概要

近年，インターネット上で様々な電子テキストが増加し，これらの電子テキストから有益な情報を取り出す技術が望まれている．

大竹ら [1] は，TF-IDF を用いて，新聞記事群から事物の関係情報を単語ネットワークとしてまとめたものを構築した．土遠ら [2] は，単語ネットワークを構築する際に，事物と無関係であるノードの削除を行った．窪ら [3] は，ノード同士の関係を示す情報としてリンクに文字列の付与を行った．

しかし，これらは複数文書を入力と場合における，発想支援を目的とした研究であり，一連の内容として書かれている単一文書に対しては適用できないということがあった．

そこで本研究では，単一文書を入力とした場合の単語ネットワークの構築を行う．単一文書を単語ネットワークとして可視化することで，文書を読む手間を省くことができる．本研究の目的は，単一文書を入力として単語ネットワークを構築することにより，読書支援に役立てることである．

実際に単一文書として「新聞記事」の単語ネットワークを構築し，単語ネットワークを利用した際の読書量に対する内容理解量を調査した．調査の結果,「新聞記事」を単語ネットワークとして出力し利用することで，入力データ全体の約 44%の読書量で，約 44%の内容を把握できることを確認した．ランダムで段落を 3 つ抽出した場合は，読書率約 42%に対して正解率 32%となり，正解率の向上を確認できた．

「小説」を入力として単語ネットワークを構築し，本文中における出現箇所を調査した．調査の結果，正規分布において約 95%の範囲に単語ネットワーク構築に用いた単語が 98%の確率で出現することを確認した．これにより，登場人物やある事柄など，単語ネットワークでノードとして出力されている単語の出現段落の推定が可能となった．

また，ノード対の有用性について調査した．調査の結果，入力データとして用いた 6 つの小説の全てで，5 つ以上の有益なノード対が獲得できることを確認した．これにより，登場人物の特徴，2 人の登場人物の関係性，物語における有益な情報のいずれかを獲得することができる．特に登場人物についての情報が多く獲得できたため，物語の大枠を捉えられる可能性から，読書支援にも有効であると考えられる．

目 次

表 目 次

図 目 次

第1章　はじめに

近年，インターネット上で様々な電子テキストが増加し，これらの電子テキストから有益な情報を取り出す技術が望まれている．大竹ら [1] は，電子テキストから特定のキーワードに基づく関係情報をネットワークとして抽出する方法を提案し,「地震」というキーワードに基づいて単語ネットワークの構築を行った．土遠ら [2] は，大竹らが構築したネットワークに関連のない事物のノードを含むことを確認し，それらのノードを削除を行った．窪ら [3] は，ノード同士の関係を示す情報としてリンクに文字列の付与を行った．しかし，これらは複数文書を入力とした場合における，発想支援を目的とした研究であり，一連した内容について書かれている単一文書に対しては適用できないということがあった．

そこで本研究では，単一文書を入力とした単語ネットワークを構築する手法を提案する．単一文書を単語ネットワークとして可視化することで，文書を読む手間を省くことができる．本研究の目的は，単一文書を入力として単語ネットワークを構築することにより，読書支援に役立てることである．

本研究の主張点を以下に示す．

- 単語ネットワークの入力として単一文書を用いることが想定されていないという問題を解決するために，システム改良により単一文書を入力として単語ネットワークの構築をする．

- 単一文書の「新聞記事」を入力として単語ネットワークを構築し，単語ネットワークの第 2 階層に出現した単語を含む段落を 3 つ抽出した場合，入力データ全体の約 44%の読書量で，約 44%の内容を把握できることを確認した．ランダムで段落を 3 つ抽出した場合は，読書率約 42%に対して正解率 32%となり，正解率の向上を確認できた．

- 「小説」を入力として単語ネットワークを構築し，本文中における出現箇所を調査し，正規分布において約 95%の範囲に単語ネットワーク構築に用いた単語が約 98%の確率で出現することを確認した．これにより，登場人物やある事柄など，単語ネットワークでノードとして出力されている単語の出現段落の推定が可能となった．

- 「小説」を入力として単語ネットワークを構築し，ノード対の有用性について調査し，入力データとして用いた 6 つの小説の全てで，5 つ以上の有益なノード対が獲得できることを確認した．これにより，登場人物の特徴，2 人の登場人物の関係性，物語における有益な情報のいずれかを獲得することができる．特に登場人物についての情報が多く獲得できたため，物語の大枠を捉えられる可能性から，読書支援にも有効であると考えられる．

本論文の構成は以下の通りである．第 2 章では，本研究の関連研究を述べ，第 3 章では，本研究の基本となるネットワーク構築の流れ及び文字列の付与手法について述べる．第 4 章では，単一文書を入力とした場合の単語ネットワーク構築手法を提案する．第 5 章では，実験条件，実験結果，評価方法，評価結果を述べる．第 6 章では，結果の考察と今後の課題を述べる．第 7 章では，本論文のまとめを述べる．

第2章　関連研究

2.1　ネットワークの関連研究

本研究では，単一文書の可視化手法として単語ネットワークを構築する．ネットワークの関連研究を以下に示す．

内山ら [4] は，大規模な出来事の要約，すなわち，複数のトピックに関する複数の文書の要約を目的としている．複数文書においてネットワークを構成し，ネットワークの各ノードの重要度を活性拡散を利用し求めている．それにより，複数文書の要約を行っている．

松尾ら [5] は，Web 上の情報から，人間関係のネットワークを抽出している．抽出手法として，氏名の関係性の強さを知るための様々な指標を用いている．

松尾ら [6] は，ノードが離れているにも関わらず，別のノードを介せば近いという Small World 構造を用いてネットワークを構築し，そのネットワークからキーワードを抽出する手法を提案した．

2.2　要約の関連研究

本研究は，単一文書の読書支援を目的としている．読書支援において，文書の要約は広く知られているため，本研究に関連する．要約の関連研究を以下に示す．

瀧川ら [7] は，入力文から名詞を抽出し，抽出した各名詞から名詞の共起語を取得している．取得した共起語を連想知識として用いることで，端的な要約を生成する手法を提案した．例として,「良い企業に内定をもらうために面接の練習を毎日行う」という入力文からは,「就職活動」という端的な要約を得ることができている．

西川ら [8] は，複数の文書から要約を作成する複数文書要約を，冗長性制約付きナップサック問題として捉えた．この問題に対し，ナップサック問題に基づく要約モデルに，冗長性を削減するための制約を加えることで，複数文書要約モデルを得ている．

森ら [9] は，複数の質問の答とその背景知識を一度に概観できる要約を生成する手法を

提案している．複数の質問文を入力し，質問応答エンジンと語の出現分布を用いて，文の重要度の計算を行った．その結果，複数の質問文の答を含む要約文書を抽出している．

2.3 問題回答システムの関連研究

本研究は，文書読解に関連するとして，問題回答システムの関連研究を以下に示す．

松井ら [10] は質問文とテキスト集合の中から，回答となる語句を探す手法としてキーワードベクトルの類似度を用いる手法を提案している．その結果，大学入試における穴埋め問題で 3 割から 5 割の正答率を得た．

2.4 可視化の関連研究

本研究は，小説を入力データとしての可視化を行うため，小説の可視化についての関連研究を以下に示す．

縣ら [11] は物語テキストの内容理解を支援するために，物語テキストの進行状況に応じた人物相関図の生成を係り受け解析とあらかじめ生成した死亡表現リストを用いて行った．これにより，登場人物の存在状態と人物感の有効敵対関係を推定することが可能となり，提案手法における死亡判定について再現率と適合率を用いて評価した結果，それぞれ平均して，84.7%と 79.4%の正解率を得た．

西原ら [12] は読書の再開前に，物語の登場人物に関する情報を整理することで円滑に読書を再開できると考え，人手構築した抽出パターンを用いて，物語テキストから登場人物の関係を自動的に抽出する手法を提案した．また，テキストから物語の登場人物を抽出すると同時に，親子，京大などの家族関係，友人，会社の同僚などの仲間関係など，様々な登場人物感の関係を推定することも可能とした．これにより，人物関係抽出の F 値 0.340 を確認し，ベースラインの手法を上回った．

第3章　先行手法

本章は，大竹ら [1]，土遠ら [2]，窪 [3] が単語ネットワークを構築した，先行手法について，窪 [3] の論文の記述を参考にして，説明する．

3.1　単語ネットワーク構築

新聞記事群のデータ（本論文では，新聞データと呼ぶ）から単語ネットワークを構築する．単語ネットワークの構築の手法は，大竹ら [1] の手法，テーマ限定抽出法，テーマ無関連削除法 [2] の 3 つの手法があるが，本研究ではテーマ限定抽出法を用いて単語ネットワークの構築を行う．単語ネットワーク構築の流れを図 3.1 に示す．また，本章では，テーマ限定抽出法のみを説明する．

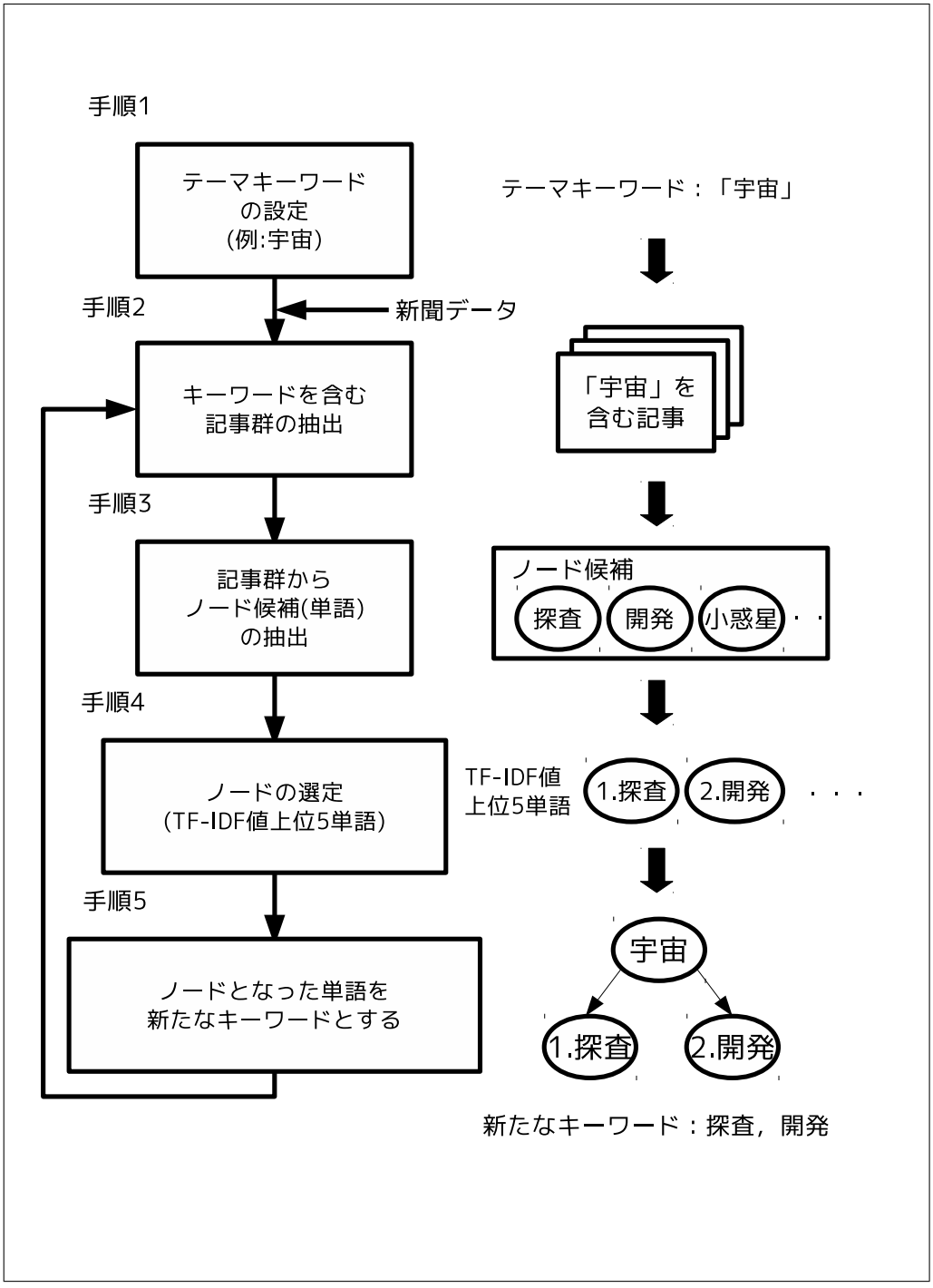


図 3.1: ネットワーク構築の流れ

3.1.1　テーマキーワードの設定

構築したいネットワークの主となる概念を，テーマキーワードとして設定する．例としては,「トヨタ」「宇宙」「ギリシャ」等の単語である．

3.1.2　キーワードを含む記事の抽出

新聞データから，キーワードを含む記事の抽出を行う．記事の抽出方法を図 3.2 に示す．

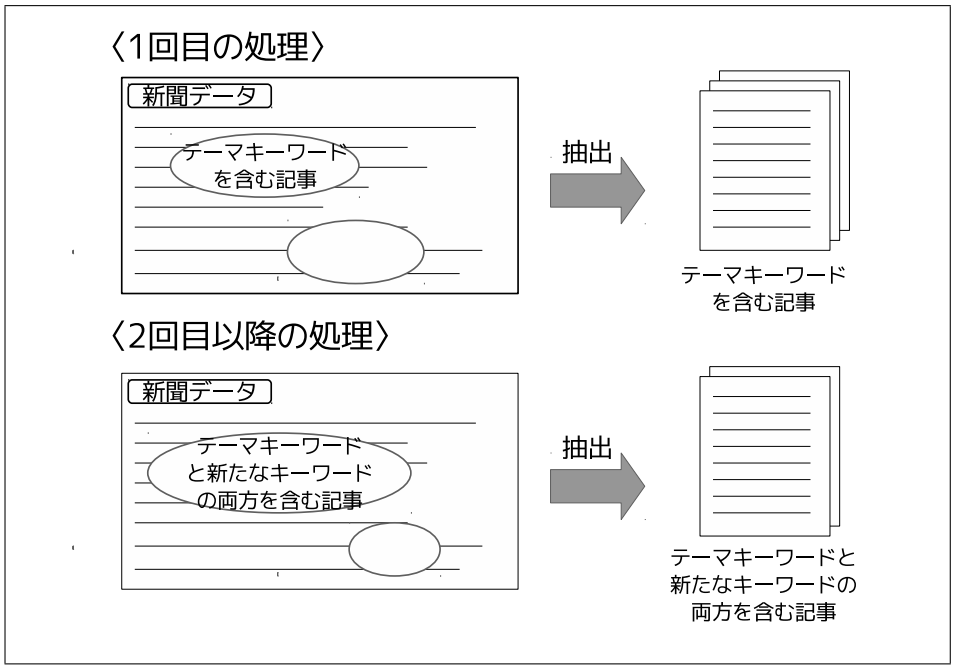


図 3.2: 記事の抽出

1 回目の記事の抽出は，テーマキーワードを含む記事とする．2 回目以降の記事の抽出は，テーマキーワードと，次のノードとなった新たなキーワードの，2 つのキーワードを含む記事とする．

3.1.3　記事の形態素解析

抽出された記事に対して，形態素解析を用いて，名詞を取り出す．

形態素解析とは，テキストを形態素と呼ばれる単位に分割することである．形態素は，厳密には単語とは違った分割の単位だが，おおよそ単語と同じようなものになり，品詞の情報を持つものである．形態素解析結果の例を図 3.3 に示す．

入力：「宇宙飛行士の若田光一さんが国際宇宙ステーションの第３９代船長に就任した」

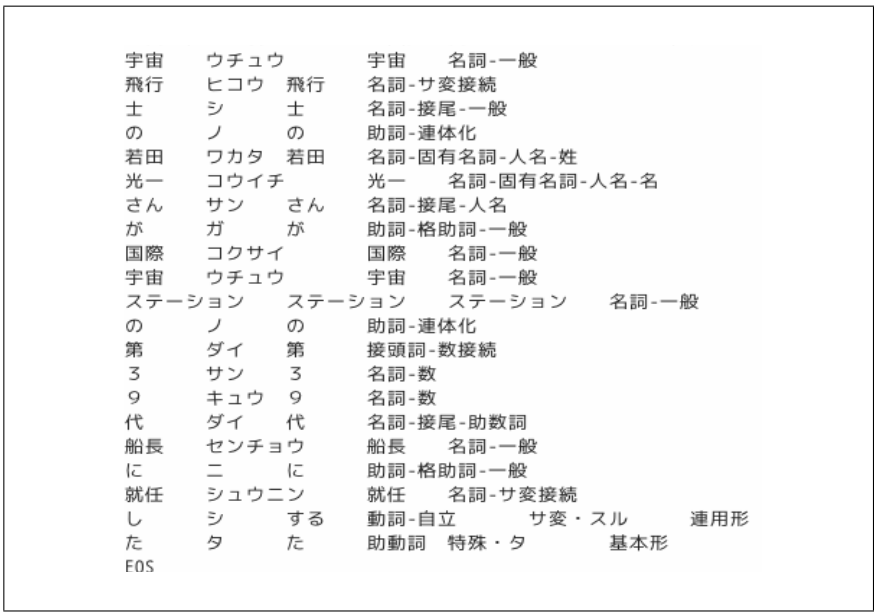


図 3.3: 形態素解析の出力例

図 3.3 のように，形態素解析を行うことで，品詞の情報を持った単語に分割する．本研究では，形態素解析に ChaSen を用いる．また，形態素解析を用いて名詞を取り出す際に，一文字，ひらがなのみ，数字のみの単語を除外する．

3.1.4　ノード候補の抽出

形態素解析を行った後，ノード候補となる単語の抽出を行う．

3.1.3 節の図 3.3 を例とすると,「宇宙」「飛行」「若田」「光一」「国際」「ステーション」「船長」「就任」といった単語がノード候補として抽出される．

3.1.5　ノード候補の選定

TF-IDF を用いて，抽出されたノード候補の中から，実際にノードに用いる単語を選定する．TF-IDF 値の上位 5 単語をキーワードと関係性の強い単語とする．

TF-IDF について説明する．TF-IDF は抽出した記事内におけるノード候補となっている単語の重要度を表す．TF-IDF は以下の式 3.1 で算出される．

$$TF\text{-}IDF = tf_t * \log \frac{N}{df_t} \tag{3.1}$$

tf_t はキーワードを含む記事群での単語 t(ノード候補) の出現回数，df_t は全記事での単語 t の出現記事数とし，N は新聞データの全記事数とする．この式からどの記事にも現れるような重要度の低い単語については低い重みを，他の記事にあまり現れないような貴重な単語には高い重みを与えることになる．

3.1.6　ネットワークの拡大

3.1.5 節で得た TF-IDF 値の上位 5 単語を，キーワードから繋がる，次のノードとする．次のノードを新たなキーワードとして設定し，3.1.2 節の 2 回目以降の処理に戻る．3.1.2 節から本節までの処理を繰り返すことにより，単語ネットワークを構築していく．構築したネットワークの例を図 3.4 に示す．図 3.4 は，テーマキーワードを「宇宙」とし，毎日新聞 2014 年度から構築したネットワークである．

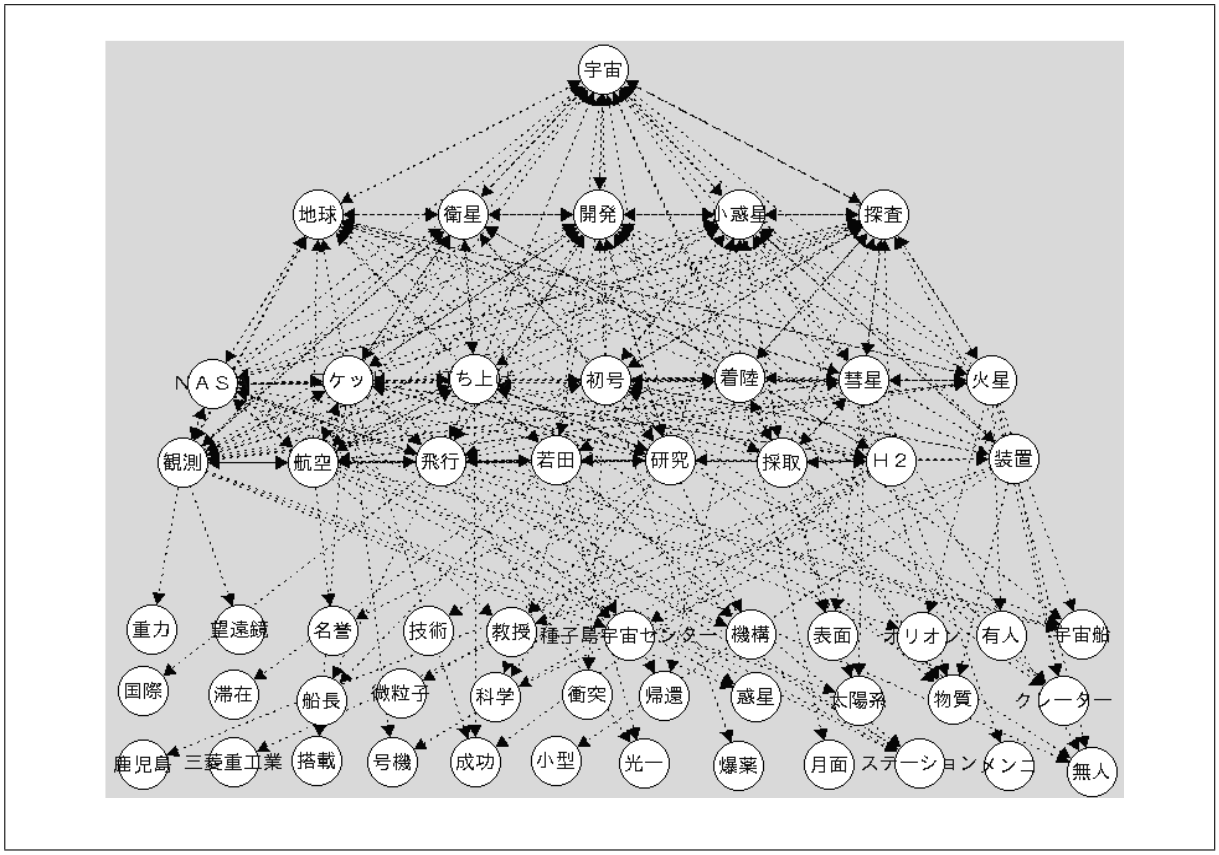


図 3.4: 構築したネットワークの例

3.2　リンクへの文字列の付与

単語ネットワークのノード間の関係性を分かりやすくするため，リンクに単語同士の関係性を示す文字列の付与を行う．

入力を新聞データと，3.1.6 節の図 3.4 の「宇宙」「探査」のような単語対データとし，出力をリンクに付与する文字列とする．付与する文字列の選定の手法を以下に示す．図 3.5 は，単語ネットワークのリンクへの文字列付与の例である．

1. 新聞データから，2 単語の間の文字列 (文字列 A と呼ぶ) を抽出する．
2. 2 単語と文字列 A の連接したものを含み，句点で区切られた文字列 (文字列 B と呼ぶ) を抽出する．文字列 A と文字列 B の抽出例を表 3.1 に示す．

表 3.1: 文字列 A と文字列 B の抽出例

単語対	文字列 A	文字列 B	元の文字列
「ギリシャ」「国債」	の	中国は財政再建に取り組むギリシャの国債を購入し	中国は財政再建に取り組むギリシャの国債を購入し、ユーロ防衛に協力する姿勢を示すなど欧州への影響力を拡大している。
「トヨタ」「水素」	自動車は	トヨタ自動車は水素で動く燃料電池車を２０１４年度に国内で販売と発表	トヨタ自動車は水素で動く燃料電池車を２０１４年度に国内で販売と発表。市販は世界初となる見通し

3. 文字列 B の中で，最も優先度が高い文字列 (出現頻度が高いものや，文字長が短いものを優先度が高い文字列とする．これを文字列 C と呼ぶ) を取得する．これを各文字列 A に対して行う．
4. 3 において取得した文字列 C のうち，優先度が最も高い文字列を選定する．
5. 選定した文字列をリンクに付与する．

優先度の算出には以下の式を用いる．3.2 式は，文字列の出現頻度を文字列の長さで割り，優先度を求める式である．

$$\text{優先度} = \frac{\text{頻度}}{\text{文字長}} \tag{3.2}$$

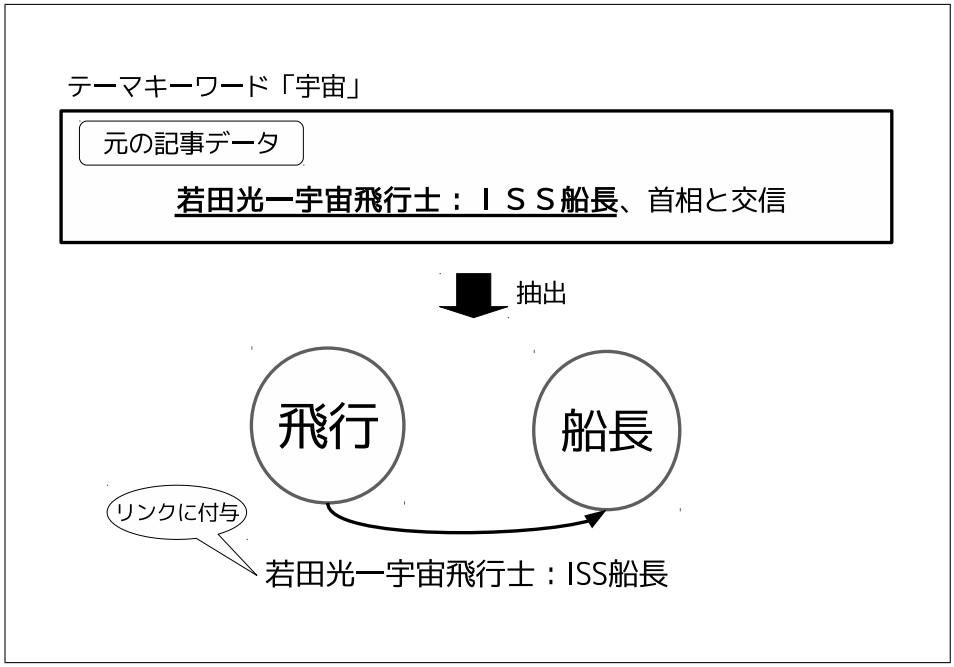


図 3.5: 単語ネットワークのリンクへの文字列付与の例

第4章　提案手法

本章では，本研究の提案手法について説明する．単一文書を入力とした単語ネットワークの構築手法について，3.1 節の手順を基本手法として 3 つの変更点を示す．

4.1　TF-IDF における DF の扱い

単一文書を入力として単語ネットワークを構築するため，重要度を算出する TF-IDF の式に用いる DF の値を新たな手法により算出する．先行手法の 3.1.5 節では，TF-IDF で用いる DF を算出する際に入力データを使用していた．本来，TF-IDF を算出する場合，入力データである文書群における単語の出現頻度 (DF) が低いほど重要語として扱われる．また，DF を算出する際には入力データとは別の文書群を用いなければいけないため，入力データで DF を算出すると，DF の値が大きくなったものが重要となる場合があり，本来の TF-IDF 手法とは矛盾が発生してしまう．入力データ内で重要であるべき単語を抽出できないという問題が発生する．そこで，あらかじめ様々な単語に対して DF の値を算出することで対応する．DF の値は入力データとは別に新聞記事群を用意し，新聞記事群での単語の出現回数から算出する．また，入力データ内で出現するが新聞記事群では出現しない単語があった場合には，DF=1 として算出する．

4.2　段落分けによる入力データの処理

単一文書を入力として単語ネットワークを構築するため，入力データに段落分けの処理を行う．先行手法の 3.1.2 節では，ノードとした単語 (キーワード) に関連する単語を次のノードにするために，次のノード候補となる単語を取り出す方法として，キーワードを含む記事の抽出を行っている．しかし，単一文書を入力とした場合にはキーワードに関する記事を限定して抽出することができない．キーワードを含む文書として入力データ全体が抽出されてしまい，ノード候補となる単語が入力データ内の全単語となってしまう．そこで，入力データを段落で分割し，文書群として扱うことで対応する．これに

より段落で分割されたキーワードを含む文書群からキーワードに関連するノード候補を獲得することが可能となる．

4.3　ノードへの段落情報の付与

単一文書を入力として単語ネットワークを構築した際に，ノードがどの段落に出現するかを提示できれば，入力データ内において，出現箇所の把握速度の向上が見込めると考えた．そこで，手法 4.2 で段落分けをした後，単語ネットワークに選出されるノードに段落番号を付与する．付与手順を以下に示す．

1. 段落で分割された入力データに，それぞれ段落番号を割り当てる．
2. 段落内にノードの単語を含む場合，ノードにその段落番号を付与する．
3. 付与された段落番号を元に，平均値，標準偏差，出現範囲をノードに付与する．
4. 段落番号の付与を行ったのち，単語ネットワークの階層内でノードに付与された平均値を降順で図の左から並べる．

なお，ノードに付与する情報が多すぎることによって，可視性を損なう恐れがあるため，入力データの文章量により，適宜付与する情報を選択する．

文章量が少ない例として「新聞記事」を入力とした場合の単語ネットワークを図 4.1，文章量が多い場合の例として「小説」を入力とした場合の単語ネットワークを図 4.2 に示す．なお,「新聞記事」の単語ネットワークに付与する情報は段落番号，平均値，標準偏差とした場合,「小説」の単語ネットワークに付与する情報は平均値とした場合である．

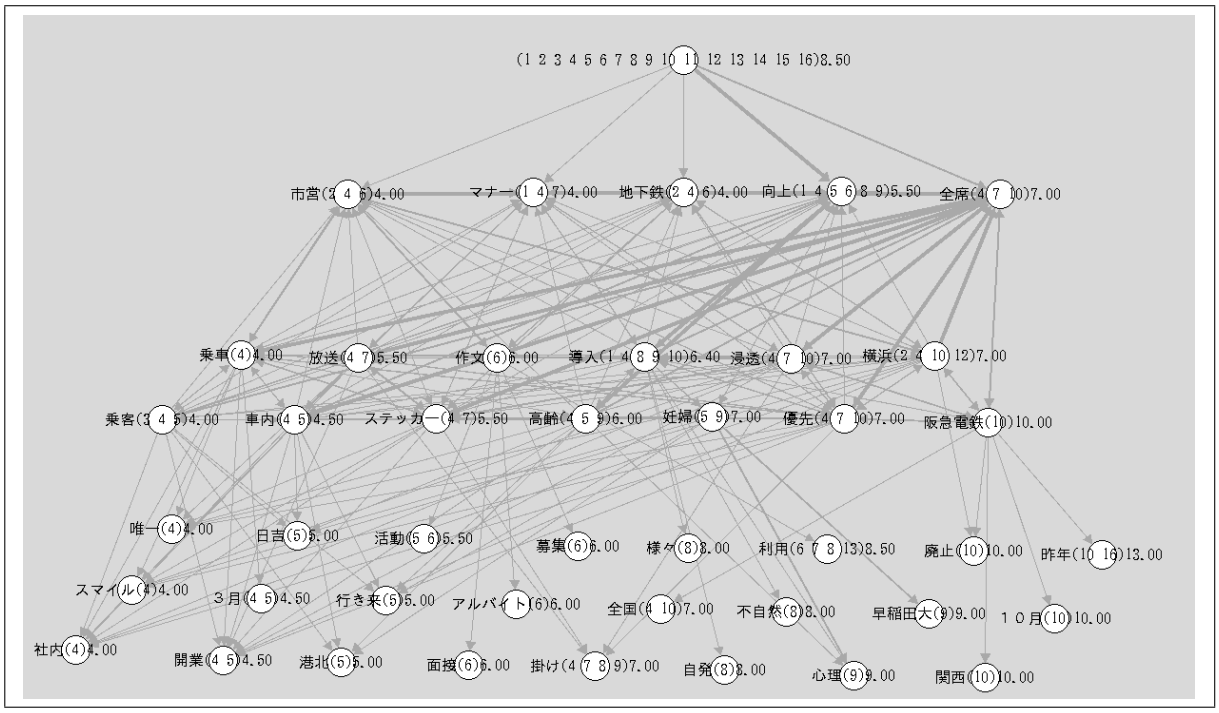


図 4.1: 「マナー」に関する新聞記事のネットワーク

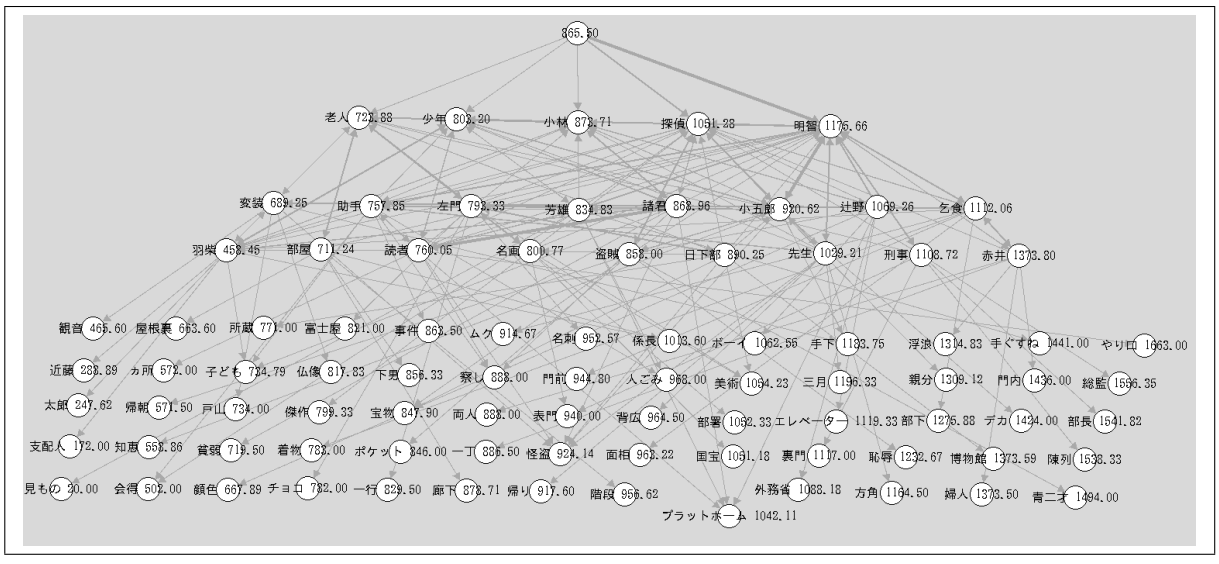


図 4.2: 小説「怪人二十面相」のネットワーク

第5章　実験

5.1　人手による特徴分析

5.1.1　実験条件

本実験では，新聞記事「エコ表示根拠示して (毎日新聞 2008 年度)」及び小説「怪人二十面相」を入力として単語ネットワークを構築する．IDF の算出には毎日新聞 2012 年度 (110,587 記事) を用いる．

5.1.2　新聞記事のネットワークの分析結果

図 5.1 の単語ネットワーク図では「マーク」や「環境」などのように上位層に出現しているノードから記事の内容をおおよそ予想できる．また，それぞれのノードの段落番号に注目してみると出現した段落に偏りがあることがわかる．今回であれば，2,6,7 段落がよく出現しているので，その段落を見ればより詳細な内容を確認できる可能性が高いと考えられる．

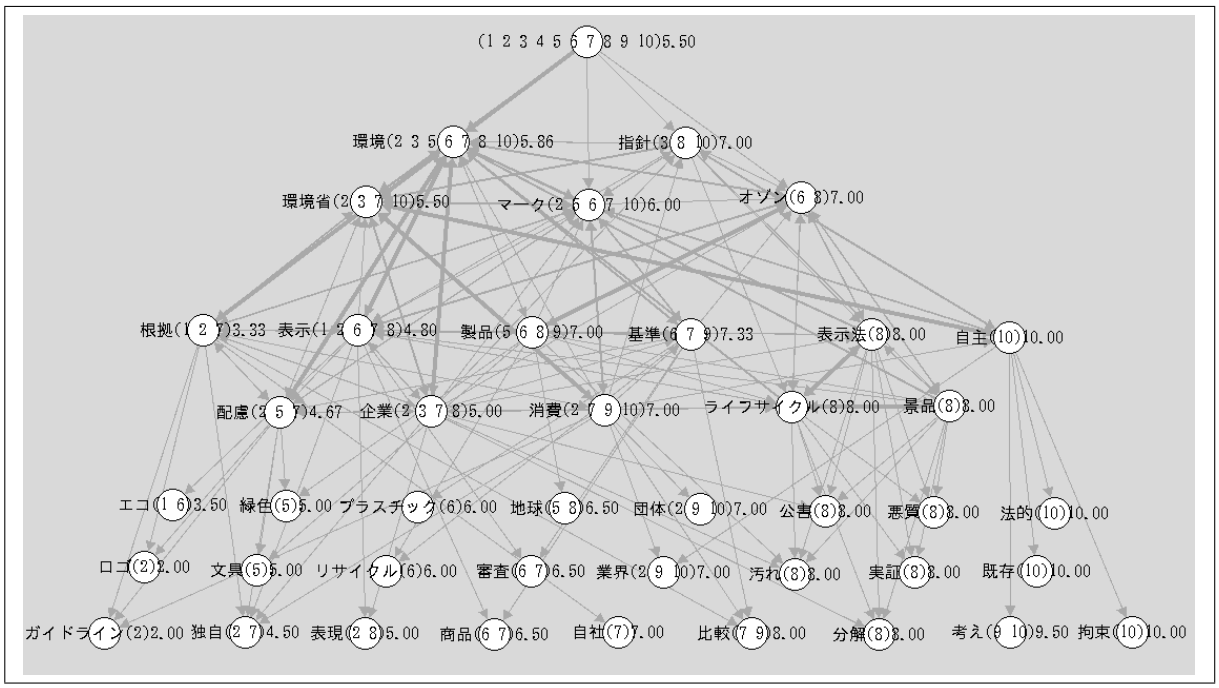


図 5.1: 「エコ表示」に関する新聞記事のネットワーク

5.1.3　小説のネットワークの分析結果

小説を入力とした場合は登場人物に関するノードを上位層にて多く獲得できた．これは文中に登場人物の名前が多く出現し，TF-IDF の値が高くなったことが要因と考えられる．また，トップノードの平均値 865.50 は段落の中央値を示している．例えば「小林」は段落の平均値 873,31 で中央値に近いため，登場人物の中でも特に中心的な人物である可能性が高いと考えられる．

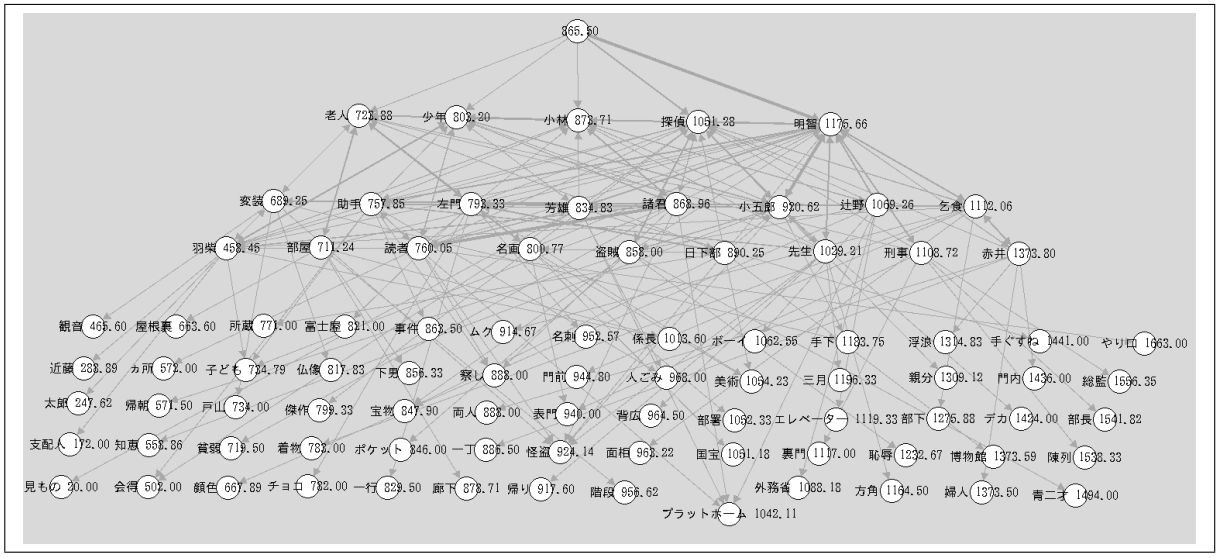


図 5.2: 小説「怪人二十面相」のネットワーク

5.2　内容把握度での評価

単一文書を入力として単語ネットワークを構築し，実際に利用した際の文書内容の把握度で評価を行う．

5.2.1　実験条件

本実験では,「構成・特徴・分野から学ぶ新聞の読解 [13]」(以下，文献 [13]) 内で用いられている，新聞記事 8 件で単語ネットワークを構築する．実験に用いる新聞記事の合計文字数は 6,663 文字である．出力された単語ネットワークの例を図 5.3 に示す．また文献 [13] より新聞記事の内容に関する問題を 1 記事につき 2 問，合計 16 問を用いて実験を行う．問題例と回答例を図 5.4，記事本文の例を図 5.5 に示す．

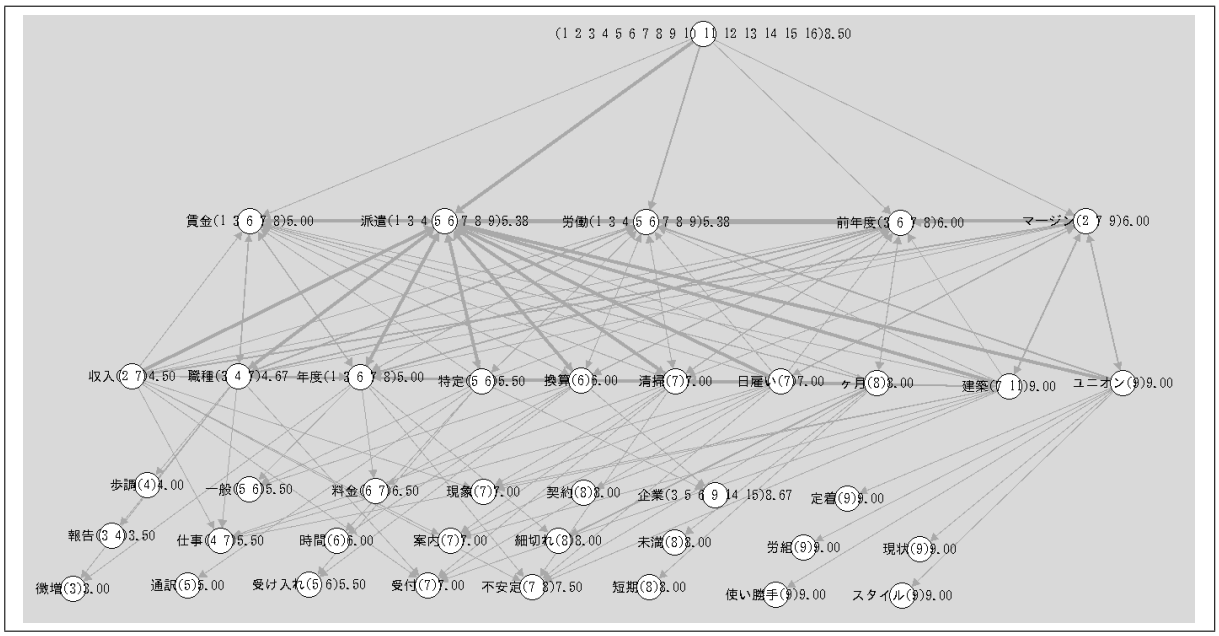


図 5.3: 「派遣労働者」に関する新聞記事のネットワーク

派遣労働者の数が増え続けているのはなぜですか。
→「安い労働力が欲しい」という企業側の需要があるから。

図 5.4: 「派遣労働者」に関する新聞記事の問題例と解答例

06年度は321万人に　派遣労働者の賃金は？
マージン膨らみ、収入は減少

派遣で働く労働者が06年度に過去最大の約321万人（前年度比26・1％増）となったことが、厚生労働省がまとめた06年度の「労働者派遣事業報告」で分かった。派遣会社の年間売上高も前年度比34％も増えて5兆円を超えた。「安い労働力が欲しい」という企業側の需要を反映した結果だが、派遣労働者が受け取る賃金は微増にとどまっており、前年度を下回る職種もある。派遣労働者の賃金事情は、どうなっているのか。

労働者派遣事業報告によると、かつては専門的な仕事に限って労働者派遣が認められていた職種が、規制緩和で次々と広がったのに歩調を合わせるように、派遣労働者の数はほぼ一貫して増え続けている。

派遣会社の事業所も、事務や軽作業などの「一般派遣」が1万8028所（同22・7％増）、通訳など専門職の「特定派遣」も2万3938所（43・6％増）に増えた。一方、派遣労働者を受け入れている企業も約86万件（同30・4％増）になった。

受け入れ企業が派遣会社に支払う派遣料金（8時間換算）は「一般派遣」が1万5577円（前年度比2・1％増）、「特定派遣」が2万2948円（同0・3％減）。これに対し、派遣労働者が受け取った賃金（8時間換算）は、「一般派遣」で1万571円（同0・5％増）、「特定派遣」では1万4156円（0．7％減）だった。

不安定な日雇い派遣が多い建築物清掃の仕事に焦点を当てると、派遣労働者が受け取った賃金は賃金調査が始まった04年度から下がり続けており、06年度では6995円と前年度より8・7％も減った。しかし、建築物清掃の仕事の派遣料金は1万1303円と前年度から2．6％上がっていた。派遣労働者の収入は減る一方で、派遣会社の収益（マージン）は膨らんでる構造だ。電話対応や案内、受付など6職種で、同じ現象が起こっている。

派遣契約が、以前より細切れになっている現実もある。3ヶ月未満の短期契約が全体の81％（前年度73％）を占めるようになった。派遣労働者が増える一方で、低賃金、不安定化が進んでいるのだ。

これらの結果に、派遣労働者を組織する労組の「派遣ユニオン」は「派遣は企業にとって使い勝手が良いという雇用スタイルが定着し、労働者にとって厳しい現状になってしまった。派遣会社が手にするマージンを規制するなどして、労働者保護を進めるべきだ」と話している。

図 5.5: 「派遣労働者」に関する新聞記事の本文

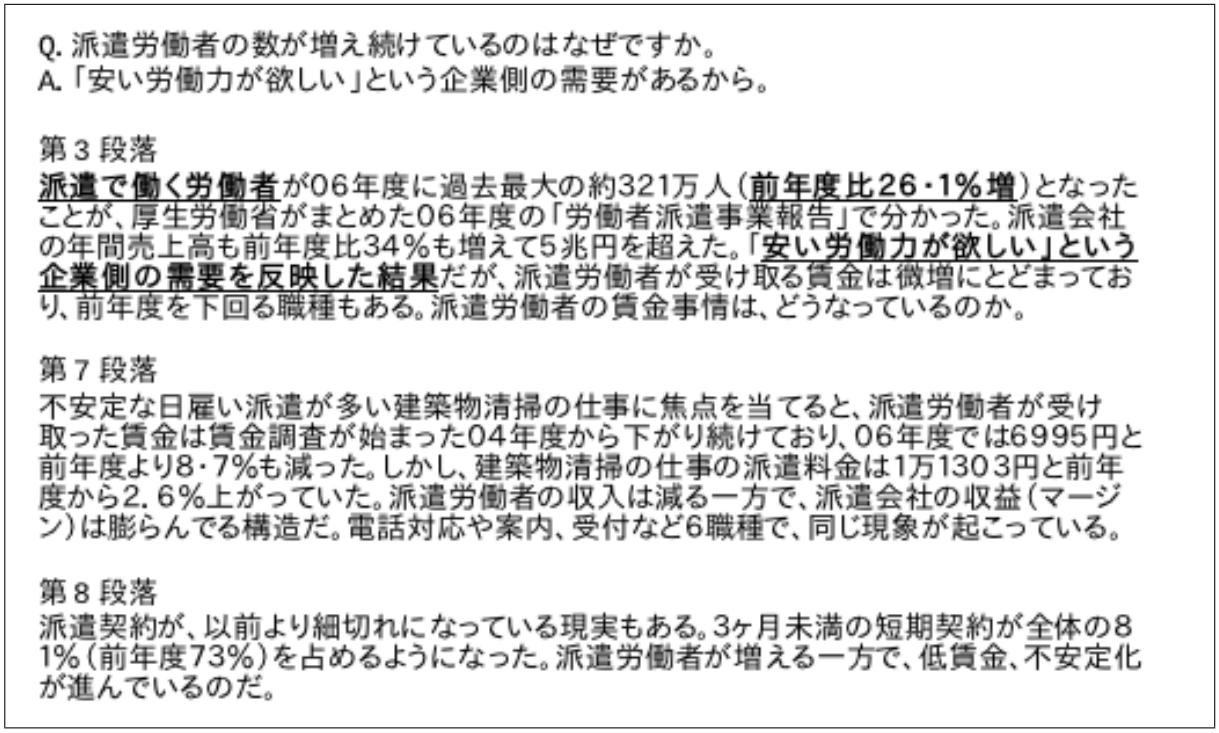
5.2.2 評価方法

単一文書を入力として単語ネットワークを構築し，実際に利用した際の文書内容の把握度で評価を行う．以下に評価方法の手順を示す．

1. 新聞記事を入力データとして単語ネットワークを構築する．
2. 構築されたネットワークの 2 階層目の 5 つのノードに注目して，そのノードに付与された段落番号の出現回数を集計する．
3. 2 で集計した段落番号で頻度が高い 3 つの段落を本文から抽出する．
4. 3 で抽出した段落の内容のみを用いて，入力データの内容に関する問題に回答する．回答する際に 3 で抽出した段落が役立った場合に正解とする．
5. 3 で抽出した段落の文字数と問題の正解数から，読書率と正解率を算出する．読書率の算出には式 5.1 を用いる．

$$\text{読書率} = 3\,\text{段落の合計文字数}/\text{入力データの合計文字数} \tag{5.1}$$

また，入力データの全段落からランダムに選出した 3 段落のみを用いた場合で読書率，正解率を算出する．段落の選出をそれぞれの記事に対し 10 回行い，その平均値を算出し手順 5 で算出した評価値と比較する．また，3 つの段落を選出した際に文字数が最小になる場合と最大になる場合，新聞記事の上から 3 つの段落を選出した場合についても読書率，正解率を算出する．なお，採点基準は文献 [13] の解答例に記されている事柄が，抽出された段落に含まれているかとする．「派遣労働者」に関する新聞記事を入力とした場合の単語ネットワークで役立つと判断した場合の例を図 5.6 に示す．図 5.6 の下線部が問題の回答に役立つと判断した理由である．


Q. 派遣労働者の数が増え続けているのはなぜですか。
A.「安い労働力が欲しい」という企業側の需要があるから。

第 3 段落
派遣で働く労働者が06年度に過去最大の約321万人（前年度比26・1%増）となったことが、厚生労働省がまとめた06年度の「労働者派遣事業報告」で分かった。派遣会社の年間売上高も前年度比34%も増えて5兆円を超えた。「安い労働力が欲しい」という企業側の需要を反映した結果だが、派遣労働者が受け取る賃金は微増にとどまっており、前年度を下回る職種もある。派遣労働者の賃金事情は、どうなっているのか。

第 7 段落
不安定な日雇い派遣が多い建築物清掃の仕事に焦点を当てると、派遣労働者が受け取った賃金は賃金調査が始まった04年度から下がり続けており、06年度では6995円と前年度より8・7%も減った。しかし、建築物清掃の仕事の派遣料金は1万1303円と前年度から2. 6%上がっていた。派遣労働者の収入は減る一方で、派遣会社の収益（マージン）は膨らんでる構造だ。電話対応や案内、受付など6職種で、同じ現象が起こっている。

第 8 段落
派遣契約が、以前より細切れになっている現実もある。3ヶ月未満の短期契約が全体の81%（前年度73%）を占めるようになった。派遣労働者が増える一方で、低賃金、不安定化が進んでいるのだ。

図 5.6: 選出した段落が役立つと判断した場合の例

5.2.3 評価結果

全 16 問の問題について，頻度上位 3 つの段落を読むことで，約 44%の読書率で約 44%の正解率を確認できた．また，ランダムについては 10 回の平均値で約 42%の読書率で約 32%の正解率を確認できた．文字数最小の場合に約 29%の読書率に対して，約 31%の正解率，文字数最大の場合に約 53%の読書率に対して，約 56%の正解率を確認できた．新聞記事の上から 3 つの段落を抽出した場合に約 44%の読書率に対して，約 44%の正解率を確認できた．

表 5.1: 問題の正解数，および文書の文字数

	正解数	文字数
頻度上位	7.0	2,901.0
ランダム	5.1	2,564.5
文字数最小	5.0	1,929.0
文字数最大	9.0	3,499.0
上から 3 段落	7.0	2,942.0

表 5.2: 問題の正解率，および読書率

	正解率	読書率	正解率／読書率
頻度上位	0.44	0.44	1.01
ランダム	0.32	0.42	0.83
文字数最小	0.31	0.29	1.08
文字数最大	0.56	0.53	1.07
上から 3 段落	0.44	0.44	0.99

5.3 単語の出現範囲の評価

単一文書において，単語の出現箇所の推定が内容把握につながると考えた．そこで，単一文書を入力として単語ネットワークを構築した際に得られる単語の本文中における出現範囲を評価する．

5.3.1 実験条件

本実験では，小説を入力とした単語ネットワークを構築する．実験に用いる小説は,「怪人二十面相」,「こころ」,「吾輩は猫である」,「人間失格」,「銀河鉄道の夜」,「坊ちゃん」の 6 つである．表 5.3 に実験に用いた小説の文字数，文数，段落数を示す．なお，IDF の算出には毎日新聞 2012 年度 (110,587 記事) を用いる．

表 5.3: 小説の詳細なデータ

タイトル	文字数 (文字)	文数 (文)	段落数 (個)
怪人二十面相	110,827	3,164	1,731
こころ	182,012	4,654	1,570
吾輩は猫である	366,158	7,487	2,234
人間失格	72,987	1,147	829
銀河鉄道の夜	41,464	1,120	487
坊ちゃん	103,221	2,451	514

5.3.2 評価方法

単語ネットワークを構築した際に得られるノードから TF-IDF 上位 5 単語を選出し，段落単位で見た場合に出現範囲が正規分布に基づいているとして，出現箇所の範囲を推定する．単語の出現範囲は以下の式で算出する．本実験では，信頼率 95%を設定して z=1.96 を 5.2 式に適用する．

$$\text{平均値} - (\text{標準偏差} * z) < \text{出現範囲} < \text{平均値} + (\text{標準偏差} * z) \tag{5.2}$$

算出された単語の出現範囲に，実際に単語が出現する割合を算出する．全 30 個のノードから得られる割合の平均値を評価値とする．

5.3.3 評価結果

95%の信頼値で出現範囲を求めると，出現範囲内に平均で約 98%の確率で単語が出現することを確認できた．実験で得られた各記事のデータを表 5.4 から表 5.9 に示す．

表 5.4: TF-IDF 上位 5 単語の出現範囲と出現割合 (入力:「怪人二十面相」)

単語	平均値	標準偏差	出現範囲	範囲内の個数	全出現個数	出現割合
明智	1,175.66	3,40.90	507.50 — 1,843.82	242	253	0.96
探偵	1,051.28	396.27	274.59 — 1,827.97	149	151	0.99
老人	723.88	269.57	195.52 — 1,252.24	102	108	0.94
少年	803.20	445.95	-70.86 — 1,677.26	104	104	1.00
小林	873.71	393.40	102.65 — 1,644.78	106	113	0.94

表 5.5: TF-IDF 上位 5 単語の出現範囲と出現割合 (入力:「こころ」)

単語	平均値	標準偏差	出現範囲	範囲内の個数	全出現個数	出現割合
奥さん	835.43	503.67	-151.76 — 1,822.62	198	198	1.00
先生	441.00	268.91	-86.06 — 968.06	289	311	0.93
お嬢さん	1,343.83	103.89	1,140.21 — 1,547.45	90	90	1.00
自分	1,064.64	411.33	258.43 — 1,870.85	177	189	0.94
心持	1,026.35	466.59	111.83 — 1,940.87	50	52	0.96

表 5.6: TF-IDF 上位 5 単語の出現範囲と出現割合 (入力:「吾輩は猫である」)

単語	平均値	標準偏差	出現範囲	範囲内の個数	全出現個数	出現割合
主人	823.49	628.31	-408.00 — 2,054.98	363	363	1.00
吾輩	544.61	508.84	-452.72 — 1,541.94	158	158	1.00
寒月	1,091.17	806.69	-489.94 — 2,672.28	161	161	1.00
先生	1,065.36	680.54	-268.50 — 2,399.21	197	197	1.00
東風	1,530.84	691.09	176.30 — 2,885.38	66	74	0.89

表 5.7: TF-IDF 上位 5 単語の出現範囲と出現割合 (入力:「人間失格」)

単語	平均値	標準偏差	出現範囲	範囲内の個数	全出現個数	出現割合
自分	373.60	254.05	-124.34 — 8,871.54	306	306	1.00
堀木	435.52	192.53	58.16 — 812.88	67	67	1.00
ヒラメ	475.41	198.07	87.19 — 863.63	32	32	1.00
道化	186.81	169.42	-145.25 — 518.87	32	32	1.00
ヨシ子	714.89	62.79	591.82 — 837.96	19	19	1.00

表 5.8: TF-IDF 上位 5 単語の出現範囲と出現割合 (入力:「銀河鉄道の夜」)

単語	平均値	標準偏差	出現範囲	範囲内の個数	全出現個数	出現割合
カム	271.24	137.69	1.37 — 541.11	78	78	1.00
パネル	271.24	137.69	1.37 — 541.11	78	78	1.00
汽車	292.52	102.6	91.42 — 493.73	30	31	0.97
向う	276.19	110.99	58.65 — 493.73	35	36	0.97
天の川	284.94	134.83	20.67 — 549.21	18	18	1.00

表 5.9: TF-IDF 上位 5 単語の出現範囲と出現割合 (入力:「坊ちゃん」)

単語	平均値	標準偏差	出現範囲	範囲内の個数	全出現個数	出現割合
山嵐	288.99	152.91	-10.71 — 588.69	78	78	1.00
シャツ	260.95	133.33	-0.38 — 522.28	91	91	1.00
校長	214.37	141.28	-62.54 — 491.28	42	43	0.97
下宿	229.22	143.05	-51.16 — 509.60	36	37	0.97
喧嘩	244.11	157.89	-65.35 — 553.57	27	27	1.00

5.4　ノード対の有用性の評価

単一文書を単語ネットワークとして可視化し，実際に単語ネットワークを見たときに有益な情報が得られたかを確認するため，3.2 節の手法で得られた，リンク情報を元に内容把握に役立つノードの評価を行う．

5.4.1　実験条件

本実験では，小説を入力とした単語ネットワークを構築する．実験に用いる小説は,「怪人二十面相」,「こころ」,「吾輩は猫である」,「人間失格」,「銀河鉄道の夜」,「坊ちゃん」の 6 つである．なお，IDF の算出には毎日新聞 2012 年度 (110,587 記事) を用いる．5.3.1 節の表 5.3 に実験に用いた小説の文字数，文数，段落数を示す．

5.4.2　評価方法

単語ネットワークを構築した際に得られるノードの内，第 2 階層と第 3 階層に出現する登場人物や事柄のノードが小説の内容を把握するために役立つかを評価する．第 2 階層と第 3 階層に出現するノードのいずれかの 2 つのみからなるノード対について，リンク情報を 3.2 節の手法により抽出する．抽出されたリンク情報を用いて，役立つかの判断を人手で行う．判断に用いるリンク情報の文は，3.2 節の優先度の式 3.2 で順位付けされた文のうち，上位 5 文とする．リンク情報が以下の 3 種類の項目のいずれかについて記載されていた場合，役立つと判断する．判断理由 1 については 2 種類のパターンを用いて，役立つかを判断する．

1. 登場人物の特徴の説明
    a). ノード A を登場人物としたとき，ノード B がノード A の特徴を示す場合
    b). ノード A を登場人物としたとき，ノード B はノード A の特徴を示さないが，リンク情報より特徴が得られた場合
2. 2 人の登場人物の関係性
3. 物語における有益な内容

1 つの単語ネットワークで 5 つ以上有益なノード対を獲得できた場合，単語ネットワークが内容把握として有効であるとする．

5.4.3 評価結果

実験に用いた小説を入力として構築した単語ネットワークの第 2 階層と第 3 階層の，ノード数，ノード対数，リンク情報を持つノード対数，リンク情報数を表 5.10 に示す．

表 5.10: 小説を入力として得られた単語ネットワークのデータ

タイトル	全ノード数 (個)	全ノード対 (個)	リンク情報を持つノード対 (個)	全リンク情報数 (文)
怪人二十面相	22	86	80	298
こころ	19	85	77	298
吾輩は猫である	20	93	77	305
人間失格	23	86	74	271
銀河鉄道の夜	17	73	60	143
坊ちゃん	22	76	59	198

有益なノード対を調査した結果，6 つの小説すべてで，有益な情報を含むノード対を 5 つ以上確認できた．表 5.11 から表 5.16 に有益なノード対として獲得できた結果を示す．表に記載されているリンク情報は，抽出された優先度上位 5 文のうち，対象のノード対が役立つと判断した例として 1 文のみを示している．

表 5.11: 小説「怪人二十面相」の有益なノード対および役立つと判断した情報と判断理由

ノード A	ノード B	リンク情報	判断理由
小林	助手	大探偵明智小五郎には、●小林●芳雄（こばやしよしお）という■少年■助手があります	1a
小林	探偵	この小学生たちは、●小林●芳雄君を団長にいただく、あの少年■探偵■団でありました	1a
明智	小林	●明智●の少年助手の■小林■芳雄とかいったっけな	2
明智	乞食	■乞食■に化けた男は、●明智●探偵誘かいのしだいと、赤井寅三を味方にひきいれた理由を、くわしく報告しました	3
盗賊	明智	■明智■小五郎とばかり思いこんでいた男が、名探偵どころか、大●盗賊●だったのです	3

表 5.12: 小説「こころ」の有益なノード対および役立つと判断した情報と判断理由

ノード A	ノード B	リンク情報	判断理由
奥さん	心持	男のように判然（はきはき）したところのある●奥さん●は、普通の女と違ってこんな場合には大変■心持■よく話のできる人でした	1a
叔父	自分	父はよく■叔父■を評して、●自分●よりも遥（はる）かに働きのある頼もしい人のようにいっていました	1b
先生	挨拶	■先生■が私に示した時々の素気（そっけ）ない●挨拶●（あいさつ）や冷淡に見える動作は、私を遠ざけようとする不快の表現ではなかったのである	1b
奥さん	自分	●奥さん●も■自分■の夫の所へ来る書生だからという好意で、私を遇していたらしい	2
奥さん	様子	帰っても何にもない、あるのはただ父と母の墓ばかりだと告げた時、●奥さん●は大変感動したらしい■様子■を見せました	3

表 5.13: 小説「吾輩は猫である」の有益なノード対および役立つと判断した情報と判断理由

ノード A	ノード B	リンク情報	判断理由
寒月	ヴァイオリン	「落第の候補者■寒月■君は●ヴァイオリン●の妙手だよ	1a
主人	逆上	前（ぜん）申す通り■主人■は立派なる●逆上●家である	1a
主人	寒月	「そうさな」と■主人■は武右衛門君の哀願に冷淡であるごとく、●寒月●君の探検にも冷淡である	1b
吾輩	主人	●吾輩●の■主人■は滅多（めった）に●吾輩●と顔を合せる事がない	2
寒月	金田	三角主義の張本■金田■君の令嬢阿倍川の富子さえ●寒月●君に恋慕したと云う噂（うわさ）である	2

表 5.14: 小説「人間失格」の有益なノード対および役立つと判断した情報と判断理由

ノード A	ノード B	リンク情報	判断理由
自分	道化	そうして●自分●は、この■道化■の一線でわずかに人間につながる事が出来たのでした	1a
自分	煙草	■自分■が、●煙草●を買いに行くたびに、笑って忠告するのでした	1a
自分	お金	「■自分■でかせいで、その●お金●で、お酒、いや、煙草を買いたい	1b
ヒラメ	自分	その男の顔が、殊に眼つきが、●ヒラメ●に似ているというので、父はいつもその男を●ヒラメ●と呼び、■自分■も、そう呼びなれていました	1b
堀木	座蒲団	■堀木■は、■堀木■の家の品物なら、●座蒲団●の糸一本でも惜しいらしく、恥じる色も無く、それこそ、眼に角（かど）を立てて、自分をとがめるのでした	1b

表 5.15: 小説「銀河鉄道の夜」の有益なノード対および役立つと判断した情報と判断理由

ノード A	ノード B	リンク情報	判断理由
パネル	女の子	カム●パネル●ラだってあんな■女の子■とおもしろそうに談（はな）しているし僕はほんとうにつらいなあ	2
硝子	汽車	すきとおった●硝子●（ガラス）のような笛が鳴って■汽車■はしずかに動き出し、カムパネルラもさびしそうに星めぐりの口笛を吹きました	3
パネル	汽車	カム●パネル●ラのうちにはアルコールラムプで走る■汽車■があったんだ	3
汽車	天の川	もうそして■天の川■は●汽車●のすぐ横手をいままでよほど激（はげ）しく流れて来たらしくときどきちらちら光ってながれているのでした	3
停車場	汽車	そのとき■汽車■はだんだんしずかになっていくつかのシグナルとてんてつ器の灯を過ぎ小さな●停車場●にとまりました	3

表 5.16: 小説「坊ちゃん」の有益なノード対および役立つと判断した情報と判断理由

ノード A	ノード B	リンク情報	判断理由
シャツ	教頭	それでこそ一校の■教頭■で、赤●シャツ●を着ている主意も立つというもんだ	1a
シャツ	馬鹿	お婆さん、あの赤●シャツ●は■馬鹿■ですぜ	1a
山嵐	生徒	淡泊（たんぱく）だと思った●山嵐●は■生徒■を煽動（せんどう）したと云うし	1b
マドンナ	シャツ	顔はふくれているが、こんな結構な男を捨てて赤■シャツ■に靡（なび）くなんて、●マドンナ●もよっぽど気の知れないおきゃんだ	2
教頭	山嵐	「●教頭●の職を持ってるものが何で角屋へ行って泊（とま）った」と■山嵐■はすぐ詰（なじ）りかけた	3

第6章　考察

6.1　内容把握についての考察

内容把握の実験の評価について考察を行う．新聞記事を入力として単語ネットワークを構築し，2階層目のノードで段落の出現数の頻度上位3つを評価したところ，ランダムに3段落を選出したものと比べ，読書率にほとんど差は見られないが，正解率が約12%高くなることを確認できた．頻度上位で段落を3つ選出した場合は読書率と正解率にほとんど差はないが，ランダムで段落を3つ選出した場合には，読書率に対して正解率が約11%低い結果となった．頻度上位では段落の選出をする際に，TF-IDFの値から得られたノードを元に段落番号を選出しているため比較的重要な段落が選出されていると考えられる．それに対し，ランダムに選出した場合には，重要度が考慮されていないため，正解率に差が出たと考えられる．

また，段落を3つ選出した際にその文字数の合計が最小になる場合と最大になる場合について評価した結果，それぞれ，読書率と正解率の差は約3%以下となった．この結果から，読書率に対して，ほとんど同等の値で正解率を獲得できることが考えられる．

なお，本実験で用いた問題は文献[13]に用いられているものであり，重要な内容についての問題であると仮定した．そのため，新聞記事で必ずしも重要な内容であるとは限らない可能性が考えられる．最も重要である内容を問題として提示することができれば，正解率が変化する可能性があるため，今後の検討が必要である．

その他の今後の課題として，段落の選出に用いるノードを2階層目以外も候補に入れ，問題文に関連する段落を取り出すことが考えられる．

6.2 単語の出現範囲の考察

単語の出現範囲についての考察を行う．95%の信頼値で出現範囲を求めると，出現範囲内に平均で約 98%の確率で単語が出現することを確認できた．

また，標準偏差の値が低いほど，推定する出現範囲が狭くなることを確認した．例えば，表 5.7 の「ヨシ子」という単語は標準偏差が 62.79 で推定した出現範囲が第 592 段落から第 837 段落となっている．全段落数が 829 段落であるので物語の後半で登場することがわかる．図 6.1 は実際に出現段落を調査した結果である．TF-IDF の値を用いて重要度を算出しているため，TF-IDF 上位 5 単語に含まれる「ヨシ子」はこの物語において重要である人物であると考えられる．

また，表 5.4 の「小林」などの標準偏差の値が高く，平均値が全段落数の中心値に近い場合は，物語の重要人物の中でも，最初から最後まで多く登場している人物であると予測できる．図 6.2 は実際に登場人物「小林」の出現段落を調査した結果である．

図 6.1，図 6.2 は対象の単語が出現した段落を 1 とした場合の分布である．図の横軸は段落番号を示す．

以上の結果から，登場人物やある事柄など，単語ネットワークでノードとして出力されている単語の出現段落の推定が可能となることが確認できた．

今後の課題として，登場人物についての情報を提示し，登場段落での行動などが把握できれば読書支援に役立つと考えられる．そのためには，名詞のみをノードとして単語ネットワークに出力するだけでは，情報に偏りが出る可能性があるため，動詞や形容詞といった他の品詞をノード候補にすることが考えられる．

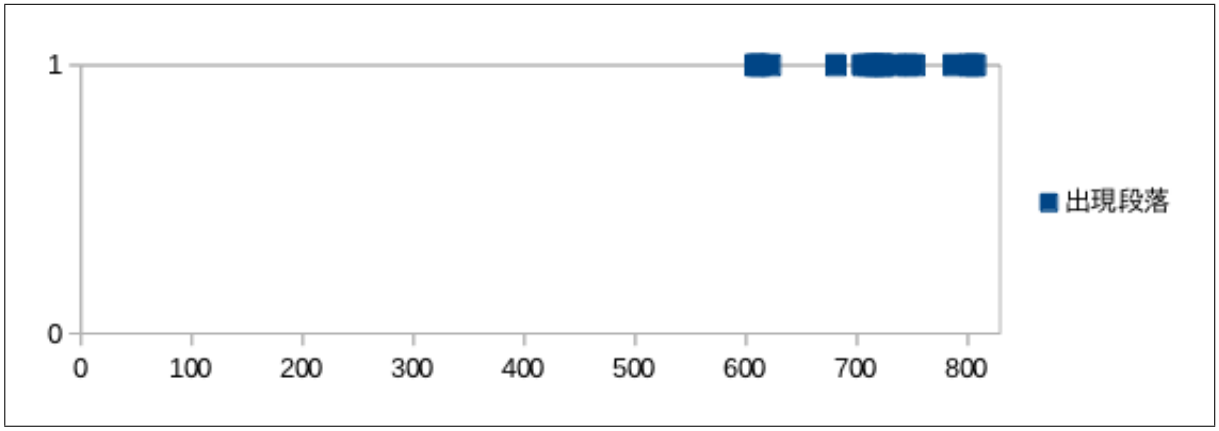


図 6.1: 小説「人間失格」における登場人物「ヨシ子」の出現段落の分布

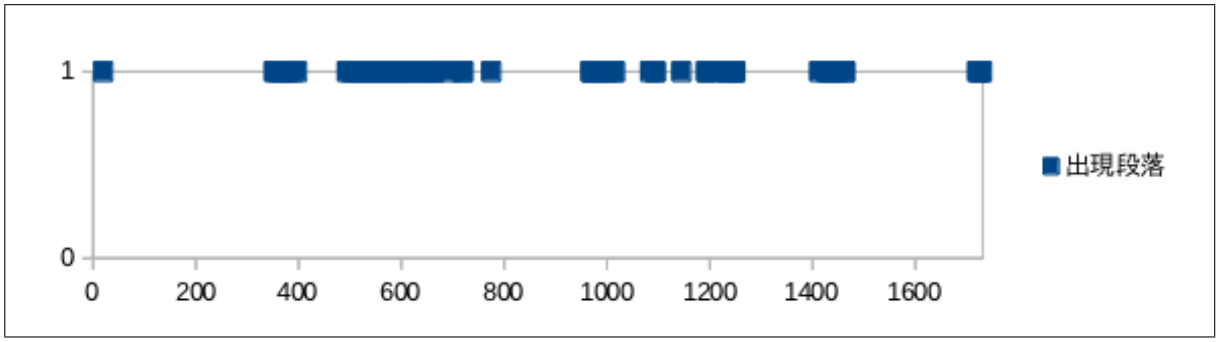


図 6.2: 小説「怪人二十面相」における登場人物「小林」の出現段落の分布

6.3　ノード対の有用性の考察

単一文書を入力として単語ネットワークを構築し，得られたノード対の有用性についての考察を行う．有益なノード対を調査した結果，入力データに用いた 6 つの小説すべてで，有益な情報を含むノード対を 5 つ以上確認できた．

判断理由として用いた 3 つの条件から，ノード対とそのリンク情報を見ることで，登場人物の特徴，2 人の登場人物の関係性，物語における有益な情報のいずれかを獲得が可能であると確認できた．

特に上位の階層に登場人物のノードが多く出現していることから，その人物の特徴について多く獲得できた．登場人物の特徴を掴むことで，物語の大枠を捉えられる可能性もあるため読書支援にも有効であると考えられる．

また，入力データを「銀河鉄道の夜」とした場合に,「カム」「パネル」というノード対が出現した．この物語には「カムパネルラ」という人物が登場する.「カム」「パネル」というノード対は人物の名前を説明しているとも捉えられる．リンク情報として「では●カム●■パネル■ラさん」という文が得られたことからも，判断理由 1b を満たし，有益なノード対とすることも考えられる．しかし，本来は「カムパネルラ」という人物が「カム」「パネル」のように 2 つに分かれてノードとして出現することは避けたいと考えたため，今回の実験では有益なノード対としては除外した．このような，名詞が連続した単語は連結して，1 つのノードとして単語ネットワークに出力することで，より正確な情報を取得できると考えられる．実装に伴い，名詞連続の連結により，連結前の単語の DF の値に変化が生じることなど，問題が発生する恐れがあるため，連結，非連結の場合の単語ネットワーク出力結果を比較し，どちらが有効かを確認する必要があると考え，今後の課題としたい．

その他の課題として，実験に用いた，判断理由 3 については，実験者の主観による部分が反映され，曖昧性を生む可能性があるため，実験者を増やして実験を行うことが考えられる．

第7章　おわりに

先行研究では，大竹ら [1] は，電子テキストから特定のキーワードに基づく関係情報をネットワークとして抽出する方法を提案し,「地震」というキーワードに基づいて単語ネットワークの構築を行った．土遠ら [2] は，大竹らが構築したネットワークに関連のない事物のノードを含むことを確認し，それらのノードを削除を行った．窪ら [3] は，土遠ら [2] が構築したネットワークのリンクに文字列を付与することで，ノード間の関係を分かりやすくした．しかし大竹ら，土遠ら，窪らが構築したネットワークは，複数文書のみを扱う手法で，単一文書の入力には対応していないという問題があった．

そこで本研究では，単一文書を入力とした単語ネットワークを構築する手法を提案した．

その結果，単一文書を入力として単語ネットワークを構築し，単一文書を可視化させることができた．

また，単一文書を入力として構築された単語ネットワークの利用について，内容把握度と単語の出現範囲の推定，ノード対の有用性で評価を行った．

内容把握度について，2 階層目の 5 つのノードにおいて，段落の出現回数頻度上位 3 つ用いた場合には，約 44%の読書率で約 44%の正解率を確認した．読書率に対する正解率の割合が 1.00 となり，ランダムで段落を 3 つ抽出した場合の割合が 0.83 であることから，提案手法で抽出した段落を読んだ方が，内容の把握度が高くなることが確認できた．

単語の出現範囲の推定について，95%の信頼値の正規分布に基づいているとして，約 98%の確率で単語が区間内に出現することを確認した．これにより，登場人物やある事柄など，単語ネットワークでノードとして出力されている単語の出現段落の推定が可能となった．

ノード対の有用性について，小説を入力とした単語ネットワークにおいて，6 つの小説の全てで，5 つ以上の有益なノード対が獲得できることを確認した．これにより，登場人物の特徴，2 人の登場人物の関係性，物語における有益な情報のいずれかを獲得することができる．特に登場人物についての情報が多く獲得できたため，物語の大枠を捉えられる可能性から，読書支援にも有効であると考えられる．

今後は，単一文書を入力として単語ネットワークを構築する際，動詞や形容詞の重要

な単語が存在する場合も考えられるため，様々な品詞をノードとして獲得できるよう改良したい．また，名詞連続の単語を連結することで，単語ネットワークとしての利便性を向上させたい．

謝辞

最後に，1 年間の間，研究を進めるに当たり，本研究のご指導を頂きました鳥取大学工学部知能情報工学科自然言語処理研究室の村田真樹教授，村上仁一准教授そして自然言語処理研究室の皆様に深く感謝するとともに心から御礼申し上げます．また，論文を引用させて頂きました窪氏を始めとして，参考にさせていただいた論文の著者の方々に対して深く感謝申し上げます.

参考文献

[1] 大竹竜太, 村田真樹, 徳久雅人. 大規模テキストデータを用いた社会構造ネットワークモデルの自動抽出. 言語処理学会第 19 回年次大会発表論文集, pp. 798–801, 2013.

[2] Y. Doen, M. Murata, R. Otake, and M. Tokuhisa. Construction of concept network from large numbers of texts for information examination using tf-idf and deletion of unrelated words. SCIS&ISIS, pp. 1108–1113, 2014.

[3] 窪雄平. テキスト処理に基づく概念ネットワークの構築におけるリンクへの文字列付与. 鳥取大学卒業論文, 2015.

[4] 内山将夫, 橋田浩一. Gda タグを利用した複数文書の要約. 言語処理学会第 6 回年次大会発表論文集, pp. 376–379, 2000.

[5] 松尾豊, 友部博教, 橋田浩一, 中島秀之, 石塚満. Web 上の情報からの人間関係ネットワークの抽出. 人工知能学会論文誌, Vol. 20, No. 1, pp. 46–56, 2005.

[6] 松尾豊, 大澤幸, 石塚満. Small world 構造に基づく文書からのキーワード抽出. 人工知能学会論文誌, Vol. 43, No. 6, pp. 1825–1833, 2002.

[7] 瀧川和樹, 村田真樹, 土田正明, De Saeger Stijn, 山本和英, 鳥澤健太郎. 連想知識を用いた端的な要約の生成. 言語処理学会第 16 回年次大会発表論文集, pp. 298–301, 2010.

[8] 西川仁, 平尾努, 牧野俊朗, 松尾義博, 松本裕治. 冗長性制約付きナップサック問題に基づく複数文書要約モデル. 自然言語処理学会論文誌, Vol. 20, No. 4, pp. 585–612, 2013.

[9] 森辰則, 野澤正憲, 浅田義昭. 質問応答エンジンを利用した複数文書要約手法. 言語処理学会第 10 回年次大会発表論文集, pp. 289–292, 2004.

[10] 松井兵庫, 阪本浩太郎, 松永詠介, 神貴久, 渋木英潔, 石下円香, 森辰則, 神門典子. 大学入試の穴埋め問題を解く質問応答システムの検討. 言語処理学会第 21 回年次大会 発表論文集, pp. 175–178, 2015.

[11] 縣啓示, 伊藤雄一, 高島和毅, 北村喜文, 岸野文郎. 物語テキストから進行状況に応じて登場人物の存在状態と関係を推定する手法. WISS, pp. 101–106, 2010.

[12] 西原弘真, 白井清昭. 物語テキストを対象とした登場人物の関係抽出. 言語処理学会第 21 回年次大会発表論文集, pp. 628–631, 2015.

[13] 内田安伊子, 内田紀子. 構成・特徴・分野から学ぶ新聞の読解. 2008.