座標位置に、ノードの記号を表示する。

【 0 0 2 6 】

なお、リンク表示部 4 とノード表示部 3 の処理は、どちらを先に行ってもよい。すなわち、処理 S 1 4 の後に処理 S 1 3 を行うようにすることもできる。

【 0 0 2 7 】

(3) : 単語辞書を用いた英語品詞間の転換に関する調査の説明

30

(はじめに)

英語の単語には品詞の転換という現象がある。これは接辞を付加せずに単語の品詞を変えることである。例えば、bottle という語は「瓶」という名詞であったが、「瓶詰にする」という動詞としても用いられるようになっていく。本調査はこのような品詞転換の現象を、英和辞典などの単語辞書を用いて調査する。具体的には各単語が持つ品詞の種類を電子化された単語辞書から自動取得し、品詞の種類ごとにそれら品詞を複数持つ単語の個数などを調査して、どういう品詞とどういう品詞の間で品詞の転換が生じやすいかを調べる。その際、自己組織化マップを利用した調査結果の可視化も行なう。これらの調査結果は、転換現象などの言語の歴史的変遷を調べる言語学の基礎データとして役に立つ可能性がある。

40

【 0 0 2 8 】

(単語辞書を用いた数量的調査)

本調査では単語辞書としてジーニアス英和辞典のデータを利用した。この辞書の見出し語の総数は、74848 個であった。次に各見出し語(単語)に出現する品詞のマークの数を計測した。その結果を以下の表 1 に示している。この個数の算出の際は、例えば、名詞と形容詞の品詞を持つ単語の個数は、名詞の出現回数と形容詞の出現回数の両方に加算される。表 1 からわかるように名詞の単語がもっとも多く次に形容詞が多い。また、この辞書では品詞のマークのつかない単語もあった。本研究の調査では、品詞のマークのつかない単語は扱わず、少なくとも一つは品詞のマークのついている 56837 個の単語を対象とした。

50

【 0 0 2 9 】

表 1 : 品詞の出現回数

品詞	出現回数	出現率
名詞	37158	0.4964
形容詞	14279	0.1908
動詞	8766	0.1171
副詞	5301	0.0708
間投詞	302	0.0040
連結形	242	0.0032
接尾辞	215	0.0029
接頭辞	157	0.0021
前置詞	145	0.0019
代名詞	126	0.0017
接続詞	71	0.0009
助動詞	32	0.0004
品詞付単語総数	56837	0.7594
単語(見出し語)総数	74848	1.0000

10

20

30

次に二つの品詞を持つ単語の割合などを調べて、どういう品詞とどういう品詞の間で品詞の転換現象がおきやすいかを調べた。例えば、品詞 A と品詞 B の両方を持つ単語の個数が多ければ品詞 A と品詞 B の間で多くの転換現象が起きていることがわかる。

【 0 0 3 0 】

この調査結果を図 6 に示している。図 6 は品詞二項間の共有性の調査結果の説明図であり、図 6 の「二項以下のみ」は、品詞を三種類以上持つ単語を除いて調査した結果で、図 6 の「すべて」は、品詞を三種類以上持つ単語を含めて品詞のマークをもつすべての単語で調査した結果である。「共起頻度」は、左の欄で示す二つの品詞をともに持つ単語の個数である。ただし、図 6 の「すべて」の場合などで生じる、品詞を三種類以上持つ単語については、 $2/n$ の重みをかけて用いる。n は、その単語が持つ品詞の個数である。これは例えば品詞を A、B、C の三種類持つ場合、転換現象は A B C、A C B、B A C、B C A、C A B、C B A の六つの順序で生じている可能性があるが、そのうち、A、B の品詞の組の間で転換が起きている割合が $2/3 (= 4/6)$ であるように、三種類以上の場合は対象とする二つの品詞の間で転換が起きている確率が 1 よりも下がるためこのような計算を行なう。図 6 の「共有率」は、例えば品詞 A と品詞 B の間でのものならば以下の式 1 で与えられる。

40

【 0 0 3 1 】

【 数 1 】

$$\text{共有率} = \frac{F_{a,b}}{F_a + F_b - F_{a,b}} \quad \text{式 1}$$

【0032】

ただし、 $F_{a,b}$ は品詞 A、B の共起頻度、 F_x は品詞 X の頻度（出現回数）である。すなわち、品詞 A、B いずれかを持つ単語に対する品詞 A、B の両方を持つ単語の割合を意味する。図 6 の「予測比」は、例えば、品詞 A と品詞 B の間でのものならば以下の式 2 で与えられる。

10

【0033】

【数 2】

$$\text{予測比} = Er(A, B) = \frac{F_{a,b}}{(F_a/Total)(F_b/Total) * Total} \quad \text{式 2}$$

【0034】

ただし、 $Er(A, B)$ は、後の式で用いる品詞 A、B の予測比を意味する記号で、 $Total$ は、調査対象の単語の総数である。これは、品詞 A、B が独立に出現すると仮定した場合、品詞 A、B の頻度のみから予測した品詞 A、B の共起頻度に対する実際の品詞 A、B の共起頻度の値の比である。つまり、この値が大きいほど予想される個数に比べて多くの共起が生じていることになる。ところでこの式は品詞二項の組を対象とする場合は $\log 2$ をつけると自己相互情報量と等しいものとなる。

20

【0035】

また、同様の調査を品詞三項の組でも行なった。その結果を図 7 に示している。図 7 は品詞三項間の共有性の調査結果の説明図であり、ここでは、品詞三項の組での実験では「すべて」の場合の結果は示さず、品詞を四種類以上持つ単語を省いた調査である「三項以下のみ」の結果しか示していない。「共起頻度」は三つの品詞がともに出現した単語の個数で「共有率」は以下の式 3 で、「予測比」は以下の式 4 で与えられる。

30

【0036】

【数 3】

$$\text{共有率} = \frac{F_{a,b,c}}{F_a + F_b + F_c - F_{a,b} - F_{b,c} - F_{c,a} + F_{a,b,c}} \quad \text{式 3}$$

$$\text{予測比} = \frac{F_{a,b,c}}{(F_a/Total)(F_b/Total)(F_c/Total) * Total} \quad \text{式 4}$$

40

【0037】

二つの調査結果の図 6、図 7 では、それぞれの尺度で上位 5 個の品詞の組の数値を太字で示している。また、表にない品詞の組は、共起頻度が 0 であったものである。

【0038】

この品詞間の共有性の調査には、「共起頻度」「共有率」「予測比」しか用いていないが、その他の尺度も共起頻度や個々の品詞の個数、調査単語総数から求まるものならば、本調査と同様、容易に単語辞書から算出可能である。また、本調査ではどういう尺度が調査に役に立つかがわからなかったため、共有性を調べる際の基本的な尺度の「共起頻度」

50

「共有率」を用い、また「共起頻度」だと頻度の大きい品詞の値が大きくなる欠点があり、また「共有率」だと比較する品詞間で頻度に差がある場合、値が小さくなる欠点があるため、品詞間で頻度に差があっても影響されない「予測比」も用いた。

【0039】

(調査結果の可視化の説明)

次に、先の数量的な調査結果を二次元のマップ上に表現した。これは先の数量的な調査結果を可視化することで、より効果的な調査結果の提示を試みるものである。この調査は、図6の「二項以下のみ」のデータで行なった。その結果を、図8～図10に示している。図8は共起頻度に基づく品詞マップの説明図、図9は共有率に基づく品詞マップの説明図、図10は予測比に基づく品詞マップの説明図である。図8～図10での線の太さは、その結ぶ品詞間での共起頻度、共有率、予測比の値の大きさに対応しており大きいものほど太い線で結ばれるようになっている。各品詞の表記は接頭辞、接尾辞のみ最初の二文字を使い、他は最初の一文字のみを用いている。各品詞の配置には Kohonenの自己組織化マップ(SOM)を用いた(Teuvo Kohonen, Self-organizing maps, 2nd Edition, (Springer, 1997)参照)。

10

【0040】

また、これらの品詞の配置のデータとしては、予測比のものを用いた。これは共起頻度、共有率を用いてできたマップよりも見やすそうだったからである。SOMの入力の行列は、前記馬らの文献を参考にして、各品詞間の非類似度(距離)を意味する行列を与えた。行列の非対角要素の品詞Aと品詞Bの非類似度としては以下の式5の値を用い、行列の対角要素は0とした。

20

【0041】

【数4】

$$\frac{1}{2} - \frac{\log_2 Er(A, B)}{2(1 + |\log_2 Er(A, B)|)} \quad \text{式 5}$$

【0042】

ただし、 $Er(A, B)$ は、式2で与えられる品詞A、Bの予測比である。この式5にはそれほど理論的根拠はなく、予測比が0のとき1、1のとき1/2、のとき0となる式として選んだ。13×13の2次元配列のSOMを用い、整列フェーズでは学習回数を10,000に学習率の初期値を0.1に近傍半径を13に、微調整フェーズでは学習回数を1,000,000に、学習率の初期値を0.01に、近傍半径を7にして自己組織化を行なった。その結果で間投詞と代名詞の位置を交換し各品詞を上下左右にすこし手で動かしたものを品詞の配置として利用した。この品詞の配置の移動は品詞間に線を結ぶ際、三つ以上のものが一直線にならぶと各線がどの品詞間のものかわからないため、そうならないように移動した。

30

【0043】

(可視化処理の説明)

図6の「二項以下のみ」の予測比のデータの例を説明する。リンクデータ作成部5に基データである「二項以下のみ」予測比のデータが入力され、リンクデータ作成部5では前記式5を用いリンクデータを作成し、データ入力部2に入力する。データ入力部2は、入力されたリンクデータを自己組織化マップ部1に渡す。リンクデータは次のようになる。

40

【0044】

リンクデータ：前 (前置詞)	接 (接続詞)	0.933777	
代 (代名詞)	接	0.889677	
副 (副詞)	接	0.835823	
副	前	0.799705	
代	間 (間投詞)	0.781262	
接尾 (接尾辞)	接頭 (接頭辞)	0.711054	
連 (連結形)	接頭	0.679438	10
動 (動詞)	助 (助動詞)	0.676753	
名 (名詞)	動	0.369459	
代	形 (形容詞)	0.257789	
副	代	0.230003	
名	間	0.213170	
前	形	0.205826	
動	間	0.204354	20
名	形	0.198059	
副	形	0.197814	
副	間	0.152055	
動	前	0.134897	
動	形	0.129903	
名	代	0.128399	
形	間	0.119920	30
名	助	0.119477	
副	接尾	0.118048	
接	形	0.110912	
名	前	0.110343	
動	代	0.101682	
名	接	0.101376	
名	副	0.070278	40
副	動	0.047658	

自己組織化マップ部 1 は、入力されたリンクデータだけを用い、各ノードの適切な二次元での配置の座標を決定する (自己組織化する)。自己組織化により、ノードの適切な二次元での配置の座標が次のように求まる。

【 0 0 4 5 】

ノードデータ :	5	13	代
	6	6	形
	1	0	助
	10	12	接
	10	0	接頭
	13	1	接尾
	13	9	前
	1	12	間
	2	3	動
	12	7	副
	0	7	名
	6	0	連

10

リンク表示部 4 は、入力されたリンク情報（リンクデータ）に基づき、リンク情報の線の太さで、そのリンクの二つのノードの座標位置を線で結ぶ。

20

【0046】

ノード表示部 3 は、入力されたノードデータに基づく、そのノードの座標位置に、ノードの記号を表示する（図 10 参照）。

【0047】

（考察）

図 6 より名詞 - 動詞の組で共起頻度が大きく、この品詞の組で多くの転換現象がおこっていることがわかる。また、名詞 - 形容詞の組でも共起頻度より転換現象が多いことがわかる。従来、あまり生産的でないとされていた名詞 - 形容詞の転換現象は、それほど少なくなくむしろ多いと、されているが、そういう名詞 - 形容詞の転換現象の多いことが本調査でもすぐにわかるのである。また、本調査では示さないが、本調査の単語辞書を用いた調査方法では、実際にその転換をしている単語の具体例も共起頻度の個数分、容易に取り出すことができる。ただし、本調査では転換の方向性はわからない。例えば、ある単語が名詞と形容詞の二つの品詞をもっている、どちらからどちらの品詞へ転換したかはわからない。これには注意しておいてほしい。

30

【0048】

共有率、予測比を見ていくとまた違った品詞の組で強い関係があることがわかる。予測比が大きいものとしては、副詞 - 接続詞や前置詞 - 接続詞がある。これらは品詞間の機能的な類似性が影響しているのであろう。

【0049】

また、図 6 では「すべて」の場合も示しているが、「すべて」の場合は、共起頻度に確率的な確からしさで重み付きで頻度を加えているものがあるので、厳密な値でなく、推測した値である。この結果を使う場合は注意が必要である。とはいえ、逆に「二項以下のみ」の場合は、すべての単語を用いていないという問題がある。

40

【0050】

図 7 より三項の組の強さとしては、共起頻度では名詞 - 形容詞 - 動詞が強く、予測比としては副詞 - 前置詞 - 接続詞が強いとわかる。ところで、これは可視化した図 8、図 10 を見てもわかることで、それぞれの図で太字の三角形を見ると上の組であることがわかる。可視化した図は、品詞三項のつながりの強さなど、より広い品詞間のつながりを考察するのに役立つのである。また、図での配置も興味深く、助動詞と動詞が近くに、また、連結形、接頭辞、接尾辞が近くに配置されるなどの構造が自動で得られている。

50

【0051】

可視化した図の作成には、自己組織化で作ったマップを基礎として用いたが、この品詞の配置も人手で行なうのは困難である。自己組織化マップの考え方は本調査のような調査研究の可視化にも役に立つのである。また、本調査では品詞をマッピングするだけでなく各品詞間を線でつなげ、各品詞間のつながりの強さを線の太さで表現したが、これにより品詞間の関係もより明確に示すことができた。

【0052】

(おわりに)

本調査では単語辞書を用いて英語品詞の転換現象の効率的な調査を行なった。本調査では詳細な調査結果の表とそれを可視化した図を示すことができた。これらの結果は転換現象などの言語の歴史的変遷を調べる研究の基礎的なデータとして役に立つものと思われる。特に可視化した図8～図10は、じっくりと見ているうちになにか深いものが見えてきそうである。

10

【0053】

(4)：形容詞が内包する抽象的意味の抽出と自動分類の試みの説明

本試みは、形容詞が内包する抽象的意味の様相を自己組織型意味マップ(SOM)を用いて求め、ノード表示部2とリンク表示部4で可視化を行うものである。また、本試みで述べる形容詞とは、形容詞と形容動詞を含める。

【0054】

形容詞の意味は、複数の抽象的な意味カテゴリーに渡る。例えば、「きつい坂道」の「きつい」は程度、感覚、場合によっては評価を表すし、また坂道のもつ特徴でもある。このように一語の形容詞でも、いろいろな意味を帯びるので、意味分類をしようと思っても、一つの意味カテゴリーに納まらないことが多い。形容詞の意味は、抽象的かつ多面的であるので意味分類することが難しい。従来形容詞の意味記述や分類は、格パターンや連体、終止、連用などの用法、あるいは接尾語をつけたり共起する名詞などの情報を利用し、類義語、対義語などを調べている(西尾寅弥, 形容詞の意味用法の記述的研究, 国立国語研究所, 秀英出版, 1972、情報処理振興事業協会技術センター, 計算機用, 日本語基本形容詞辞書IPAL 解説編 1991、参照)。

20

【0055】

「属性」や「感情」などは、一般的によく知られた形容詞の意味的特徴であるのだが、そのような語を用いずに、抽象名詞を利用することで、形容詞の抽象的な意味、つまり意味カテゴリーにあたるものをコーパスから抽出し自動分類する。

30

【0056】

従来我々の試みでは、限られた語数でSOMの配置に従ってその類似性を検討してきた。また、SOMの分類能力に関しても他手法との比較により劣っていないことも明らかにした(前記馬らの文献参照)。今回、構築した意味マップは、類似尺度として補完類似度を導入した。補完類似度に関しては山本・梅村(山本英子, 梅村恭司「コーパス中の一対多関係を推定する問題における類似尺度」自然言語処理, Vol. 9 No. 2, 2002 参照)で詳細に述べられているように、ある事象と事象が包含関係にある場合に強い類似尺度である。語どうしの関係も包含関係で関係付けてくれる可能性が高い。これと従来類似尺度を用いた意味マップとを比較、検討する。そして、これらの意味マップにおいて、語がどのような分布になっているのかを、マップ上での語全体の位置関係と類似度の関係の高い二語間の関係とを合わせて検討する。

40

【0057】

(形容詞の抽象的な意味を探るための意味の手がかりについて)

抽象的な名詞の統語的役割に着目した先行研究には、根本、高橋(根本今朝男「「が格」の名詞と形容詞とのくみあわせ」「電子計算機のための国語研究II」国立国語研究所, 1969、高橋太郎「文中にあらわれる所属関係の種々相」国語学103, 国語学会, 1975、参照)などがあげられる。例えば前記研究高橋(1975)においては、

1 やぎは性質がおとなしい

50

2 ぞうは鼻が長い

の二例を比較し、1 を側面語、2 を部分語と仮に呼び、文中の役割が異なることを述べている。側面語になる単語は主語の示すものや人の側面を表すとともに、述語の示す属性の類概念（上位概念）を表す単語である。また、前記研究根本（1969）においても「色が白い」「速さがはやい」「年が若い」「背が高い」などは、「顔が赤い」などのような状態の持ち主を表す場合と違って同義反復的な性格が強いと述べている。このように、我々の言語活動の中にも、形容詞の上位概念を示すような用法がみられる。

【0058】

神崎（神崎享子，井佐原均「形容詞類の連体用法にみられる連用的な意味」計量国語学，Vol. 22 No. 2，計量国語学会，1999、参照）において形容詞の意味関係の分類を行ったところ、主語述語関係に変換できるものの中で、「ゆるやかな傾斜」にみられる「名詞+「が」+形容詞」というタイプと主語述語に変換できないものの中で「悲しい気持ち」のようなタイプの2種類に、上記のような関係に近いものがみられる。つまり、形容詞が抽象名詞の具体的表現になっているものである。このタイプは、形容詞と抽象名詞が語彙的意味をお互いに共有していると考えられる。例えば、「白い色」は「色」という意味を共有し、「白い」が「色」の属性値であり、「色」は「白い」の上位概念である。また、「悲しい気持ち」においても、「悲しい」は「気持ち」の属性値であり、「気持ち」は「悲しい」の上位概念である。このようなパターンは、形容詞の抽象的意味をコーパスから探るのに重要な手がかりになるのではないかと考える。

【0059】

（データについて）

抽象名詞は、94、95年分の毎日新聞2年分から取り出した。抽象名詞と共起する形容詞、形容動詞は、毎日新聞11年分、日本経済新聞10年分、産業金融流通新聞7年分、読売新聞14年分、新潮文庫100選、新書版100冊の中から用例を調べた。抽出された抽象名詞は365語、形容詞の異なり語が10525語、のべ語数は35173語であった。最大共起語数1594語である。最初に作成されるのは、以下のような表である。

【0060】

思い： うれしい 楽しい 悲しい... ..
 気持ち： 楽しい 嬉しい 幸せな... ..
 観点： 医学的な 歴史的な 学術的な... ..

（入力データの符号化について）

上記データの単語のリストを可視化装置の自己組織化マップ部1であるSOMの入力とするには、符号化する必要がある（Qing Ma, Kyoko Kanzaki, Masaki Murata, Kiyotaka Uchimoto, and Hitoshi Isahara (2000) Self-Organization Semantic Maps of Japanese Noun in Terms of Adnominal Constituents, In Proceedings of IJCNN'2000, Como, Italy, vol. VI. 参照）。

まず、従来型の符号化について述べ、次に補完類似度の符号化について述べる。

【0061】

（前記Ma(2000)の符号化について）

SOMへの入力とするためには、単語のリストを符号化する必要がある（前記Ma(2000)）。

【0062】

ここで、一般に種類の名詞 w_i ($i = 1, \dots, n$)が存在し、それらの意味マップを構築すると仮定する。このような場合、名詞 w_i は以下のように連体修飾要素のセットで定義される。

【0063】

思い = { 悲しい、楽しい、幸せな、... }

【0064】

【数5】

$$w_i = \{ a_1^{(i)}, a_2^{(i)}, \dots, a_{\alpha_i}^{(i)} \}$$

【0065】

ただし、 $a_j^{(i)}$ は、 w_i と共起する j 番目の連体修飾要素で、 i は w_i と共起する連体修飾要素の数である。これを符号化するためにここでは、名詞 w_i と w_j 間の距離 d_{ij} を以下のような計算式によって求めることとした。

【0066】

【数6】

$$d_{ij} = \begin{cases} \frac{(\alpha_i - c_{ij}) + (\alpha_j - c_{ij})}{\alpha_i + \alpha_j - c_{ij}} & i \neq j \text{ の場合} \\ 0 & i = j \text{ の場合} \end{cases}$$

20

【0067】

ここで、 i と j はそれぞれ w_i 、 w_j と共起する連体修飾要素の総数で、 c_{ij} は w_i 、 w_j に共通する連体修飾要素の数である。上の式は、意味的關係 d_{ij} は、 w_i 、 w_j の間にどのくらい共通する連体修飾要素があるのかということを表す正規化された距離である。すなわち、 d_{ij} が大きければ意味的な距離は遠く、 d_{ij} が小さければ意味的な距離は近くなる。前記 Ma (2000) では、この類似計算のあと「相関コーディング法」を用いる。ここで提案する相関コーディング法では、名詞 w_i をこの行列を用いて以下のような多次元ベクトルに符号化する。

【0068】

【数7】

$$V(w_i) = [d_{i1}, d_{i2}, \dots, d_{i\omega}]^T$$

30

【0069】

$v(w_i)$ は SOM への入力であり、この多次元ベクトルを自己組織化によって、それらの間に存在する意味関係を顕在化し二次元空間に表現する。

【0070】

(補完類似度を用いた符号化について)

前記山本・梅村(2002)の補完類似度は、包含関係を取り出すことを得意とする類似尺度である。これを SOM の入力データの符号化に用いた。前記山本・梅村(2002)によれば、補完類似度は以下のような式になる。

【0071】

【数8】

40

今、 $\vec{F} = (f_1, f_2, \dots, f_i, \dots, f_n)$ ($f_i = 0$ または1)、
 $\vec{T} = (t_1, t_2, \dots, t_i, \dots, t_n)$ ($t_i = 0$ または1)
 とする。

$$Sc(\vec{F}, \vec{T}) = \frac{ad - bc}{\sqrt{(a+c)(b+d)}}$$

10

【0072】

ここで、aは、二つのラベルが同時に現れるデータの数、bは1 a b 1が現れ、1 a b 2は現れないデータの数、cは、1 a b 2が現れ、1 a b 1は現れないデータの数、dは、二つのラベルがどちらとも現れないデータの数である。本データに対してこの尺度を用いる際には、1 a b にあたるのが抽象名詞となり、aは、ある形容詞が二語の抽象名詞と共起しているパターン、bとcは、ある形容詞がそれぞれ一方の抽象名詞とだけ共起しているパターン、dは形容詞が両者とも共起していないパターンということになる。補完類似度の数値を正規化し、それをもとに前記Ma(2000)の符号化について述べた相関コーディング法を用いて抽象名詞を多次元ベクトルに変換して符号化する。

20

【0073】

(ベースラインの意味マップと補完類似度を符号化に用いた意味マップについて)

ここでは、従来型の意味マップ(ベースラインと呼ぶ)と補完類似度をを用いた符号化による意味マップとを示す。

【0074】

まず、分類結果の評価に関して、一つの座標上に配置された名詞の、共通する共起形容詞数で判断した。意味マップを見たときに、少なくとも同じ座標上にあるものは、かなり類似していなければ、直感的に類似している単語が近くに配置されているとは考えにくい。

【0075】

次に両マップでグループの変わらない単語を調べる。両方のマップで、単語どうしがいつも同じ位置あるいは比較的近い位置にあれば、その単語の類似度は高い可能性がある。

30

【0076】

最後に、単語がどのようにマップ上に配置されているのかについて調べる。意味マップ上に配置されている単語は、全体の中での類似性の位置付けであるが、そこに、SOMの符号化の際に計算した二語間の類似度計算で値の高いものから語どうしにリンクをはることにした。これによってある語の周辺の語との関係の強さがわかるので、マップ上の語どうしの関係性が視覚的にわかる。

【0077】

意味マップは整列フェーズと微調整フェーズからなる。ベースラインの意味マップでは、整列フェーズが1万回、微調整フェーズは5万回の学習で得られたものである。座標は45 × 45で、半径7である。

40

【0078】

一方、補完類似度をを用いた意味マップの方は、ベースラインと同じ学習回数だと学習不足で、365語の抽象名詞が意味マップ全体にランダムに散らばった状態になった。そこで学習回数を増やし、整列フェーズは3万回、微調整フェーズは10万回にした。座標と半径の数値はベースラインと同じ45 × 45で、半径7である(図11、図12参照)。

【0079】

(同じ座標上の語数の説明)

図11はベースラインの意味マップの説明図であり、図12は補完類似度をを用いた意味マップの説明図である。同じ座標に集まっている名詞をトータルで見るとベースラインの意

50

味マップの方が圧倒的に名詞がまとまっている。ベースラインの意味マップでは140語、補完類似度を用いた意味マップでは104語の名詞がところどころで座標を同じくしている。ベースラインの意味マップは、複数の抽象名詞が固まっているのに対し、補完類似度のマップは抽象名詞が散在している、というイメージになる。

【0080】

では、同じ座標上に複数の抽象名詞がある場合、どれくらい意味的に類似しているのか。それを調べるために、同じ座標上の複数の抽象名詞に共通する共起語数は、一つの抽象名詞の全共起語数の何割を占めているかを求めた(表2参照)。同じ座標上であれば、未分類であるか、とても意味が近いかのどちらかであると思われる。ベースラインも補完類似度も、共起語の重なる割合がほとんどが0%から30%の間である。これは、お互いの関係が薄いか、あるいは、同じ座標上の複数の名詞の中で、共通する形容詞はあるものの比較的共起語が多いので重なる割合が低くなる名詞であるかのどちらかの可能性がある。

10

【0081】

例えば、ベースラインの意味マップで、同じ座標に複数の名詞がありながら、すべてに共通する形容詞がない場合(0%)をみると、「間隔・思い込み」や「圧力・線・値・都合」などがある。これらの名詞の意味を考えると直感的にも座標上同じ位置を許すほどの類義語であるとは思にくい。同じ座標上の複数の名詞の中で、比較的共起語が多いために重なる割合が低くなっている例としては、たとえば「色・色彩・彩り」は一つのグループであるが、「色」の共起語が比較的他の2つの抽象名詞と比べて多いため、形容詞の重なる割合が10%程度となっている。全体的には、共通する共起語が10%未満の場合には、名詞は直感的にもうまく分類されていない場合が多い。ベースラインの意味マップと補完類似度を類似尺度として用いた意味マップとを比べると、両者とも、同じ座標上に複数の単語が分類されていても共通する共起語の割合がそれほど高くはないが、両者の比較をするならば、ベースラインの方が、同じ座標上に共起語に重なりのある単語が位置しているので、うまく分類されているイメージになる。

20

【0082】

ここで、もう一度、前の2つのマップ(図11、図12)を比較すると、補完類似度の意味マップはそもそも単語が散在している。もし、単語を一つ一つ厳密にマップ上に配置するとしたら、同義語以外には同じ座標に置かれなわけなので、そういう意味では、補完類似度の意味マップをもう少し検討してみる必要がある。

30

【0083】

表2 同じ座標に位置する抽象名詞の、重なる割合ごとの個数

重なる割合(%)	ベースライン	補完類似度
0-1	22	16
1-10	27	21
10-20	40	22
20-30	17	21
30-40	10	10
40-50	4	2
50-60	10	5
60-70	0	0
70-80	0	2
80-90	1	1
90-100	0	0
100	9	4

10

20

【0084】

(両方のマップに共通する語のセットの説明)

ベースラインの意味マップ上で同じ座標もしくは近辺にある単語群は、補完類似度を用いた意味マップ上では、どのようになっているかを調査する。どの尺度を使っても安定している単語のグループは、強い類似性を示す可能性が高い。前述のように、ベースラインの意味マップで同じ座標に位置する抽象名詞の数は140語であった。この140語の名詞のうち補完類似度の意味マップにおいて近い位置にプロットされたもの(半径2以内)は87語あった。半径3以内を含めると96語になる。両方のマップに共通して近い位置にある87語をあげると、次のようになる。

30

【0085】

「幸福感 心遣い 情愛 配慮 思いやり 温かさ 気立て」「願い 情熱」「言葉 意見 評価」「触感 手触り 感受性 感性 舌ざわり」「におい2 香り1 味覚 味」「色2 色彩2 彩り1」「活力 若さ」「緊張 緊張関係」「かかわり つながり」「傾斜 勾配」「核 核心 眼目 急所 骨子 重み 正念場」「家系 血筋 血統 家柄」「昔 大昔 老舗」「円 曲線 図形 面3」「奥行き 空間 面積」「歳月 格好2」「勢い 速度 時間 時刻 期間」「強度 力2 語気」「品格 階級」「数量」「音程 角度2」「利点 順番」「順 程度 可能性」「うち1 中1 一方」「時 状態 方」「ところ1 イメージ 印象 面」「美しさ 魅力 人柄」

40

以上の結果をみると、補完類似度を用いた意味マップでは、同じ座標上の単語は少ないが、類似している単語は基本的には近くに位置していることがわかる。

【0086】

(意味マップの単語間の関係の説明)

名詞の意味マップは相関行列値に基づいた類似度によるので、ある名詞の全体の中での類似関係が求められたのだが、この単語のマップを二語間の類似度(リンクデータ)に従って更にリンクさせると、ある語の周辺の語との直接的な関係の強さがわかる。ここに、リンク表示部4によるリンクつきのベースラインのマップと補完類似度を用いたマップをあ

50

げる。ただし、類似度の低いものもすべてリンクしてしまうとマップは線で埋め尽くされてしまうので、見やすさの便宜のため類似度の値が高いものだけリンクを張っている。そして、このリンクは類似度の値により太さを変えて表示することもできる。

【0087】

図13はベースラインの二語間の類似関係の説明図であり、図14は補完類似度の二語間の類似関係の説明図である。図13と図14を比較すると、ベースラインの意味マップは右上と右下のコーナーのあたりに線が密集しているが、基本的には、全体の単語の配置と二語間の類似度との対応は説明しがたい。補完類似度を用いた意味マップの方は右下を基点に左上へ放射状に単語が配置されている。補完類似度の特徴である包含関係が、マップ全体の単語の配置に反映されていることがわかる。先に同じ座標軸上の複数の名詞の共起語を比較した結果、両マップに劇的な分類結果の差が出たわけではなかった。同じ座標上になくても、近くの座標に類似した語はプロットされていた。従って、意味マップ上の名詞の分類がどのような結果になったかについては、マップ全体の単語の配置に包含関係を反映した補完類似度を用いた意味マップがわかりやすい。

10

【0088】

(まとめ)

図14の補完類似度による意味マップは、右下が共起語が最多の「こと」という抽象名詞が配置されており、「こと」を基点として単語の広がり方に方向性がみられる。「状態」「方向」「意味」「様子」「感覚」「印象」などの抽象名詞が「こと」の近くに位置しており、更にそれを基点に放射線状に広がっている。右下から離れるほど名詞が具体化していく。マップの右上には、「観点」や「立場」などの抽象名詞があるが、これらは、「意味」と強い包含関係を示している。抽象名詞の広がりを解説すると次のようになる。「状態」から「段階」「程度」「数」「量」「時間」「距離」「空間」「奥行き」と分布していく。「方向」から「傾向」「兆候」「影響」「評価」などへと広がり、「状態」と「方向」の広がりの中には「局面」「状況」「情勢」などが位置している。「様子」の方向には、「顔つき」「しぐさ」「そぶり」などがあり、「感覚」は、「気持ち」と「感触」、「印象」などと関係を強くし、「印象」は「感覚」や「人柄」などの人やものの特徴などと関係していく。一般的に、程度を表す「傾斜」などは「評価」の近くに、「温度」は「思いやり」「情熱」「気候」などの近くに、「弾力性」は「強さ」や「根性」などの近くに、「触感」や「味覚」「色」などは「感受性」の近くに「感受性」は人の特徴を表す抽象名詞の近くに位置している。得られたマップから形容詞は、「状態」「方向」「意味」「様子」「感覚」「印象」などの上位概念ではほとんどの形容詞が共起しうることから、形容詞は、一般的にこれらの抽象概念を複合的に持っていると考えられる。マップ上では、語の散らばり方に方向性があり、上記の抽象概念を基点に右下隅から左上へと放射線上に上位から下位概念へと広がっている。

20

30

【0089】

(5) : 方向性のあるデータの可視化の説明

国間の貿易(輸出、輸入)等のように、リンクデータに方向性のある場合がある。例えば、ノードA、Bがあって、ノードAからノードBへの量が x 、ノードBからノードAへの量が y の場合、ノードA側では x の太さで、ノードB側では y の太さの、台形状の線でノードA、Bを結ぶ。これにより、二つのノード間の方向と量を一つの線で表現することができる。

40

【0090】

図15は方向性のある量の表現の説明図である。図15(a)~図15(c)は、方向性のある量を表現する3つの例を示しており、ノードAからノードBへの量が線の太さ x 、ノードBからノードAへの量が線の太さ y となるものである。図15(a)は、ノードの端における線の太さで方向性のある量を表現する場合の例である。図15(b)は、ノードの中心における線の太さで量を表現する場合の例であり、ノードを前面(優先して)表示したものである。図15(c)は、ノードの中心における線の太さで量を表現する場合の例であり、線を前面(優先して)表示したものである。このように、図15の例だけに

50

限られるものではないが、図15(a)～図15(c)のいずれか一つを用いて方向性のある量の表現を行うことができる。

【0091】

以下、鉄鋼の輸出の例で説明する。図16は鉄鋼の輸出の説明図である。図16において、データ入力部2にノードデータ(ノード情報)とリンクデータ(リンク情報)が次のように与えられ、ノード表示部3とリンク表示部4で表示したものである。ここでノードデータは、横軸の座標、縦軸の座標とノードの記号となり、リンクデータは、ノードの記号1、ノードの記号2と線の太さ1(ノードの記号1側での太さ)、線の太さ2(ノードの記号2側での太さ)となる。

【0092】

ノードデータ：14 1 カナダ
 14 4 米国
 10 5 日本
 8 5 韓国
 7 7 中国
 4 1 ロシア
 3 7 中東
 1 3 EC

リンクデータ：カナダ 米国 0.2 0
 日本 米国 0.5 0
 EC 米国 0.7 0
 韓国 日本 0.2 0.2
 日本 ロシア 0.2 0
 日本 中国 1 0
 日本 中東 0.3 0
 EC ロシア 0.5 0
 EC 中国 0.2 0
 EC 中東 0.5 0

なお、例えば、上記カナダと米国のリンクデータは、カナダから米国への鉄鋼の輸出量は0.2、米国からカナダへの鉄鋼の輸出量は0であることを示している。

【0093】

この例では、鉄鋼の輸出のだいたいの流れが人目で見れて便利である。また、逆方向のデータも一つの線で表現でき便利です。例えば、日本と韓国の間は、輸出は同じで同じだけ輸出して同じだけ輸入しあっていることが、線の太さが一定の線で結ばれていることからわかる。逆方向の線を一つの線で表さない場合は、それぞれ二つの線が必要になり図が煩雑になりみにくくなる。

【0094】

図17は方向性のあるデータの可視化処理フローチャートである。以下、図17の処理S21～S23に従って説明する。

【0095】

S21：データ入力部2に、次のような書式で、ノードデータ及びリンクデータ(方向性

10

20

30

40

50

がある)を入力する。

【0096】

ノードデータ(ノード情報)：

横軸の座標、縦軸の座標、ノードの記号

横軸の座標、縦軸の座標、ノードの記号

...

横軸の座標、縦軸の座標、ノードの記号

リンクデータ(リンク情報)：

ノードの記号1、ノードの記号2、線の太さ1、線の太さ2

ノードの記号1、ノードの記号2、線の太さ1、線の太さ2

ノードの記号1、ノードの記号2、線の太さ1、線の太さ2

...

ノードの記号1、ノードの記号2、線の太さ1、線の太さ2

S22：リンク表示部4は、入力されたリンク情報に基づき、そのリンクの二つのノードの座標位置を線で結ぶ。ノード座標位置は、ノード情報から取得する。リンクの線の太さは、そのリンクのノードの記号1の場所で線の太さ1に、そのリンクのノードの記号2の場所で線の太さ2になるように設定し、リンクの線はそれらの太さを満足する台形により描く。

【0097】

S23：ノード表示部3は、入力されたノード情報に基づく、そのノードの座標位置に、ノードの記号を表示する。

【0098】

なお、方向性のあるデータの場合も、データ入力部2にリンクデータのみを入力し、自己組織化マップ部1によりノードデータを作成することもできる。

【0099】

(方向性ありのデータの場合の自己組織化マップの利用の説明)

方向性ありのデータの場合はリンクのデータとしては、片一方からもう一方への値と、もう一方から片一方への値の二つがあり、方向性なしのデータでの自己組織化マップの利用とは状況がやや異なる。自己組織化マップ部1の利用方法としては、単純には以下の1、2の二つが考えられる。

【0100】

1 二つの値を一つの値に融合してから方向性なしのデータの場合と同じ方法を用いる。

【0101】

例えば、二つの値の相加平均値(足して2で割ったもの)をそれぞれのリンクの値として方向性なしのデータの場合と同じ処理をすればよい。なお、二つの値の融合方法には、この相加平均値以外に、相乗平均値(かけて平方根をとったもの)を使うなど、種々の方法が考えられる。

【0102】

2 リンクデータとして二つの値を用いて、これを自己組織化マップ部1の入力とする。

【0103】

例えば、ノードiからノードjへの値が $w(i, j)$ である場合は、入力の行列は以下のようになる。

【0104】

【数9】

10

20

30

40

$$\left(\begin{array}{cccccccc} 0 & 0 & w(1,2) & w(2,1) & w(1,3) & w(3,1) & \dots & w(1,n) & w(n,1) \\ w(2,1) & w(1,2) & 0 & 0 & w(2,3) & w(3,2) & \dots & w(2,n) & w(n,2) \\ w(3,1) & w(1,3) & w(3,2) & w(2,3) & 0 & 0 & \dots & w(3,n) & w(n,3) \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ w(n,1) & w(1,n) & w(n,2) & w(2,n) & w(n,3) & w(3,n) & \dots & 0 & 0 \end{array} \right)$$

10

【0105】

各行ベクトルをそれぞれのノードのベクトルとして自己組織化マップのプログラムを動作させ、各ノードの配置位置を求める（なお、自己組織化マップアルゴリズムの入力の行列は正方行列でないといけないということはない）。

【0106】

また、ノードの位置を自動で求めるのに Kohonen の自己組織化マップを用いる説明をしたが、これ以外に、ノード間のリンクがもつ値の大きさをバネの強さと仮定して求めるバネモデル方式、多変量解析配置方法等を用いることができる。

【0107】

(6) : プログラムインストールの説明

自己組織化マップ部 1、自己組織化マップ手段 1 a、データ入力部 2、入力手段 2 a、ノード表示部 3、ノード表示手段 3 a、リンク表示部 4、リンク表示手段 4 a、リンクデータ作成部 5 等は、プログラムで構成でき、主制御部 (CPU) が実行するものであり、主記憶に格納されているものである。このプログラムは、一般的な、コンピュータで処理されるものである。このコンピュータは、主制御部、主記憶、ファイル装置、表示装置、キーボード等の入力手段である入力装置などのハードウェアで構成されている。このコンピュータに、本発明のプログラムをインストールする。このインストールは、フロッピー、光磁気ディスク等の可搬型の記録 (記憶) 媒体に、これらのプログラムを記憶させておき、コンピュータが備えている記録媒体に対して、アクセスするためのドライブ装置を介して、或いは、LAN 等のネットワークを介して、コンピュータに設けられたファイル装置にインストールされる。そして、このファイル装置から処理に必要なプログラムステップを主記憶に読み出し、主制御部が実行するものである。

20

30

【0108】

【発明の効果】

以上説明したように、本発明によれば、次のような効果がある。

【0109】

(1) : 自己組織化マップ手段でリンクデータにより各ノードの配置座標位置を求め、リンク表示手段で前記リンクデータのリンクがもつ値の大きさによる線の太さで二つのノードの配置座標位置を線で結んで表示し、ノード表示手段で前記求めた各ノードの配置座標位置にノードを表示するため、複数のノード間のリンク関係とリンクの大きさを明確に示すことができる。

40

【0110】

(2) : リンク表示手段で、リンクデータのリンクがもつ値が一定値以上のもののみ、二つのノードの配置座標位置を線で結んで表示するため、表示するノードが多い場合のノード間のリンク関係とリンクの大きさを明確に示すことができる。

【0111】

(3) : リンク表示手段で、二つのノードを線で結ぶ際、一方のノード側の線の太さを一方のノードから他方のノードへのリンクがもつ値に基づいた線の太さにし、他方のノード側の線の太さを他方のノードから一方のノードへのリンクがもつ値に基づいた線の太さに

50

するため、ノード間で方向性のあるリンクデータを一つの線で表示することができる。

【0112】

(4)：自己組織化マップ手段でリンクデータにより各ノードの配置座標位置を求め、リンク表示手段で二つのノードを線で結ぶ際、一方のノード側の線の太さを一方のノードから他方のノードへのリンクがもつ値に基づいた線の太さにし、他方のノード側の線の太さを他方のノードから一方のノードへのリンクがもつ値に基づいた線の太さにするため、入力手段にリンクデータを入力するだけで、ノードの表示とノード間で方向性のあるデータを一つの線で表示することができる。

【0113】

(5)：複数のノード間のリンクデータを入力する入力手段と、前記リンクデータにより各ノードの配置座標位置を求める自己組織化マップ手段と、前記リンクデータのリンクがもつ値の大きさによる線の太さで二つのノードの配置座標位置を線で結んで表示するリンク表示手段と、前記求めた各ノードの配置座標位置にノードを表示するノード表示手段として、コンピュータを機能させるためのプログラム又はプログラムを記録したコンピュータ読取可能な記録媒体とするため、このプログラムをコンピュータにインストールすることで、複数のノード間のリンク関係とリンクの大きさを明確に示すことができる可視化装置を容易に提供することができる。

【0114】

(6)：複数のノード間のリンクデータと前記各ノードの配置座標を入力する入力手段と、前記入力された各ノードの配置座標位置にノードを表示するノード表示手段と、二つのノードを線で結ぶ際、一方のノード側の線の太さを一方のノードから他方のノードへのリンクがもつ値に基づいた線の太さにし、他方のノード側の線の太さを他方のノードから一方のノードへのリンクがもつ値に基づいた線の太さにする、リンクをもつ二つのノードの配置座標位置を線で結んでリンクを表示するリンク表示手段として、コンピュータを機能させるためのプログラム又はプログラムを記録したコンピュータ読取可能な記録媒体とするため、このプログラムをコンピュータにインストールすることで、ノード間で方向性のあるデータを一つの線で表示することができる可視化装置を容易に提供することができる。

【図面の簡単な説明】

【図1】本発明の原理説明図である。

【図2】実施の形態における可視化装置の説明図である。

【図3】実施の形態における可視化処理フローチャートである。

【図4】実施の形態における自己組織化マップ部を備える可視化装置の説明図である。

【図5】実施の形態における自己組織化マップ部を備える可視化処理フローチャートである。

【図6】実施の形態における品詞二項間の共有性の調査結果の説明図である。

【図7】実施の形態における品詞三項間の共有性の調査結果の説明図である。

【図8】実施の形態における共起頻度に基づく品詞マップの説明図である。

【図9】実施の形態における共有率に基づく品詞マップの説明図である。

【図10】実施の形態における予測比に基づく品詞マップの説明図である。

【図11】実施の形態におけるベースラインの意味マップの説明図である。

【図12】実施の形態における補完類似度を用いた意味マップの説明図である。

【図13】実施の形態におけるベースラインの二語間の類似関係の説明図である。

【図14】実施の形態における補完類似度の二語間の類似関係の説明図である。

【図15】実施の形態における方向性のある量の表現の説明図である。

【図16】実施の形態における鉄鋼の輸出の説明図である。

【図17】実施の形態における方向性のあるデータの可視化処理フローチャートである。

【符号の説明】

1 a 自己組織化マップ手段

2 a 入力手段

10

20

30

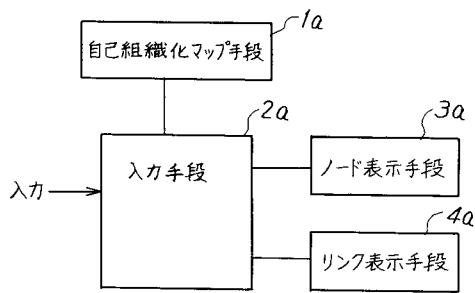
40

50

- 3 a ノード表示手段
- 4 a リンク表示手段

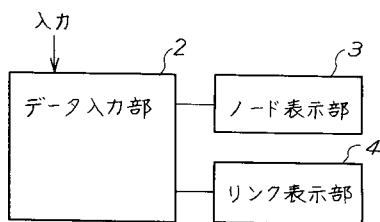
【図1】

本発明の原理説明図



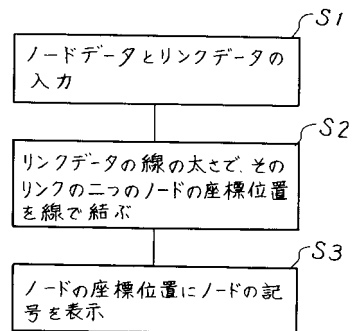
【図2】

可視化装置の説明図



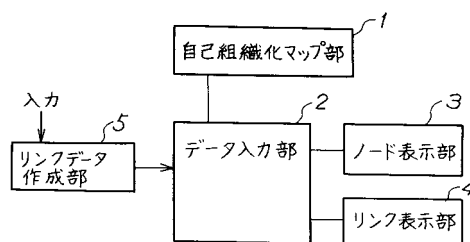
【図3】

可視化処理フローチャート



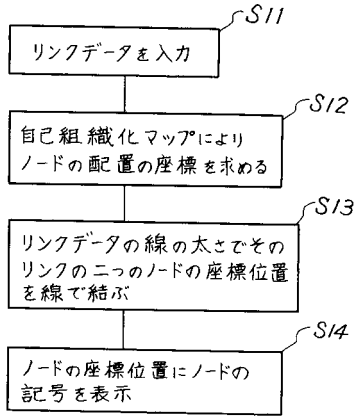
【図4】

自己組織化マップ部を備える可視化装置の説明図



【図5】

自己組織化マップ部を備える可視化処理フローチャート



【図6】

品詞二項間の共有性の調査結果の説明図

品詞二項の組	二項以下のみ			すべて		
	共起頻度	共有率 (%)	予測比	共起頻度	共有率 (%)	予測比
名詞 形容詞	3061	6.53	0.35	3494.9	7.29	0.37
名詞 動詞	4177	10.33	0.78	4514.6	10.90	0.79
名詞 副詞	46	0.11	0.01	255.6	0.61	0.07
名詞 間投詞	57	0.16	0.39	102.8	0.28	0.52
名詞 前置詞	5	0.01	0.09	24.9	0.07	0.26
名詞 代名詞	9	0.02	0.13	14.8	0.04	0.18
名詞 接続詞	2	0.01	0.07	6.8	0.02	0.15
名詞 助動詞	2	0.01	0.11	4.7	0.01	0.22
形容詞 動詞	276	1.28	0.14	570.1	2.54	0.26
形容詞 副詞	412	2.28	0.35	619.7	3.27	0.47
形容詞 間投詞	6	0.04	0.11	18.3	0.13	0.24
形容詞 前置詞	8	0.06	0.37	27.6	0.19	0.76
形容詞 代名詞	13	0.10	0.52	23.0	0.16	0.73
形容詞 接続詞	1	0.01	0.09	5.8	0.04	0.33
動詞 副詞	1	0.01	0.00	70.4	0.50	0.09
動詞 間投詞	12	0.14	0.37	51.3	0.57	1.10
動詞 前置詞	2	0.02	0.15	6.9	0.08	0.31
動詞 代名詞	1	0.01	0.07	2.0	0.02	0.10
動詞 接続詞	0	0.00	0.00	1.8	0.02	0.17
動詞 助動詞	6	0.07	1.46	8.7	0.10	1.76
副詞 間投詞	4	0.08	0.20	14.7	0.26	0.52
副詞 接尾辞	2	0.04	0.11	2.0	0.04	0.10
副詞 前置詞	22	0.44	2.82	47.7	0.88	3.53
副詞 代名詞	4	0.08	0.44	15.2	0.28	1.29
副詞 接続詞	17	0.34	4.13	28.3	0.53	4.28
間投詞 代名詞	1	0.31	2.44	1.5	0.35	2.24
間投詞 接続詞	0	0.00	0.00	0.7	0.18	1.77
連結形 接頭辞	1	0.25	1.47	1.0	0.25	1.50
接尾辞 接頭辞	1	0.27	1.66	1.0	0.27	1.68
前置詞 代名詞	0	0.00	0.00	1.3	0.49	4.15
前置詞 接続詞	7	5.43	93.72	15.5	7.73	85.57
代名詞 接続詞	1	0.67	11.57	3.3	1.72	21.18
調査単語総数	56004			56837		

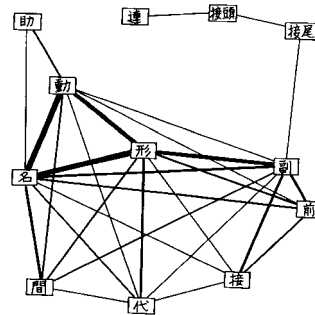
【図7】

品詞三項間の共有性の調査結果の説明図

品詞三項の組	三項以下のみ		
	共起頻度	共有率 (%)	予測比
名詞 形容詞 動詞	362	0.697	0.26
名詞 形容詞 副詞	198	0.376	0.23
名詞 形容詞 間投詞	4	0.008	0.09
名詞 形容詞 前置詞	6	0.012	0.31
名詞 形容詞 代名詞	3	0.006	0.16
名詞 形容詞 接続詞	1	0.002	0.10
名詞 動詞 副詞	27	0.059	0.05
名詞 動詞 間投詞	49	0.119	1.74
名詞 動詞 前置詞	1	0.002	0.08
名詞 動詞 代名詞	1	0.002	0.08
名詞 動詞 助動詞	4	0.010	1.25
名詞 副詞 間投詞	3	0.007	0.18
名詞 副詞 前置詞	7	0.017	0.98
名詞 副詞 代名詞	2	0.005	0.28
形容詞 動詞 副詞	17	0.064	0.09
形容詞 動詞 間投詞	1	0.004	0.09
形容詞 動詞 前置詞	8	0.043	2.94
形容詞 動詞 代名詞	9	0.048	3.34
形容詞 動詞 接続詞	2	0.011	1.48
動詞 前置詞 接続詞	1	0.011	52.80
副詞 間投詞 接続詞	1	0.018	37.40
副詞 前置詞 接続詞	7	0.132	618.22
調査単語総数	56718		

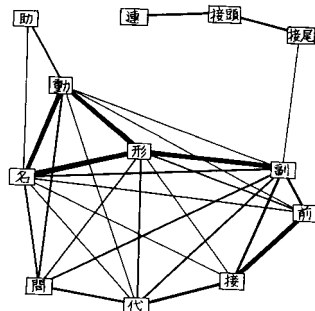
【図8】

共起頻度に基づく品詞マップの説明図



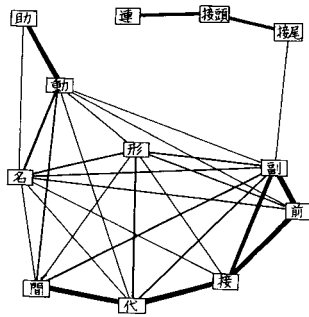
【図9】

共有率に基づく品詞マップの説明図



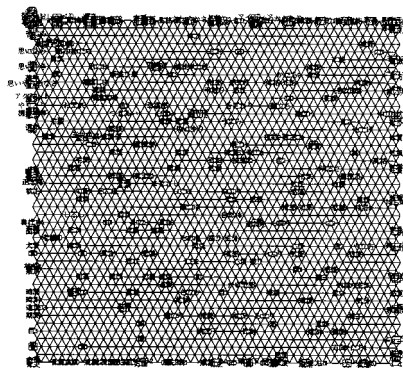
【 図 1 0 】

予測比に基づく品詞マップの説明図



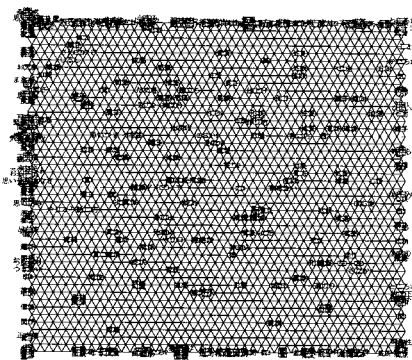
【 図 1 2 】

補完類似度を用いた意味マップの説明図



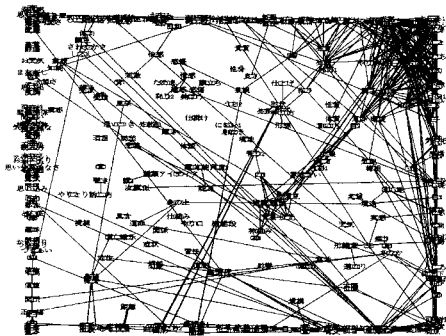
【 図 1 1 】

ベースラインの意味マップの説明図



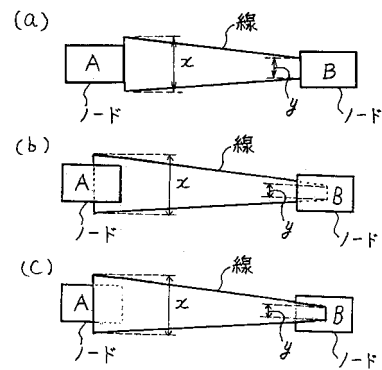
【 図 1 3 】

ベースラインの二語間の類似関係の説明図



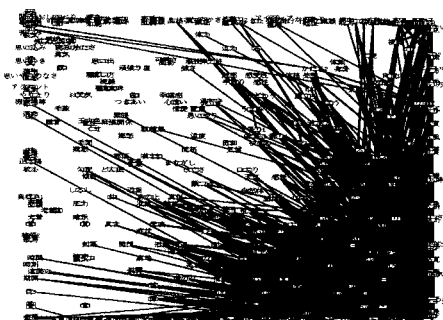
【 図 1 5 】

方向性のある量の表現の説明図



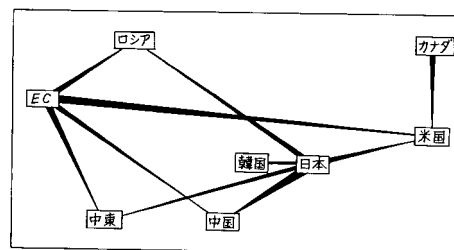
【 図 1 4 】

補完類似度の二語間の類似関係の説明図



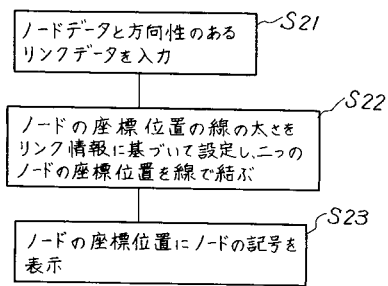
【 図 1 6 】

鉄鋼の輸出の説明図



【 図 1 7 】

方向性のあるデータの可視化処理フローチャート



フロントページの続き

- (72)発明者 馬 青
東京都小金井市貫井北町4 - 2 - 1 独立行政法人通信総合研究所内
- (72)発明者 進藤 三佳
東京都小金井市貫井北町4 - 2 - 1 独立行政法人通信総合研究所内
- (72)発明者 山本 英子
東京都小金井市貫井北町4 - 2 - 1 独立行政法人通信総合研究所内
- (72)発明者 井佐原 均
東京都小金井市貫井北町4 - 2 - 1 独立行政法人通信総合研究所内
- Fターム(参考) 5B056 BB22 BB62 BB66