

複素重み付き意味ネットワークの提案とテキスト要約への応用

佐藤 裕介[†] 松田 聖^{††}

[†] 日本大学大学院生産工学研究科数理工学専攻

^{††} 日本大学生産工学部数理情報工学科

概要：1968年にQuillianによって意味ネットワークが提案されて以来、様々な形式の意味ネットワークモデルが提案されてきた。しかしながら、それぞれのモデルが表現する arc は一つの関係しか表現できず、多数の関係を持つネットワークモデルでは arc の数が増大してしまう。また、多数の意味ネットワークモデルが存在し、一般的なネットワークモデルはあまり提案されてきていない。そのため、本論では、意味ネットワークの arc に対して重みを付加したネットワークモデルを提案する。この重みは複素数を利用して表現された複素重みであり、複素重みを使用することで、ノード間の連結を複数表現することが出来る。これらの複素重みは無作為に選んだテキストデータから得る。これにより、一般的な知識の獲得を目指す。さらに、提案したモデルの有効性を検証するためにテキスト要約に応用する。

Complex Weighted Semantic Network and Its application to text wummurization

Yusuke SATO[†] Satoshi MATSUDA^{††}

[†] Graduate Department of Mathematical Engineering, Postgraduate of Industrial Technology Nihon University

^{††} Department of Mathematical information Engineering, College of Industrial Technology Nihon University

Abstract: As for the Semantic Network proposed by Quillian in 1968, the various forms which should be done were proposed. However, each model can be expressed with arc that it is expressed only one relation. The number of arc increases in the network model which has many relations. And, Though many meaning network models were proposed, a general network model hasn't been proposed very much. Thus, we proposed the new network model which was used the weight given to arc of the semantic network with main subject. This weight is a complex number. Accordingly not only arc can show strength related to the meaning to express it but also we can evaluate the transition cause from the transition point. It aims at the acquisition of the general knowledge by getting this complex weight from the text-data chosen at random. And we do the text summary which used a semantic network with the weight to verify the validity of the network model.

1 はじめに

人間が生活を営む上で、「知識」は必要不可欠なものである。「知識」とは、物を知りそれを生活に利用するための情報である。この情報を持っているものは環境に対してすばやく対応し、自分を変化させることによって、環境に適応させていくことが出来る。また「知識」とは「知」を保持し、有効活用するための形式であると言える。計算機上でそのような知識を実現させることが人工

知能における最終目的と言われる。人工知能の研究分野のひとつとして知識表現と呼ばれる分野がある。知識表現とは知識を計算機上で表現するという分野であり、さまざまな表現形式が提案、研究されてきた。その中のひとつの手法として意味ネットワーク (Semantic Network: SN) が挙げられる。意味ネットワークは、概念同士をその意味関係で連結しネットワーク状に表現した手法であり、近年ではハードウェア化 [1] [2] [3] や文章生成 [4] など様々な技術に応用されている。また、

様々なテキストから意味的な関係を抽出する手法として Li らの研究 [11] が挙げられる。これらに対し、本論文では意味ネットワークの arc に対して付加する重みとして複素重みを用いる。これにより、node 間に複数の意味的關係が存在する場合でも一つの arc で表現することが出来る。また、無作為に選定したテキストデータから、知識を獲得することで一般的な知識の獲得を目指す。

第 2 節では、本論文で提案するモデルについて詳しく説明し、獲得した重みの評価を行なう。また、第 3 節では提案するモデルの実用例としてテキスト要約に応用した。これは、意味ネットワークは文章理解、文章作成などの応用分野でも研究され、意味ネットワークが自然言語処理を行なう上で、十分に有効であるということが分かっているため、本研究で提案するモデルが意味ネットワークが有効であるフィールドで有効であるかどうかを判断する。第 3 節では、そのテキスト要約に関する基本概念、システム構成、結果を提示し、最後に考察する。

2 複素重み付き意味ネットワーク

2.1 基本概念

本論文で提案する複素重み付き意味ネットワーク (Complex Weighted Semantic Network: CWSN) は、意味ネットワークの arc に対して、重みを付加させたネットワークモデルである。図 6 に複素重み付き意味ネットワークの基本構造を示す。

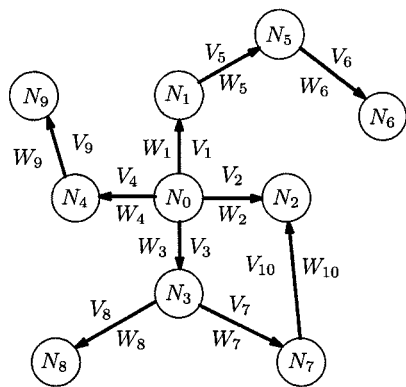


図 1: 重み付き意味ネットワーク

ここで、 N_i は node、 V_i は arc、 W_i は V_i に対する重みのことである。ここでいう重みとは、概念同士の関係の強さ

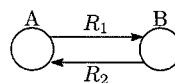
を数値化したものである。その値は、複素数

$$w = a + bi \quad (0 \leq a, 0 \leq b)$$

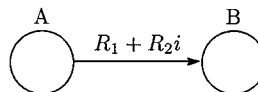
で表現される。複素数にすることによりまた、複素数の大きさ $|w|$ は

$$|w| = \sqrt{a^2 + b^2}$$

で表される。この値が大きければ大きいほど、その値が付加された arc が表す関係が強い事を意味する。また、意味ネットワークは人間の連想能力をモデル化したものであるため、関係が強いということは連想する確率が高いことを意味している。このように意味ネットワークに複素数の重みを付加することにより、ただ関係があるということだけでなく、どの程度の関係にあるかということも表現できるだけでなく、二つの重みを同一の重みとして表現できるため、例えば



と言うような関係を複素重みを利用して、



と表現することによってノード間の相互関係を一つの重みとして表現することが可能になる。また

$$w^2 = (R_1^2 - R_2^2) + 2R_1R_2i$$

を観察することで、 R_1 と R_2 の関係を考察することが出来る。例えば、 w^2 の実部については、 R_1 が R_2 よりもどれ程重要であるかを示していると言える。また、重み付き意味ネットワークによる推論時においてこれらの重みを会して推論を行う事が出来る。この重みは関係の強さを表しているが、意味ネットワークは連想能力をモデル化したものであるため、重みも連想のしやすさと定義することができる。次に簡単な複素重み付き意味ネットワークの生成例を示す。

2.2 生成例

図 2 のようなネットワークがあるとする。図 2 において、node は楕円で表現されており、arc は矢印で表現されている。基本的に意味ネットワークを使って推論を行

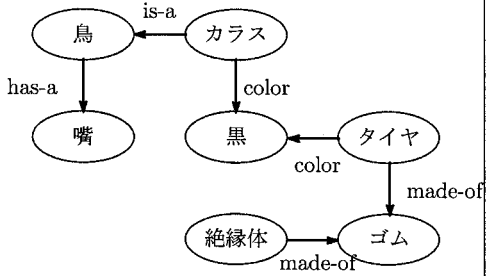


図 2: 意味ネットワーク例

なう場合,node から node へ arc をたどることによって推論を行なう. 次に表現された arc の逆の遷移を考えるために arc の意味合いを考える. 所有とラベル付けされた arc が示すものは遷移元が遷移先を所有しているということである. 逆の遷移で表わされる意味は「所有される」という受動的な遷移と考えられる. 普通, 鳥は嘴を持っているので, 重みは 1.0 と考えることができる. 基本的に重みは, 関係の強さを表すため, 関係があれば 1, 関係がなければ 0 となる. このため, 0~1 の実数値を取るといえる. 同様にして全ての arc に対し重みを与える」と以下の図のようになる.

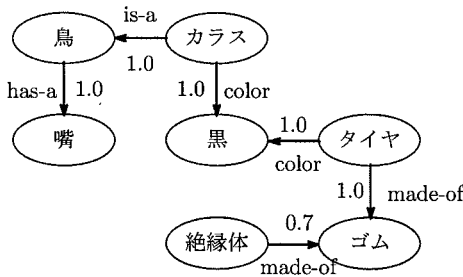


図 3: 重み付き意味ネットワーク例

逆に推論すると嘴は鳥に所有されている. これも 1 対 1 対応しているため, 逆遷移の重みも 1.0 と考えられる. しかしながら, このように遷移先から遷移元への遷移と逆の遷移が同じような関係にあるものもあるが, そうでない関係も存在する. 例えば, 黒という概念を中心に考えてみよう. カラス, タイヤから黒という色を連想することはたやすい. つまり両方の概念から黒に対する関係が強いことを意味している. よって, 両方の arc の値は 1.0 となることが考えられる. しかし, 逆に黒という概念から連想するとすると「カラス」「タイヤ」とどちらを

連想するかは人によってまちまちであろう. このように arc の指し示す方向への遷移と逆の遷移において関係の強さが異なる場合が存在する. このように逆の遷移を考えた時, arc の指し示す関係の強さのみで逆の関係を想定できない場合は逆の遷移を記述する必要がある. 以下に双方向の遷移を持つネットワークモデルを示す

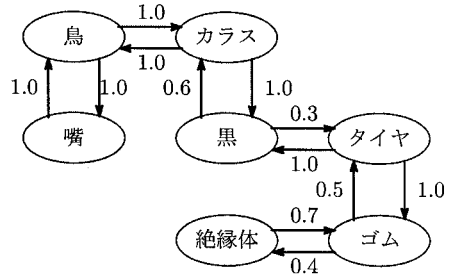


図 4: 重み付き意味ネットワークの相互結合

しかし, このようなモデルで表現すると, arc の数が倍増してしまう. このため我々は重みに与える値として複素数を使用した. 複素数を重みとすることで逆の遷移に対しても一つの重みとして表現することが出来る. 上記の図を複素重みを利用したネットワークモデルに変換すると,

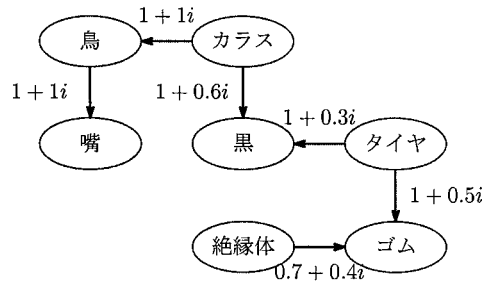


図 5: 複素重み付き意味ネットワーク例

となる. このようにして複素重み付き意味ネットワークを構築する.

2.3 複素重みの獲得

複素重みの獲得はインターネットから収集したテキストより, 以下の処理を繰り返すことで逐次的に獲得する.

- 1 テキストを取得

- 2 形態素解析により動詞と名詞を取り出す
- 3 取り出された名詞と動詞から意味ネットワークを生成する
- 4 tf・idf法により単語に重要度を割り振る
- 5 意味ネットワークを利用しその文で得た全ての arc の重みを獲得する
- 6 以前までの重みを今得られた重みを使って更新する

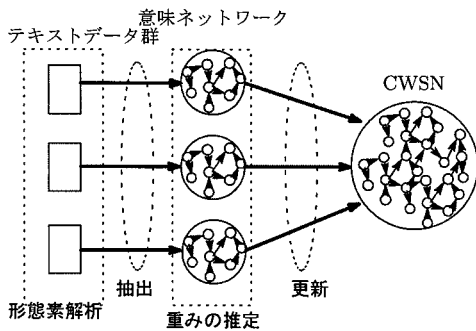


図 6: 重みの獲得

まず、インターネットよりテキストを取得する。取得するテキストの種類については、考察において言及するが、本論文ではジャンルわけをした三種類のテキスト群に分類し、それぞれの種類を取得する。次に、取得されたテキストから形態素解析を行い名詞と動詞を取り出す。形態素解析には Tree-Tagger を使用する。三つ目の工程では抽出された動詞と名詞から意味ネットワークを生成する。処理 4 は、単語の重要度を決定する工程である。重要度は tf・idf 法によって決定する。このときの重要度とは、その単語が文章の中でどれ程重要なかを示した値であり、名詞のみに対して与えるものである。処理 5 は獲得された意味ネットワークを利用し重みを推測することである。ここで、重要になるのが、動詞の結合が受動態結合であるのか、能動態結合であるのかということである。本論で扱う複素重みは実部が能動態重み、虚部が受動態重みとなっているため、基本的には能動態結合で実部の重みを、受動態結合で虚部重みを更新する。また、虚部の重みも実部の重みも 0~1 の実数値を取るため、遷移は値の大きいほうから小さいほうへ遷移する。よって、遷移元の重要度を I_c 、遷移先の重要度を I_p とすると、($I_c > I_p$) のとき、能動態結合は受動態結合と見なし

す。その逆に受動態結合である時は能動態結合と見なし重みの獲得を行なう。まとめると各文における各々の動詞の重みは以下のように定義できる。

●能動態結合時

$$W_i = \frac{I_c}{I_p} + 0i \quad (I_c > I_p)$$

$$W_i = 0 + \frac{I_p}{I_c}i \quad (I_c < I_p)$$

●受動態結合時

$$W_i = 0 + \frac{I_p}{I_c}i \quad (I_c > I_p)$$

$$W_i = \frac{I_c}{I_p} + 0i \quad (I_c < I_p)$$

このようにして得られた重みを同じ単語同士で平均を取り、そのテキストにおける重みとする。このようにして、テキストより生成された意味ネットワークから重みを推測し、複素重み付き意味ネットワークの重みを更新させていく。更新式は以下のようになる。

$$W = \tau * w_{old} + (1 - \tau) * w_{new} \quad (0 < \tau < 1)$$

ここで、 w_{old} は更新前の重み、 w_{new} は新しく入力されたテキストから得られた重みを表す。また、 τ は減衰率であり、 τ の値が大きければ大きいほど、新しい知識ではなく古い知識を重要視すると言うことを指し示し、 τ が小さければ小さいほど新しい知識を重要視するということになる。今回は名詞の重要度の算出として tf・idf 法を用いたが、他の重要度算出法でも対応は可能である。

2.4 複素重みの検証

獲得した重みを検証するために、3 つのテキスト群を用意した。括弧の中の数字はテキスト数を意味する。

- A 群: ジャンルを指定したテキスト群 (10)
- B 群: A 群以外のジャンルのテキスト群 (20)
- C 群: A 群と B 群の両ジャンルを混合したテキスト群 (30)

これらのテキストデータから、複素重みを獲得しそれらを検証する。

2.5 学習結果

ここではA,B,Cのそれぞれのテキスト群から得られた重みの大きさの順に上位15単語を示す。

表 1: A 群における複素重み

単語	実部	虚部	大きさ
quote	0.50000	1.00000	1.11803
report	0.50000	0.92857	1.05463
epitomize	0.33333	1.00000	1.05409
display	0.32510	1.00000	1.05152
differ	0.52912	0.90349	1.04702
edit	0.25000	1.00000	1.03078
spring	0.33333	0.97193	1.02750
depend	0.98237	0.20351	1.00322
meet	0.00000	1.00000	1.00000
call	0.45270	0.87500	0.98517
drive	0.00000	0.90000	0.90000
advance	0.31653	0.83588	0.89380
attend	0.00000	0.87500	0.87500
carve	0.63909	0.59062	0.87021
programme	0.42424	0.75000	0.86167

またA,B,C群のそれぞれから得られた重みの大きさと複素数平面的の図は以下ようになる。

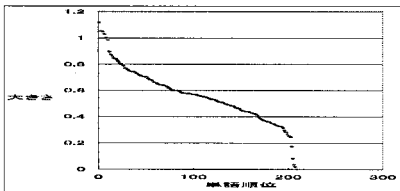


図 7: 重みの大きさ (A 群)

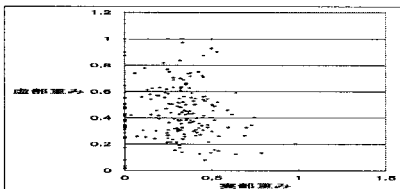


図 8: 重みの複素平面 (A 群)

図 7, 図 9, 図 11 は, テキスト数, ジャンルの異なったテキスト群から得られたにもかかわらず同じような分布をしていることがわかる。により, 一般的な重みが得ら

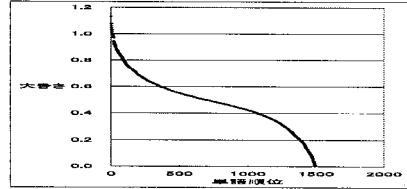


図 9: 重みの大きさ (B 群)

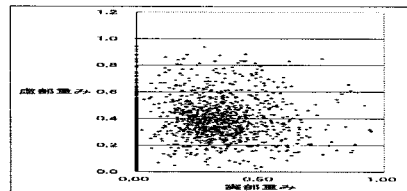


図 10: 重みの複素平面 (B 群)

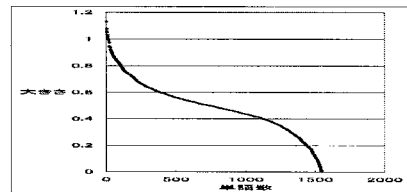


図 11: 重みの大きさ (C 群)

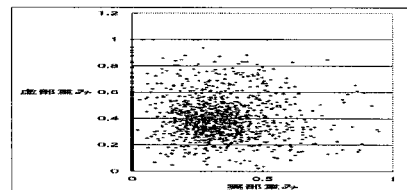


図 12: 重みの複素平面 (C 群)

れたとすることができる。複素平面上でも、図 10、図 12、
 においては同じような分布をしている。このことから
 一般的な知識の獲得が出来たと思われる。

3 複素重み付き意味ネットワークの テキスト要約への応用

本論文では、提案するモデルの応用例として英文のテ
 キスト要約に応用する。この章ではまず、英文テキスト
 要約と複素重み付き意味ネットワークの関係について
 述べる。

テキストデータ内において同一の文に出現する単語
 同士は意味関係が深いものと思われる。このため、複素
 重み付き意味ネットワークの node には名詞,arc に動詞
 を割り当て、文中に共記されている名詞同士をその文中
 の動詞によって連結し、ひとつの枝とする。この枝を連
 結することでネットワークを生成する。この意味ネット
 ワークに 2 節において獲得した重みを付加することで
 複素重み付き意味ネットワークとする。

このようにして生成された複素重み付き意味ネット
 ワークを利用しテキスト要約を行なう。

3.1 要約システム

テキスト要約工程について説明する。テキスト要約工
 程は以下の 6 つの処理から構成される。

- 1 テキストを取得
- 2 形態素解析により動詞と名詞を取り出す
- 3 取り出された名詞と動詞から意味ネットワークを
 生成する
- 4 意味ネットワークの arc に対し重みを付加し、テキ
 スト自体の重み付き意味ネットワークを作成する
- 5 テキストから最重要単語を取り出し、重み付き意
 味ネットワークをたどり、各単語の重要度を算出
 する
- 6 算出された重要度から文の貢献度を算出し、貢献
 度の高い文を抽出して要約文とする。

図 13 に上記の要約手法の流れ図を表示する。テキス
 トを形態素解析し意味ネットワークを抽出するところ(処
 理 1~3)までは、重みの獲得手法の時と同じ手順で行わ
 れる。ここで意味ネットワークの生成に着目する。同一
 の文に共起されている名詞同士を一つの枝としてネット
 ワークを生成するため、生成される意味ネットワーク

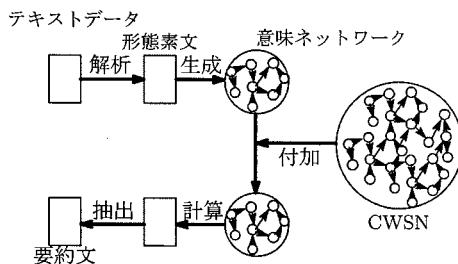


図 13: 要約手法の流れ図

の数は一つのテキストに対して、ひとつ以上生成される。
 これにより、テキスト中に関連性のまったくない文章が
 存在する場合(例えば、複数のテキストを無作為に一つ
 のテキストにまとめるなど)においても、関係のあるも
 の同士ネットワークで連結される事になる。つまり、最
 重要単語が決定できるのであれば、複数存在するテキ
 ストの中から自動的に関連のあるテキストのみを選択し、
 要約することも可能であるといえる。

重みの獲得法と異なるのは、4 つ目の工程からである。
 要約の行程では今まで得た知識を使用するため、抽出さ
 れた意味ネットワークの arc に対し、これまでに更新(学
 習)してきた重みを割り当て要約するテキストに対する
 重み付き意味ネットワークを作成する。また同時にテキ
 ストにおける最重要単語を決定する。ここで最重要単語
 とはそのテキストにおける最も重要な単語であり、 $tf \cdot$
 idf 値が最大である単語とする。最重要単語とその $tf \cdot$
 idf 値を基準に重み付き意味ネットワークをたどって各
 単語の重要度を決定する。 $tf \cdot idf$ 法で決定する重要度は
 最重要単語の重要度のみである。

3.1.1 重要度の決定

名詞の重要度は生成された意味ネットワークを最重
 要単語の重要度 I_{max} を初期値としてたどり、その arc
 の遷移先 node の重要度 I_p は遷移元 node の重要度 I_c
 と重み w_{ij} を使用し決定する。まず、最重要単語と結合
 している node の重要度 I_i は

$$I_i = \begin{cases} I_{max} * (\frac{w+w}{2}) & \text{能動態結合} \\ I_{max} * (\frac{w-w}{2i}) & \text{受動態結合} \end{cases}$$

となる。ここで w は最重要単語と node を結合している
 arc に割り当てられている重みである。更に、重要度の決
 定した node から結合されている node の重要度を決定
 し、これを全 node の重要度を決定するまで繰り返す。つ
 まり最重要単語より意味ネットワーク内で接続されて
 いる順に、順次個々の重要度が決定していく。このとき、

一度重要度を決定した node が出現した場合、再計算し、大きいほうをその node の重要度とする。一般的な I_a と I_b の関係式は以下ようになる、

$$I_b = \begin{cases} I_a * (\frac{w+\overline{w}}{2}) & \text{能動態結合} \\ I_a * (\frac{w-\overline{w}}{-2\overline{w}}) & \text{受動態結合} \end{cases}$$

このようにして算出された重要度から、文の貢献度を算出する。また、未知動詞つまり重みの獲得・更新時において獲得されなかった動詞の重みはすべて $0+0i$ とする。

3.1.2 文の貢献度の決定

貢献度とは単語の重要度を文単位に拡張した値であり、その文がどれだけ要約に貢献しているかという度合いを数値化したものである。そのため、その文中に存在する単語の重要度から算出される。文 i の貢献度 C_i は以下の式で定義される

$$C_i = \frac{\sum I_{ij}}{n}$$

ここで、 I_{ij} は i 番目の文の j 番目の名詞の重要度であり、 n はその文に存在する名詞の総数である。文中に存在する名詞の重要度の合計を名詞の総数で割ることにより、正規化を行なっている。このようにして得られた文の貢献度を利用し要約文を抽出する。

3.1.3 要約文の抽出

貢献度の高い文から抽出し、要約文とする。抽出する文の数は要約率に依存し、要約率を超えるまで貢献度の高い文を順に抽出しつづける。要約率は以下の式で与えられる

$$\text{要約率} = \frac{\text{抽出された総単語数}}{\text{テキスト内の総単語数}}$$

抽出された文を原文の出現順に並び替え要約文とする。

3.2 要約結果

tf・idf 法のみを使った要約手法との一致率から本手法の要約文を評価する。一致率とは、どれだけ同じ文が要約されているのかという比率であり、

$$\text{一致率} = \frac{\text{二つの要約文で共に出現した文の数}}{\text{二つの要約文の文の平均}}$$

で与えられる。一致率が高いと二つの要約手法は同様の要約をしているといえる。要約率 10%~50%における A 群、B 群、C 群のそれぞれの重みを使用した要約文と tf・idf 法を使用した要約文との一致率は以下ようになる。

表 2: tf・idf 法との一致率

	10%	20%	30%	40%	50%
A 群	33.3	36.3	50.0	52.4	65.2
B 群	16.7	36.3	43.3	42.9	65.0
C 群	33.3	36.3	50.0	52.4	65.2

ここで、原文は単語数 1843 文字、文数 52 文であるので、要約率 10% では約 5 文、180 文字程度である。

要約率 10% では、どの重みでも低い一致率であるが、要約率が上がるにつれて、一致率も向上している。ここで重要なのは、本当に重要な文を抽出でき、不要な文を抽出していないことである。本研究では要約に対して重要であると思われる 3 つの文について、どの要約率で抽出されているかという観点から本手法を評価する。

表 3: 重要文の出現要約率

	tf・idf	A 群	B 群	C 群
1 文目	20%	10%	30%	10%
2 文目	30%	20%	10%	30%
3 文目	20%	10%	40%	10%

要約するテキストに関連する知識(専門知識)を獲得した場合(A 群)の要約文では、重要文が早い要約率で出現している。これは tf・idf 法のみでの要約よりも、重要文の貢献度が上位に来ていることを示している。tf・idf 値を意味ネットワークの知識によって算出しなおすことでより良い要約ができていると言える。また、C 群と tf・idf よりもより良い結果となっていることから、少しでも関係のある文章を知識として取り込んでいるならば、tf・idf 法のみでの要約よりも有効な要約ができているといえる。このことにより本要約手法において、複素重み付き意味ネットワークが要約に対する知識として有効であることが解った。早い意思決定が求められる現代社会において、要約時間が長ければいくら良い要約をしたところで意味のない事である。そのため要約時間について考察する。本手法では tf・idf 値を求める上に意味ネットワークを形成しなければならないため、余計に時間がかかってしまう。また、意味ネットワークは文字数が多ければ多いほど枝数が増え、巨大なものになってしまう。本手法と tf・idf 法のみでの手法のそれぞれの要約時間と文字数、文数、意味ネットワークの枝数の関係について比較すると以下ようになる。

5000 文字程度のテキストであれば、55 秒と 1 分を超えない程度で要約ができているので、時間的には tf・idf 法と比べてもさほどかかっているとはいえない。しかし、20000 文字を越えたあたりから、tf・idf 法のみでの手法

表 4: 要約時間

文字数	文数	枝数	本手法	tfidf 法
72749	5167	71728	1632	1161
72211	5021	94774	1414	837
25088	372	289418	1091	110
5733	169	21861	55	31
3539	144	2870	5	5
2661	177	2134	5	4
1032	40	1360	3	2
707	57	183	2	2
323	25	215	2	2

に比べ、10分以上余分にかかってしまう。これは、使用品詞を動詞と名詞のみに限定しているとはいえ、文中に存在するすべての名詞と動詞に対して枝を形成するため、単語数が増大すると枝の数も増大してしまい、名詞の重要度決定の時に参照する対象が増加してしまう為である。しかしながら、これに対しては、段落ごとに処理を行うなど、逐次的に重要度を決定することにより対処できると思われる。

4 終わりに

本論文では、より効率的に処理するために、意味ネットワークの arc に対して複素数の重みを付けた知識表現の意味ネットワークモデルを提案した。また、応用例としてテキスト要約に適用し現行の手法と比べることにより評価した。この結果、重みを複素数にすることによって、受動態、能動態の区別をつけとることができより高度な要約が行なえるようになっていえる。また、複数のテキストデータより知識を獲得し、その知識を使った要約ができ、現行の手法と比較すると遜色ない要約が行なわれていることが判明した。更には学習するテキストをジャンルによって限定することで、そのジャンルに特化した重みを得られている。また、これらのことよりテキストデータより得た知識を複素数を用いた複素重み付き意味ネットワークとして表現することができたと言え。さらに応用事例として、得た知識を利用したテキスト要約を行い要約結果より現行の手法よりよりよい要約が出来たとことから、複素重み付き意味ネットワークの実用性、有効性が実証できたといえる。

今回は応用事例としてテキスト要約を用いたが、その発展としては要約で得た意味ネットワークを利用した文章生成に適用することが期待される。日本語文テキストにおける文章生成は、京都大学大学院の尾崎らによって試みられている。そこでは体言と用言に言語を分類し、そのつながりによって文章の生成を行なっている。この

ようなことを実現するためには意味ネットワークで使われる要素について再考が必要である。本論文では単純化を図るため、動詞と名詞のみを意味ネットワークの要素とした。しかし、文章を生成するためには形容詞や副詞、その他の品詞も要素として使用する必要がある。また、単語に対する意味関係などの背景知識を留意する必要がある。

参考文献

- [1] Fahlman.S.E. : "Design sketch for a million - element NETL machine", Proc . of First Annual National Conf. on AI(Aug 1980)
- [2] 古谷立美, 樋口哲野, 半田剣一, 国分明男:意味記憶システム IX(イクス)-意味ネットワークのハードウェア化について-, 情報処理学会, 研究報告,1984-ICS-039 pp41-48,1985
- [3] Hillis.G.E. : "The connection machine", TR-646, Cambridge,MIT AI Lab
- [4] 長尾正太郎, 黒橋禎夫, 長尾真 :意味ネットワークからの文章生成, 情報処理学会, 自然言語処理研究会,120-20, pp.133-140,1997
- [5] 使いやすくなった自然言語処理のフリーソフト, 情報処理学会誌, Vol.41, No.11, pp.1202-1238(2000).
- [6] 小嶋秀樹 : 単語の意味的な類似度の計算, 電子情報通信学会技術研究報告, AI92-100, pp.81-88, 1993
- [7] 北 研二, 津田和彦, 獅子堀正幹 : 情報検索アルゴリズム, 共立出版, (2002)
- [8] Tree-Tagger,<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>
- [9] 高木 徹, 木谷 勉:単語出現共起関係を用いた文書重要度付与の検討, 情報処理学会研究報告 96-FI-41-8, (1996)
- [10] 荻原, コネクションリストモデルを用いた知識ベースシステム, 研究報告,1989-ICS-064,pp.1-10,1989
- [11] W.Li,M.Wu,Q.Lu,W.Xu and C.Yuan, "Extractive Summarization using Inter- and Intra- Event Pelevance," Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, pp.369-376, 2006.