修 士 論 文 の 和 文 要 旨

研究科・専攻	大学院情報システム学研究科　　社会知能情報学専攻　博士前期課程		
氏　　名	横尾　亮平	学籍番号	1351020
論 文 題 目	語句間の意味構造に基づくニュース記事推薦システムの提案		

要　旨

近年，ユーザの行動やソーシャルメディア上での発言を興味関心として分析し，ニュース記事を推薦するキュレーションサービスが普及している．膨大な情報から自分で必要なものを探さなくても，自身の興味に沿った情報が手に入ることで利用者が増加している．

既存のコンテンツベースの情報推薦システムに関する研究では記事推薦のために各語句を特徴としているが，頻出する語句を重要視しており語句間の関係を特徴として用いていない．

本研究は，ユーザが興味関心を示す記事に表れる語句間の意味構造を用いることで，ユーザが面白いと感じることができるニュース記事を収集，推薦するシステムを提案する．本研究では面白いニュース記事をユーザが興味を示すことができ，意外な情報が得られるものと定義した．語句間の意味構造 Linked Data で表現する．同ニュース記事の同文脈に表れる複数の語句間の意味構造を文構造と定義する．ユーザが興味・関心を示す記事文の文構造の部分グラフを用いることでインターネット上のニュース記事を推薦する手法を提案する．

本手法の有効性を確かめるため，20 人の被験者に提案手法，ベースライン手法それぞれによるニュース記事推薦をして評価を得る比較実験を行った．ベースライン手法は単語の重要度を出現頻度から計算する tf-idf を用いた．提案手法によるニュース記事推薦での関連度の指標の平均値は 4 点満点中 3.06，興味度は 3.30，意外度は 2.93 という結果であった．ベースライン手法では関連度が 3.22，興味度が 3.03，意外度が 2.79 という結果であった．

ベースライン手法との比較実験により，提案手法は推薦するニュース記事の関連度は下がるものの，ユーザが興味を持つことができ，また意外と感じることができるニュース記事推薦手法であることがわかった．これによりユーザに面白い記事を推薦できる手法として提案手法は有効であることが明らかになった．

平成２６年度修士論文

語句間の意味構造に基づく
ニュース記事推薦システムの提案

電気通信大学 大学院情報システム学研究科
社会知能情報学専攻

氏名： 横尾 亮平
学籍番号：1351020

主任指導教員： 大須賀 昭彦 教授
指導教員： 植野 真臣 教授
指導教員： 田原 康之 准教授

平成27年2月27日（金） 提出

概 要

近年，ユーザの行動やソーシャルメディア上での発言を興味·関心として分析し，ニュース記事を推薦するキュレーションサービスが普及している．膨大な情報から自分で必要なものを探さなくても，自身の興味に沿った情報が手に入ることで利用者が増加している．既存のコンテンツベースの情報推薦システムに関する研究では記事推薦のために各語句を特徴としているが，頻出する語句を重要視しており語句間の関係を特徴として用いていない．本研究は，ユーザが興味·関心を示す記事に表れる語句間の意味構造を用いることで，ユーザの興味·関心をより具体化することでユーザが面白いと感じることができるニュース記事を推薦する手法を提案する．本研究では面白いニュース記事の定義をユーザが興味を示すことができ，意外な情報が得られるものと定義した．語句間の意味構造は Linked Data で表現する．ユーザが興味·関心を示す記事文章からインターネット上のニュース記事を推薦する手法を提案する．本手法の有効性を確かめるため，20 人の被験者に提案手法，ベースライン手法それぞれによりニュース記事を 5 件推薦をして評価を得る比較実験を行った．ベースライン手法は単語の重要度を出現頻度から計算する tf-idf を用いた．提案手法によるニュース記事推薦での関連度の指標の平均値は 4 点満点中 3.06，興味度は 3.30，意外度は 2.93 という結果であった．ベースライン手法では関連度が 3.22，興味度が 3.03，意外度が 2.79 という結果であった．ベースライン手法との比較実験により，提案手法は推薦するニュース記事の関連度は下がるものの，ユーザが興味を持つことができ，また意外と感じることができる手法であることがわかった．また，興味度・意外度が最高値で評価された記事数はベースライン手法が平均 0.65 件に対し，提案手法は 1.15 件であった．以上より，ユーザに面白い記事を推薦できる手法として提案手法は有効であることが明らかになった．

目 次

図 目 次

表 目 次

第1章　はじめに

近年，Gunosy[1] や Vingow[2] などのニュース記事を自動で収集し，推薦するキュレーションサービスが普及し，膨大な情報から自分で必要なものを探さなくても，簡単に関心に合う情報が手に入るようになりつつある．多くは，ソーシャルメディアである Facebook やマイクロブログの Twitter と連携利用することで取得したユーザの発言から興味 · 関心を抽出，またサービス上でのユーザの記事閲覧履歴を学習することで，ユーザに最適な情報を配信している．キュレーションサービスの出現を背景に，ニュース記事は必要な情報をユーザが能動的に取得する存在から，ユーザへ自動で配信される存在へ形を変えつつある．

本研究では，ユーザが興味 · 関心を示すニュース記事文章内の語句間の意味構造に着目し，ユーザが面白いと感じることができるニュース記事を推薦することを目指す．本論文では語句間の意味構造を Linked Data で表現する．インターネット上から取得したニュース記事群とユーザが興味 · 関心を示すニュース記事群からそれぞれ Linked Data を構築する．2 つの Linked Data 間の類似する部分グラフを用いることで，ユーザが面白いと感じるニュース記事を提供する手法を提案する．例として，“クリミアの美人すぎる検事総長”[1] というニュースが存在する．昨今，インターネット上やマスコミの報道で美人すぎる市議や美人すぎる海女といった記事の出現により，美人すぎる○○という言葉が生まれ，興味 · 関心が集まっている．これには,「美人 →（職業）→ 職業名」という意味構造が存在する．美人と意外な組み合わせの検事総長という語句の意味構造から，クリミア美人検事総長に興味の矛先が向けられた．それまでは日本であまり知られていない存在であったクリミアという地名にも興味を引く結果となった．このことは他国の美人検事やクリミアでは他にどのような美人が存在するかなど，意味構造「美人 →（職業）→ 検事総長，検事総長 →（地名）→ クリミア」により連想される他ニュースへも多くの関心が集まると想像される．

[1] http://gunosy.com/
[2] https://vingow.com/

文章表現に Bag-of-Words ベクトルを用いる既存手法の場合，ユーザが興味 · 関心を示す文「政情不安の続くクリミアにおいて美人すぎる検事総長が大人気」から 3 つのユーザの興味 · 関心語を抽出すると，ユーザが興味 · 関心を示す正しい語句の組み合わせが「美人，検事総長，クリミア」であっても,「政情不安，続く，大人気」となりえる．また，記事全体から「美人，検事総長，クリミア」や「政情不安，続く，大人気」という 3 語句だけで探索すると，語句間の関係を特徴に用いていないので関係のないニュース記事まで探してしまう可能性がある．そのため，語句間の意味構造に基づきユーザの興味 · 関心事を Linked Data[3] を用いて表現することにより，ユーザの興味 · 関心の具体化を試みる．Linked Data とはデータを再利用しやすいような形で構造化し，公開 · 共有するための Web 技術である．Linked Data に変換することで計算機が扱いやすい整理された情報として利用できる他，類似する部分グラフの検索が容易になる．

本論文は以下のように構成される．2 章にて関連研究を紹介する．3 章で提案手法の概要について述べる．4 章で提案手法の有用性を示すための評価実験について示す．5 章では，本研究のまとめと今後の展望を述べる．

[3]http://linkeddata.org/

第2章　関連研究

既存のコンテンツベースの情報推薦システムに関する研究 [2][3][4] では文章内の各語句が特徴語として用いられている．特徴語を使った文章表現に Bag-of-Words ベクトルが一般的に用いられている．Bag-of-Words モデルを用いた推薦システムでは tf-idf 法やページランク法，トピックモデル法といったアルゴリズムを用いて，頻出語の特徴語に重み付けをし，重み付けが大きい語句を含む文章を推薦する．ユーザの興味 · 関心を示す文章内で頻出する語句を含む文章が推薦されているが，語句間の意味構造を用いてニュース記事を推薦しているわけではない．

本研究では日本語文章を形態素解析し，ラベルを付与し，語句と語句間の意味構造を抽出する．この解析は一般的に意味的役割付与 (Semantic Role Labeling) または述語項構造解析と呼ばれる．松林 [5] らは自動意味役割付与における意味役割の汎化を行っている．意味役割付与をする際に利用するコーパスには出現率が低い意味役割の存在が見られるため，学習の妨げになることが問題となっている．類似する意味役割を共通化することでこれを解決している．また，意味的役割付与のコーパスには概念辞書 WordNet[1] や語彙意味構造辞書 VerbNet[2]，SemLink[3] が用いられている．音声対話システムの研究分野では，記事の文章を述語項構造解析を用いて情報を抽出し，述語項構造が類似する情報推薦や検索を行う手法が提案されている [6]．Bag-of-Words モデルと比較しても，より的確な応答生成が確認されている．今村 [7] らは日本語の雑談対話に対して，ゼロ代名詞照応付き述語項構造解析を行っている．ゼロ代名詞とは代名詞のない文中に表れる音形のない代名詞のことである．

語句のセマンティクスを考慮したニュース記事の推薦手法として，Capelle ら [8] の研究

[1] http://wordnet.princeton.edu/
[2] http://verbs.colorado.edu/~mpalmer/projects/verbnet.html
[3] http://verbs.colorado.edu/semlink/

がある．Capelle らは，ユーザが読んだニュース，読んでないニュースごとに出現する語句の類似度を考慮した上でユーザにニュース記事を推薦するためのシステムを構築している．類似度算出には WordNet[4] と検索エンジンの Bing を用いている．

近年様々な情報が Linked Open Data 化 [9][10] されている．そして Linked Data に関連する研究も数多く提案されている．

ニュース記事から Linked Data を構築している研究として，Kira ら [11] は 150 年分のニュース記事の自然言語から高精度にニュース記事のトピックを抽出し，因果関係に着目した Linked Data を構築し，Linked Data の構造を用いて今後起こるであろうニュース予測を可能にする研究をしている．Ohsawa ら [12] は Facebook ページの Like 数を予測するために，DBPedia の情報を用いている．対象となる Facebook ページと他の Facebook ページの情報との類似度を DBPedia によって算出し，類似する他の Facebook ページの Like 数を用いて対象 Facebook ページの Like 数を予測するモデルを構築している．

Linked Data を用いた推薦システムとして，Elahi ら [13] は Lined Data を用いた写真推薦を行っている．Facebook と flickr のユーザ情報を RDF に変換し，ユーザの興味と写真タグに DBPedia の情報を付与し，ユーザの興味をリッチに表現した上で，ユーザの興味を推論し，写真の推薦を行うシステムを提案している．Khrouf ら [14] はイベント情報サイトのメタ情報 (場所，時間，タグ，ジャンルなど) を Linked Data 化し，データ構造の類似度を用いた手法と協調フィルタリング手法とのハイブリットによりイベント情報推薦システムを構築している．また，Mirizzi ら [15] は Wikipedia の情報を Linked Open Data 化した DBPedia 内の映画に関する情報のグラフにベクトル空間モデルを適応し，映画間の情報 (ジャンル，監督など) 類似度を算出することでユーザが興味 · 関心を示す映画を推薦システムを構築している．Passant ら [16] は LDSD(Linked Data Semantic Distance) という Linked Data 間の意味的な距離を計算する手法を提案し，DBPedia 上のバンドや歌手の情報の類似性を計算することで，ユーザが興味 · 興味を示すバンドや歌手と類似するバンドや歌手を推薦するシステムを構築し，有効性を確かめている．また Linked Data データ間の類似度を図る指標に様々な提案がされている [17].

このように，Linked Data を用いた推薦システムが多く提案されている．しかし，文章

[4]http://wordnet.princeton.edu/

内に存在する語句間の意味構造を Linked Data に変換してニュース記事推薦を行っている研究はない．本論では語句間の意味構造を特徴に利用することで Bag-of-Words モデルを用いた既存手法よりも，ユーザの興味 · 関心を具体化することができ，ユーザが面白いと感じるニュース記事を推薦することを示す.

第3章　語句間の意味構造を用いた記事推薦手法の概要

本研究は，ユーザの興味·関心を示すニュース記事の語句間の意味構造を抽出する．語句間の意味構造を用いることでユーザが面白いと感じることができるニュース記事をユーザに推薦することを目的とする．また同ニュース記事同文脈に複数の語句間の意味構造が存在する場合，これを文構造と呼ぶ．提案手法の概要を図 3.1 に示す．(1)-まず，ユーザの興味·関心を示す 1 つのニュース記事を取得する．(2)-取得した記事を用いて，ニュース記事を構成する文ごとにユーザ嗜好 Linked Data を構築する．Linked Data は記事の文から抽出できる語句と意味構造の組み合わせとする．(3)-続いて，ユーザに推薦するための記事をインターネット上から収集する．(4)-同様にニュース記事 Linked Data を構築する．(5)-2 つの Linked Data 間で類似する部分グラフを探索する．(6)-もし，ユーザ嗜好 Linked Data と類似する部分グラフがニュース記事 Linked Data に存在する場合はこの類似する部分グラフに紐づくニュース記事をユーザに推薦する．

本研究では 2 つの Property で繋がれた 3 語句ノードの部分グラフを利用し，類似する部分グラフを用いてのニュース記事検索をしている．文中のいずれかの箇所から 3 語句を拾うのではなく，類似する部分グラフの繋がりがある 3 語句を拾うほうが，より「関連度」の高い記事を選ぶことが出来，且つその内の一箇所 (1 語句) を変数として部分グラフ検索を行うことで，関連度を維持しながら，ユーザが面白いと感じる記事を推薦できると想定した．これはユーザの興味は関連度の高い記事の隣にある（元々，興味のある内容に近く（関連度が高く），僅かに異なる内容であることが定番である）と考えたからである．例えば，元々，クリミアの美人検事に興味がある人ならば，他の国の美人検事についても同様に興味があるだろうといえる．提案手法はこれを実現するための手法である．

提案手法の詳細として，記事からの Linked Data 構築に関しては 3.1 に，類似する部分

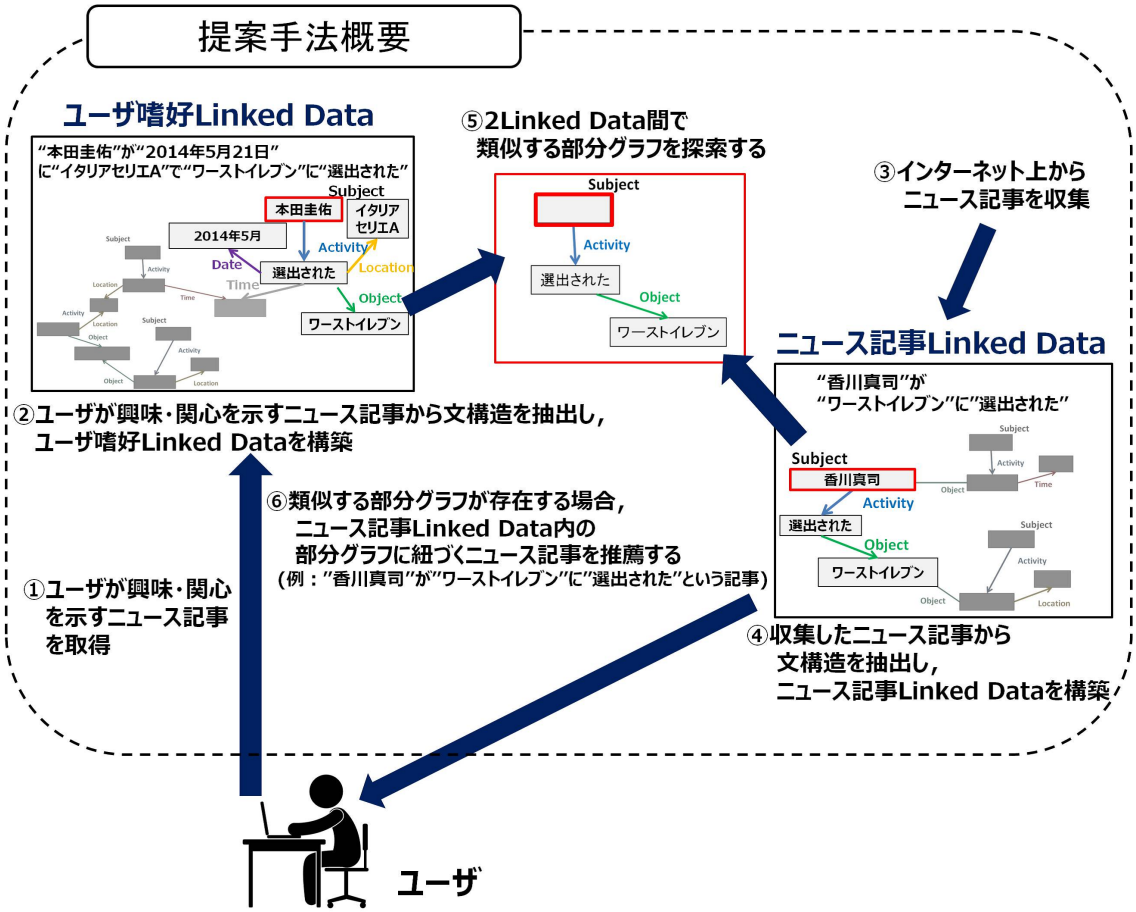


図 3.1: 提案手法の概要

グラフ検索の手法と例を 3.2 にそれぞれ示す．また，類似する部分グラフ検索数増加のための Entity Linking について 3.3 に示す．

3.1　Linked Data の構築

本研究においての Linked Data は 1 つの文脈から生成されたトリプル集合またはその部分集合を指す．トリプルとは「選出された (Activity) → Property:Object →ワーストイレブン (Object)」のように「主語 (Subject) →述語 (Property) →目的語 (Value)」の 3 つの要素でリソースに関する関係情報を表現しているメタデータモデルである．本論文では語句間の意味構造を Linked Data で表現する．

3.1.1　文構造の定義

Linked Data を構築するために，ニュース記事文から文構造を抽出する．記事文内で出現する事象が表れる語句を対象とし，語句と語句間の関係性を組み合わせたものを文構造とする．Nguyen ら [18] は，Twitter や Web ページから得られる情報リソース内の文で表現されている人間の行動意図を認識するため，行動属性が表す語句を取得している．行動属性は (行動主: Who / 動作: Action / 対象: What / 場所: Where / 時刻及び場面:When / 行動間の遷移ラベル次: Next，後: After / 行動間の因果関係: Because of) のように定義されている．また，越川ら [19] はソーシャルメディアとマスメディアの対比をするため，ニュース記事，Twitter の文中から事象情報を抽出し，Linked Data 化することで見える化するサービスを提案している．越川らは事象を世の中で起こっている事柄とし，事象を表現するための属性を (主題:Subject，動作:Activity，動作の目的語:Object，事象が起こる時刻及び場面:Time，事象が起こる場所:Location，事象が起こる原因:Cause，ある事象の次の事象:Next，動作の対象主:Target，主題の状態:Status，情報の発信元:Quoted source，事象を捉える立場/観点:Regard，修飾句:Modifier，事象が起こる条件:Case) と定義している．

本研究では Nguyen ら，越川らの事象属性定義を参考に，文構造を構成する属性を以下に定義する．

- Subject (主語)
- Activity (動作)
- Object (動作の対象)
- Date (日付)
- Time (時間)
- Location (場所)

Nguyen ら，越川らが定義した属性をそのまま用いると，文構造が細分化され，ユーザに推薦するニュース記事の件数が少なくなると判断し 6 属性に絞ることとした．今回定義した属性を用いて「本田圭佑が 2014 年 5 月 21 日にイタリアセリエ A でワーストイレブン

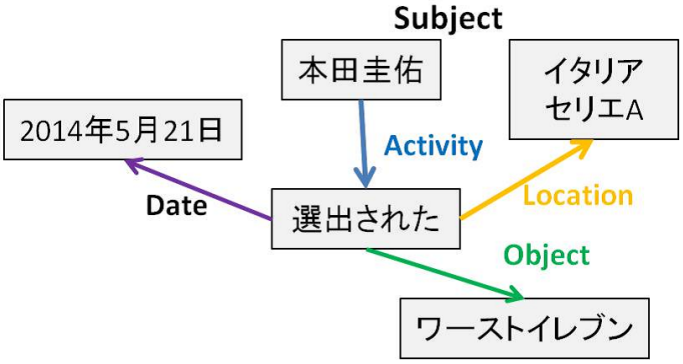


図 3.2: 文構造の例 (1)

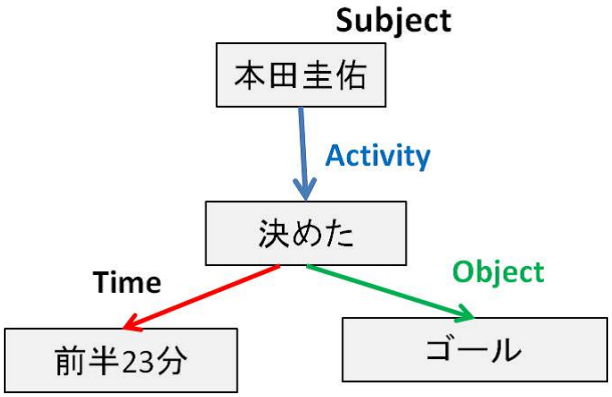


図 3.3: 文構造の例 (2)

に選出された」という文を本文に持つニュース記事を取得した場合，抽出した文構造は図 3.2 のように表現される．同様に「本田圭佑が前半 23 分にゴールを決めた」という文を本文に持つニュース記事の文構造は図 3.3 のように表現される．

3.1.2　CRF による文構造のラベリング

ニュース記事から文構造を抽出するためにで定義した文構造ラベルが付与されたニュース記事文が必要である．ニュース記事本文に文構造ラベルを自動ラベリングをするために CRF(Conditional Random Field) を利用した．本論文でのラベリング手法は同じく CRF を用いて，ニュース記事や twitter などの文から動作を表す語句を事象として抽出を行った

越川らの研究のアプローチで行う．まず，ニュース記事文章に自動ラベリングを行うために，ニュース記事文章を CRF による自動ラベリングに適した形にフォーマット変換を行う．次に学習モデルを構築する際に必要となる訓練データを作成するために，フォーマット変換を行ったニュース記事文章に対して，人手で 3.1.2 で定義したラベル付けを行う．今回は訓練データを用いた特徴モデル構築，及び CRF による自動ラベリングに CRF++[1] というツールを利用する．

本研究でニュース記事をフォーマット変換する際に行った前処理について下記に述べる．

辞書の構築

本研究では自然言語の形態素解析 · 構造解析に日本語形態素解析エンジン Mecab[2]，日本語係り受け解析器 Cabocha[3] をそれぞれ用いている．ニュース記事中に含まれる名詞句を適切に探索し，解析できるよう 2014 年 12 月 11 日時点での Wikipedia の見出し語 1509897 件を MeCab のユーザ辞書として加えた．また，地名の情報を適切に探索するために，日本の地名 7572 件を同様に MeCab のユーザ辞書として加えた．

文中に現れる括弧の処理

ニュース記事の文内には，丸括弧や読者の目を引くためにカギ括弧 (「」, 【】,[ ] 等) が頻出する．文構造が複雑になるため，抽出が難しくなる．文構造抽出精度向上のために事前処理を適用した．丸括弧は略語の注釈，引用元の情報，著者名など，ほとんどの情報が 3.1.2 で定義した文構造を構成する属性として扱う必要がないため括弧内の文字列ごと除去した．カギ括弧の処理については越川らの手法を用いた．カギ括弧に対する事前処理では文構造を単純化するために「括弧内の文字列」と「括弧外の文字列」を分割する．括弧内の文字列は対象となる文と連続する他の文として扱い，括弧外の文字列と括弧内の文字列を抜き出した箇所を “[SYSTEM:KAKKO1]” として置換する．

[1] http://crfpp.googlecode.com/svn/trunk/doc/index.html?source=navbar
[2] http://mecab.googlecode.com/
[3] https://code.google.com/p/cabocha/

照応・共参照解析

本研究では照応 · 共参照には対応していない．文脈依存性の高い日本語記事では照応 · 共参照解析が必要となる場合ある．CRF モデル作成のために訓練データ作成の際に利用した 10 月 3 日のニュース記事 13 件に含まれる 98 文を対象とし，照応 · 共参照解析が必要となる箇所を探した結果，対象記事含まれる 2479 語中 “それ · その · これ · このうち” などの指示語の出現数は 8 回であった “その”·“領収書” という連続する指示語句が記事文に含まれている場合でも，本手法では CRF によるラベリングを用いることにより “その領収書” という形で語句を抽出することができる．そして，他の記事文に含まれる “領収書” とマッチすることができる．よって照応 · 共参照解析は本論で利用しているニュース記事に限定すると考慮する必要性は薄いと考えたからである．

CRF とは

John D. Lafferty ら [20] が提案した CRF は系列ラベリング問題を解くことができ，重複する特徴をモデルに組み込むことができる識別モデルである．通常の識別モデルとは異なり，出力が出力集合の部分集合ではなく，系列となる特徴がある．形態素解析，品詞タグ付与などの系列ラベリング問題に利用されている．文中の語の出現位置，直前 · 直後に出現する単語の品詞によりラベリング対象となる単語の品詞が変わるように，系列ラベリング問題では要素のラベルが系列内の他のラベルにも依存している．そのため，系列ラベリングでは可能なラベル付が多く存在する．可能な品詞列をすべて列挙し，それぞれについて分類器を作成するような形では系列ラベリング問題は解けない．単語を要素とする系列となる文章を対象とする自然言語処理において非常に重要な問題である系列ラベリングを行うためには CRF のような特別な手法が必要である [21]．そのため本論では CRF を用いて文構造のラベリングを行う．

入力データを x，出力データを y とするとき CRF のグラフィカルモデルは図 3.4 のように表現することが出来る．出力データ y は入力データ x に対する条件付き確率である．例として「This is a thesis」という文を対象に品詞分類を CRF により行う場合，入力データ x は “This is a thesis” に対し，出力データ y は “(This →) 代名詞，(is →) 動詞，(a →) 不定冠詞，(thesis →) 名詞” のように x に対し y が正しく出力できるような特徴モデルを構築す

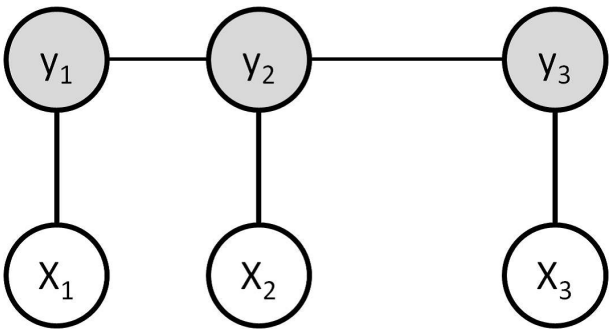


図 3.4: CRF のグラフィカルモデル

る．入力データ $x = x_1, x_2, \ldots, x_n$，出力データ $y = y_1, y_2, \ldots, y_n$ に対する条件付き確率は式 3.1 のように表すこと出来る．

$$P(y|x) = \frac{1}{Z_{x,w}} exp(w \cdot \phi(x, y)) \tag{3.1}$$

$\phi(x, y)$ は素性関数であり，すべての素性ベクトルを足しあわせたものを出力とする．w は素性に対する荷重値である．$Z_{x,w}$ は $\sum_y P(y|x)$=1 を保証する係数である．CRF では素性に対する荷重値 w を訓練データから学習し，これを用いることで入力データ x が与えられたときの出力データ y の確率 $P(y|x)$ を最大化するようにラベリングを行う．

本研究において，どの素性をどの程度用いるかについてデザインする素性テンプレートは後述する 3.1.3 で述べる．また，CRF は品詞分類や形態素解析以外の系列ラベリング問題にも広く扱われている．例えば，固有表現抽出 [22]，集団行動認識 [23] などの研究分野にも用いられている．

テストデータの作成

越川らはニュース記事のテキストを CRF の入力となるテストデータ形式にフォーマット変換するために以下の情報を用いた．

- データ元参照情報
- 括弧内文章との対応情報
- 文章情報 (文脈 ID/係り受け先文脈 ID)

- 表層形
- 品詞 ID

“データ元参照情報” はニュース記事群から処理対象とするテキストを読み込むときに文章情報と共に取得している，どの記事のどの文であるかを参照するための情報である．“文脈情報”，“表層形”，“品詞 ID” についてはテキストに対して，形態素解析及び構文解析した結果から得ている．文脈 ID と係り受け先の文脈 ID を Cabocha，品詞 ID を MeCab からそれぞれ得ている．形態素情報は表層系と品詞 ID に分けることができ，文字通り単語の表面の形である．文脈情報とは表層系の語句がどの文脈に存在し，またどの文脈にかかり受けているかの情報である．MeCab が出力する形態素の品詞は複雑な形式となるため，機械的に処理しやすいように形態素に対応する品詞に対し品詞 ID が付与される．例として，「本田圭佑が 2014 年 5 月にイタリアセリエ A でワーストイレブンに選出された」という文のニュース記事のテキストはテストデータ形式にフォーマット変換すると図 3.5 のようになる．テストデータ形式にフォーマット変換したデータの一部を CRF を用いた特徴モデル生成のための訓練用データとして利用している．

3.1.3　訓練データの作成

CRF を用いた特徴モデル生成のためにテストデータ形式にフォーマット変換を行った文に対して，属性をラベルとして人手で付与する．図 3.5 の文に対して人手でラベル付けした訓練データの例を図 3.6 に示す．B はチャンクの先頭，I は内部，O は外部をそれぞれ示す．チャンクは文の語句表層形をまとめた固まりを示す．構築したモデルを利用して，テスト用データの属性ラベル抽出を行う．出力は CRF により連続する同一の属性ラベルが付与された表層形のチャンクとする．

訓練データを用いた CRF のための特徴モデルの構築

CRF のための特徴モデルには 3.1.3 で人手でラベル付を行った文を利用して構築する．越川らは訓練データに与える素性として “表層形”，“品詞 ID”，“ラベル付けした属性” の 3 つを用いている．訓練データのどの素性を採用するかを選定する CRF 素性テンプレート

文脈ID	係り受け先ID	表層形	品詞ID
0	4	本田	43
0	4	圭	44
0	4	佑	44
0	4	が	13
1	2	2014	48
1	2	年	53
2	5	5	48
2	5	月	38
2	5	に	13
3	5	イタリア	46
3	5	セリエA	38
3	5	で	13
4	5	ワースト	38
4	5	イレブン	38
4	5	に	13
5	-1	選出	36
5	-1	さ	31
5	-1	れ	32
5	-1	た	25

図 3.5: テストデータの例

を作成している．CRF 素性テンプレートでは対象となる形態素の情報と前後 3 形態素の情報を素性と設定している．例えば，図 3.6 の “イタリア” を対象とする場合，“5，月，に，**イタリア**，セリエ A，ワースト，イレブン” までの形態素の情報を 1 組みとして，CRF に与える 1 つの素性とする．本論でも越川らと同様のテンプレートを用いて CRF のための特徴モデルを訓練データから構築し，自動ラベリングに利用する．

訓練データのラベリング精度

本研究では，インターネット上のニュースメディアである朝日新聞デジタル [4] の記事のうち 10 月 3 日に掲載された 13 件の日本語ニュース記事を取得し，訓練用データに利用した．訓練データの概要を表 3.1 に示す．10 交差検定により算出した文構造の各属性ラベルの推測精度の平均とすべてのラベルに対する推測精度を示す加重平均を表 3.2 に示す．加

[4] http://www.asahi.com/

表層形	品詞ID	属性ラベル
本田	43	B-Subject
圭	44	I-Subject
佑	44	I-Subject
が	13	O
2014	48	B-Date
年	53	I-Date
5	48	I-Date
月	38	I-Date
に	13	O
イタリア	46	B-Location
セリエA	38	I-Location
で	13	O
ワースト	38	B-Object
イレブン	38	I-Object
に	13	O
選出	36	B-Activity
さ	31	I-Activity
れ	32	I-Activity
た	25	I-Activity

図 3.6: 訓練データの例

重平均のは全ラベルにおける推測精度である．10 交差検定において，各ラベルの推測精度に各ラベルの正解データ数を掛けたものの総和を正解データ数の総和で割ったものを平均した．日本語ニュース記事のテキストに対しては本論で定義したラベル付を施すことにより，8 割以上の精度で全体のラベルを推測できることを示す．

表 3.1: 訓練データの概要

文の数	語句数	ラベル数	Subject	Activity	Object	Date	Time	Location
98	2479	1888	265	718	754	79	37	35

表 3.2: CRF による文構造ラベルの推測精度

	Subject	Activity	Object	Date	Time	Location	Weighted Average
Precision	67.80%	91.22%	87.08%	81.23%	72.93%	45.00%	85.79%
Recall	83.46%	87.20%	79.60%	90.03%	87.50%	80.00%	85.28%
F-measure	74.82%	89.16%	83.17%	85.40%	79.55%	57.60%	85.54%

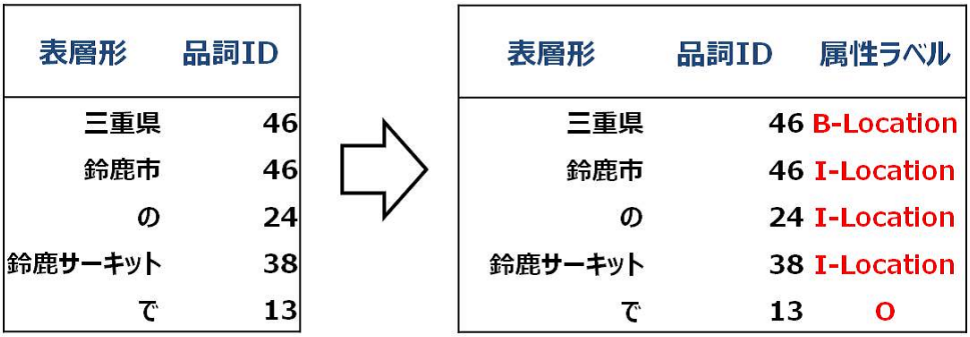


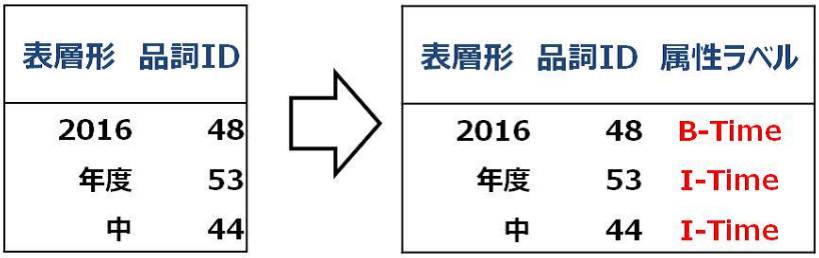
図 3.7: Location の推測精度を向上させるためのルールの例

ラベリング精度向上のためのルール適用

しかし，CRF による自動ラベル付だけでは Subject, Time, Location の推測精度が 80%を超えていない．十分な推測精度とはいえない．特に Location に関しては 60%を下回っている．このうち Time と Location に関しては品詞 ID と表層形を手がかりにヒューリスティックルールを設けることで推測精度の向上が見込めるパターンを発見した．また，CRF による自動ラベリングを行った際に，チャンクの先頭語句に誤ったラベルが付けられたケースを解決するヒューリスティックを設けた．以下に今回定義した，ラベルの推測精度を向上させるための 3 つのヒューリスティックルールの詳細を述べる．

Location 推測の精度を向上させるためのヒューリスティックルール

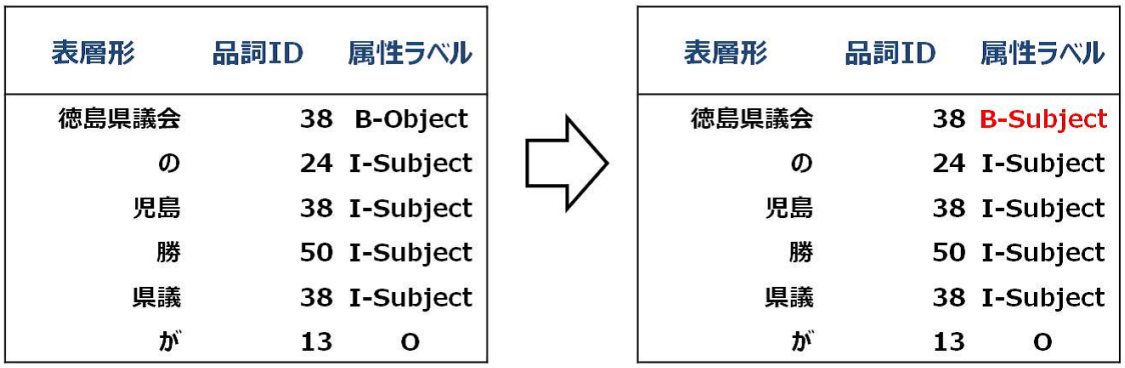
Location の推測精度を向上させるためのルールにおいて，地名を表す語句の品詞 ID に注目する．MeCab では形態素解析結果の語句に対して，品詞 ID を割り振っている．地名を表す語句に対しては「名詞, 固有名詞, 地域, 一般」という意味を持つ品詞 ID「46」を割り振られている．また地名を表すチャンクの末尾には「〜で」,「〜では」という言葉が頻出する．今回は地名を示す品詞 ID「46」の語句の直後 8 語句に



表層形	品詞ID
2016	48
年度	53
中	44

表層形	品詞ID	属性ラベル
2016	48	B-Time
年度	53	I-Time
中	44	I-Time

図 3.8: Time の推測精度を向上させるためのルールの例



表層形	品詞ID	属性ラベル
徳島県議会	38	B-Object
の	24	I-Subject
児島	38	I-Subject
勝	50	I-Subject
県議	38	I-Subject
が	13	O

表層形	品詞ID	属性ラベル
徳島県議会	38	B-Subject
の	24	I-Subject
児島	38	I-Subject
勝	50	I-Subject
県議	38	I-Subject
が	13	O

図 3.9: チャンクの先頭を正しいものに置換するルールの例

「〜で」,「〜では」という語句が存在する場合，それより前の語句までを Location とラベル付をすることとした．また，品詞 ID「46」の語句の直後 8 語句以内に「は」や「が」が存在する場合．Subject のチャンクである可能性がある．この場合は Location のラベル付はしない．Location の推測精度を向上させるためのヒューリスティックルールの例を 3.7 に示す．

Time の推測精度を向上させるためのヒューリスティックルール

Time の推測精度を向上させるためのルールとして，時相を表す語句の品詞 ID に注目する．MeCab では時相を表すチャンクの接尾語句に対し,「名詞, 接尾, 副詞可能,*」という意味を持つ品詞 ID「58」が割り振られる．例として,「2016 年度中」の “中” や「今週末」の今週 “末” に品詞 ID「58」が割り振られる．品詞 ID「58」を持つ語句の前に数値や時相を表す語句が存在する場合，Time とラベル付することとした．今回は「名詞, 数,*,* 」という意味を持つ品詞 ID「48」,「名詞, 接尾, 助数詞,*」という意

表 3.3: ヒューリスティックルールを適用した文構造ラベルの推測精度

	Subject	Activity	Object	Date	Time	Location	Weighted Average
Precision	67.80%	91.22%	87.41%	81.23%	82.46%	97.77%	86.48%
Recall	85.61%	87.20%	82.22%	90.03%	87.50%	85.71%	86.59%
F-measure	75.67%	89.16%	84.74%	85.40%	84.90%	91.34%	86.53%

味を持つ品詞 ID「53」,「名詞, 副詞可能,*,*」という意味を持つ品詞 ID「67」が割り振られた語句を数値や時相を表す語句とする．Time の推測精度を向上させるためのヒューリスティックルールの例を 3.8 に示す．

チャンクの先頭を正しいものに置換するヒューリスティックルール

CRF による自動ラベリングではチャンクの先頭語句に正しくないラベリングをしている場合がある．チャンクとして，正しい可能性が高い場合のみ，先頭のラベルを正しいと思われるラベルに置換するルールを設けた．例として，B-Subject が正しいラベルであり B-Object が付与されている場合を下記に述べる．B-Object 以下に I-Subject が 3 つ以上連続してつながっているチャンクの場合，Subject のチャンクである可能性が高いと考える．この時，B-Object を B-Subject に置換し，Subject のチャンクとして扱うこととする．チャンクの先頭を正しいものに変更するヒューリスティックルールの例を 3.9 に示す．

上記で定義した 3 つのヒューリスティックルールを CRF による自動ラベリング後に適用し，10 交差検定により算出した文構造の各属性ラベルの推測精度の平均とすべてのラベルに対する推測精度を示す加重平均を表 3.3 に示す．この結果，Time と Location のラベルの精度を大幅に向上させることができた．

3.1.4　文構造の抽出

本論文での文構造抽出は，CRF により自動ラベリングを行った文章に対して事象情報の抽出を行った越川らの研究で提案されているルーチンを利用する．文構造抽出の対象となるデータの例を図 3.10 示す．形態素解析 · 構文解析により得られた，“表層系”，“品詞 ID”，

文脈ID	係り受け先ID	表層形	品詞ID	属性ラベル
0	4	本田	43	B-Subject
0	4	圭	44	I-Subject
0	4	佑	44	I-Subject
0	4	が	13	O
1	2	2014	48	B-Date
1	2	年	53	I-Date
2	5	5	48	I-Date
2	5	月	38	I-Date
2	5	に	13	O
3	5	イタリア	46	B-Location
3	5	セリエA	38	I-Location
3	5	で	13	O
4	5	ワースト	38	B-Object
4	5	イレブン	38	I-Object
4	5	に	13	O
5	-1	選出	36	B-Activity
5	-1	さ	31	I-Activity
5	-1	れ	32	I-Activity
5	-1	た	25	I-Activity

図 3.10: ラベル付与されたテストデータの例

“文脈 ID”，“係り受け先文脈 ID”，“付与された文構造ラベル” を手がかりに文構造を抽出する．

まずはじめに，属性ラベルに付与されている先頭の文字「B」,「I」,「O」に着目し，文構造を構成する語句を抽出する．抽出する属性及び語句の対象は BIO タグの B から始まる単独もしくは B-○○，I-○○，...，I-○○と複数の連続する属性ラベルが付与された形態素とする．図 3.10 のデータを例とすると，文構造を構成する語句は図 3.4 のように抽出される．

つぎに文脈 ID と係り受け先文脈 ID を用いて，Activity の属性ラベルが付与された語句に着目し，同一文脈に存在する文構造の属性語句をまとめる．以下，手順である．

1. Activity の属性ラベルが付与された語句が出現する文節を文意の区切りとし，文脈 ID リストを作成する．(文脈 ID リスト=[5]:選出された)

表 3.4: 抽出された文構造を構成する語句の例

文構造ラベル	対応語句
Subject	本田圭佑
Date	2014 年 5 月
Location	イタリアセリエ A
Object	ワーストイレブン
Activity	選出された

2. この文脈 ID リストに係り受けしている係り受け先 ID を持つ語句の組み合わせを同一文脈に存在する文構造の語句とみなし，まとめる．
   (文脈 ID 5：本田圭佑，2014 年 5 月，イタリアセリエ A，ワーストイレブン，選出された)

3. 文構造の語句とそれぞれの属性ラベルをリンクさせることで文構造をグラフ化する．(図 3.2) またこの際に，文構造グラフにどのニュース記事，文，文脈から構築された情報化を示す情報を紐付ける．

これだけでも文構造を図 3.2 のようにグラフ化してまとめることができるが，さらに越川らはよりリッチな情報にすべく，事象間の関係情報の抽出をヒューリスティックルールにより導出している．越川らが定義した，事象間の関係情報を表す修飾句，目的語句，要因句，条件句，並列句，主語句を同定するヒューリスティックルールは全 6 件である．このうち目的語句を同定するためのヒューリスティックルールで得た関係情報のみを本研究で用いることとした．今回定義した文構造ラベルは他の 5 件のヒューリスティックルールで得られる関係情報には対応していないため，利用しなかった．下記に目的語句と同定するヒューリスティックルールを越川らが示した例文と共に記す．

目的語句になる条件を定義するヒューリスティックルール

例文：天気が回復するかわからない

→ (目的語句 Object:(Object:天気，Activity:回復する)，Activity:わからない)

Object の末尾の係り受け先が “Activity” の文脈であり，且つこの Activity の末尾の

係り受け先が他の Activity の場合，前方の Object，Activity の組を目的語句として同定する．

越川らは属性ラベルに修飾語を意味する M(Modifier) から始まる事象属性ラベルを定義していたが，本研究では BIO タグのみを用いた．実際には 3.1.3 で構築したモデルを用いて，フォーマット変換した他のニュース記事群を対象として文構造の抽出を行い，ユーザ嗜好 Linked Data，ニュース記事 Linked Data をそれぞれ構築する．

3.2　類似部分グラフ検索によるニュース記事の推薦

ニュース記事を推薦するために，Linked Data 間で一致する 1 つ以上のトリプルが連結する部分グラフを類似する部分グラフと定義する．類似部分グラフは Linked Data 間で一致するトリプルから検索する．一致トリプルの主語または目的語を含む別トリプルが両部分グラフに存在し，尚且つ該当主語または目的語が持つ述語も一致するトリプルが存在する部分グラフを探索する．次にユーザ嗜好 Linked Data の部分グラフと類似するニュース記事 Linked Data 内の部分グラフに紐づくニュース記事をユーザに推薦する．

3.2.1　類似部分グラフ検索の例

提案手法の概要図 3.1 においてユーザ嗜好 Linked Data 内の “本田圭佑 (Subject)” を主語とする部分グラフとニュース記事 Linked Data 内の “香川真司 (Subject)” を主語とする類似部分グラフの具体例を図 3.11 に記す．2 つの部分グラフはトリプル「選出された (Activity) → Property:Object → ワーストイレブン (Object)」を持つため，一致トリプルであるといえる．また，一致トリプルの主語である「選出された (Activity)」を含むトリプル「主語 → Property:Activity → 選出された (Activity)」で部分一致する．このとき，ニュース記事 Linked Data の該当トリプルの主語部分は「香川真司 (Subject)」である．以上から類似する部分グラフ「香川真司 (Subject) → Property:Activity　→　選出された (Activity) → Property:Object →ワーストイレブン (Object)」に紐づくニュース記事（この場合,「香川真司がワーストイレブンに選出された」）をユーザに推薦する．

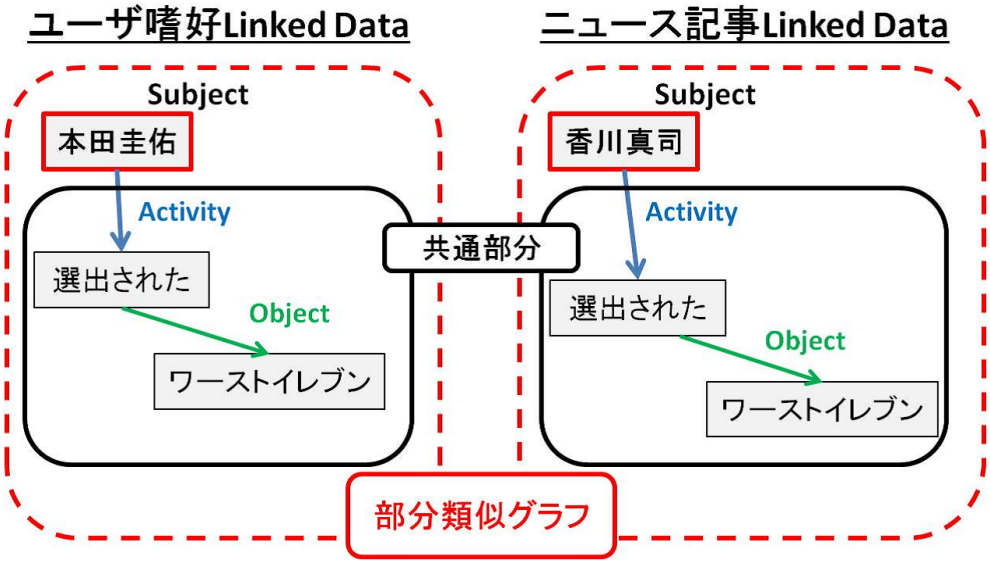


図 3.11: 類似部分グラフの例

3.2.2　類似部分グラフ検索アルゴリズム

Linked Data の類似部分グラフを取得するためのアルゴリズムを Algorithm1 に示す．ユーザの嗜好 Linked Data 内のトリプル *user_triple* の集合を *UserGraph*，ニュース記事 Linked Data 内のトリプル *news_triple* の集合を *NewsGraph* とし，それぞれを入力値とする．まず，*user_triple* と *news_triple* が 2 つの Linked Data 間で一致するトリプルであるかどうかを *SIMTRIPLE* によりチェックする．*SIMTRIPLE* は Algorithm2 で述べる．一致するトリプルであった場合，該当トリプルの主語または目的語を含むトリプル集合を *CollectGraph* により取得し，それぞれ *u_graph*，*n_graph* とする．*PatialMatch* により *u_graph*，*n_graph* 間で同一の Property を含むトリプルが存在する場合，*n_triple* を変数 x として，類似部分グラフ構築のために収集する．最後に，*n_triple* と変数 x を連結させて類似部分グラフを構築し，*Subgraph* を出力する．そして，出力した *Subgraph* に紐づくニュース記事をニュース記事 Linked Data から取得し，ユーザに推薦する．

提案手法では正しいトリプル集合を完全に取得できていなくても，部分的にトリプルが取得できていれば，類似文構造検索を行い，類似部分グラフを取得することができる．本研究では 2 つの Property で繋がれた 3 ノードで構成される 2 トリプルで類似する部分グラ

フを探索し，類似する部分グラフを用いてニュース記事を検索している．そのため，連結した 2 トリプルを持つ文構造が存在していれば類似部分グラフ検索は可能である．但し，現在筆者らが定義する Linked Data のスキーマの場合は Activity を示す語句間の意味構造が取得できない場合は類似文章構造検索を行い，類似部分グラフの取得ができない．

3.3　Entity Linking

ユーザ嗜好 Linked Data の部分グラフと類似する部分グラフをニュース記事 Linked Data から検索するためには，通常，主語 (Subject)· 述語 (Property)· 目的語 (Value) の 3 要素がすべて一致するトリプルを含む部分グラフを探索する必要がある．しかし，トリプルの主語 · 目的語の語句表層形が完全一致することに限定すると，一致するトリプル数は非常に少ないと予想される．また，探索機会損失に繋がり，ユーザに推薦されるべき記事が推薦されない問題にもつながる．

そのため，Linked Data の各ノード (Subject，Value) の語句に対して Entity Linking を行い，探索機会を増加させることを試みる．Entity Linking とは文章中に現れる Entity(語句) への参照表現を認識し，参照表現辞書の該当する意味にリンクするタスクのことである．例えば「ワーストイレブンに選出された」という文章があるときに「選出された」は「選ばれた」,「選抜される」などと同様の参照表現を持つ．語句の表層形一致よりも多くの探索機会を得ることができる．Bunnescu[24] らの研究は Entity Linking において草分け的存在である．Bunnescu らは Wikipedia の記事間のハイパーリンク構造を用いることで，固有名を同定し，曖昧性を解消する手法を提案している．また，Hoffart[25] らは Wikipedia の語の文脈類似度を計算し，Entity Linking を行い，AIDA という固有表現抽出，曖昧性解消のためのフレームワークを開発している．

本研究での Entity Linking の手法には日本語 WordNet[5] と最も基本的な手法である Jaccard 係数を利用して，文字列の類似度を算出する．Jaccard 係数は式 (3.2) により定義される．Jaccard 係数とは 2 つの文字の集合 A，B の共通要素の割合を表す．入力を 2 つの文字の集合とすると，出力の値域は 0 から 1 の間を示し，1 に近づくほど 2 つの文字列集合間の類似度が大きくなる．

[5]http://nlpwww.nict.go.jp/wn-ja/

$$Jaccard(A, B) = \frac{A \cap B}{A \cup B} \tag{3.2}$$

本論で示す「選出された」「選ばれた」「選抜される」は一文字ごとの共起関係しかないため，Jaccard 係数による Entity Linking は，特に「選出された」「選ばれた」「選抜される」等の活用や語尾変化が存在する動詞においては脆弱な手法である．そのため対象となる語句を，一旦，MeCab により形態素を基本形に戻し，活用や語尾変換を統一し，更に日本語 WordNet を参照し類似語を検索することで脆弱性を軽減している．事前実験では日本語 WordNet を利用することで新たに検索できた語句組のうち，ランダムに 100 件の語句組を抽出して筆者らが確認したところ，意味が同様であると判断できた語句組は 80 件であった．日本語 WordNet で検索しきれなかった部分については Jaccard 係数を用いて一致トリプルの検索を行う．

Entity Linking による一致トリプルを取得するためのアルゴリズムを Algorithm2 に示す．入力値をユーザ嗜好 Linked Data を構成する 1 つのトリプル *u_triple* とニュース記事 Linked Data を構成する 1 つのトリプル *n_triple* とする．この時，トリプル *u_triple* とトリプル *n_triple* の Property は同一のものを扱う．それぞれのトリプルが持つ Subject 同士 (*u_triple.subject,n_triple.subject*), Value 同士 (*u_triple.value,n_triple.value*) の文字列の類似を完全一致，日本語 WordNet を参照することによる類似語検索による一致 *WordNet*, Jaccard 係数による一致 *Jaccard* の順に判断する．また，Jaccard 係数の閾値を 0.5 とし，Subject または Value の片方でも下回るトリプルは一致していないとして類似部分グラフ検索に利用しない．また，本研究で Jaccard 係数により類似度を算出する際には，鍵括弧の表現に用いている “[SYSTEM:KAKKO1]” は 1 文字の記号に置き換えている．

Algorithm 1 類似部分グラフ検索 Search Subgraph

Input: $UserGraph, NewsGraph$
Output: $Subgraph$

function PatialSearch(u_graph, n_graph)
  for all $u\_triple \in u\_graph$ **do**
    for all $n\_triple \in n\_graph$ **do**
      if PatialMatch(u_triple, n_triple) **then**
        Push n_triple **into array** X
      end if
    end for
  end for
  return X
end function

function CollectSubgraph($news\_triple, X$)
  for all $x \in X$ **do**
    Push $news\_triple + x$
      into array $Subgraph$
  end for
  return $Subgraph$
end function

for all $user\_triple \in UserGraph$ **do**
  for all $news\_triple \in NewsGraph$ **do**
    if SimTriple($user\_triple, news\_triple$) **then**
      $u\_graph \leftarrow CollectGraph(user\_triple)$
      $n\_graph \leftarrow CollectGraph(news\_triple)$
      $X \leftarrow$ PatialSearch(u_graph,n_graph)
      Push CollectSubgraph($news\_triple, X$)
        into array $Subgraph$
    end if
  end for
end for
return $Subgraph$

Algorithm 2 一致トリプル検索 Search Triple

Input: u_triple, n_triple
Output: $Bool$

```
1: function SimWords(u_word, n_word)
2:     if u_word == n_word then
3:         return True
4:     end if
5:     if WordNet(u_word, n_word) then
6:         return True
7:     end if
8:     if Jaccard(u_word, n_word) ≥ 0.5 then
9:         return True
10:    end if
11:    return False
12: end function
13:
14: function SimTriple(u_triple, n_triple)
15:    if u_triple.property != n_triple.property then
16:        return False
17:    end if
18:    if SimWords(u_triple.subject, n_triple.subject) then
19:        if SimWords(u_triple.value, n_triple.value) then
20:            return True
21:        end if
22:    end if
23:    return False
24: end function
25:
```

第4章　評価実験

本章ではユーザが面白いと感じることができるニュース記事を推薦できることを確認することを目的とする．本研究では面白いの定義を「興味があり，意外と感じることができる」としている．また，本研究では2つのPropertyで繋がれた3語句ノードの部分グラフを利用し，類似する部分グラフを用いてのニュース記事検索を行う．ニュース記事検索結果を被験者に推薦したうえで提案手法の有効性を確かめる．文中のいずれかの箇所から3語句を拾うのではなく，類似する部分グラフの繋がりがある3語句を拾うほうが，より「関連度」の高い記事を選ぶことが出来ると考えたからである．加えて，その内の一箇所 (1語句) を変数として部分グラフ検索を行うことで，関連度を維持しながら，ユーザが面白いと感じるニュース記事を推薦できると想定した．今回の指標としては被験者の興味度の値さえ高ければ良いといえるが，参考情報として関連度（推薦されたニュース記事が興味·関心を示したニュース記事と内容が関連しているか）も評価指標に加えた．また，関連度のとても高いニュース記事が得られたからといってそれがユーザにとって本当に有益であるとは言い切れない．推薦されたニュース記事の内容がユーザにとって意外な情報が含まれることでユーザが何か新しい情報が得られることがユーザにとっての利益だと考えた．このため，評価指標に意外度も加えた．

4.1　データセット

ニュース記事Linked Data構築のためにインターネット上のニュースメディアの朝日新聞デジタルから日本語文章のニュース記事を収集した．記事収集期間は2014年10月3日から2015年1月5日まで収集した21105件のニュース記事をデータセットに利用する．ユーザ嗜好Linked Dataの構築には朝日新聞デジタルから収集したニュース記事1471件を利用した．収集期間は2015年1月6日～2015年1月13日である．表4.1，表4.2にそれぞれの

データセットの概要として記事数，Linked Data を構成するユニークなノード数を属性ラベルごとに示す．

表 4.1: ニュース記事 Linked Data のデータセット

記事数	ノード数	ラベル計	Subject	Activity	Object	Date	Time	Location
21105	42890	44869	10892	12040	17994	1761	749	1433

表 4.2: ユーザ嗜好 Linked Data のデータセット

記事数	ノード数	ラベル計	Subject	Activity	Object	Date	Time	Location
1471	4526	4617	1612	1612	1548	172	84	117

4.2　実験環境

本研究では「主語 (Subject) →述語 (Property) →目的語 (Value)」の 3 つの要素でリソースの関係情報を表現しているトリプル集合を Linked Data として扱っている．トリプル集合を Linked Data として表現する方法は様々であるが，RDF(Resource Description Framework) という形にフォーマット変換して Linked Data を表現している．Linked Data を容易に扱えるような環境が必要である．RDF 等に変換した Linked Data は RDF データストアにアップロードして，SPARQL(SPARQL Protocol and RDF Query Language) を用いて情報を検索することが一般的である．今回は最も多く利用されている RDF データストアの 1 つである Virtuoso[1] の Open Source Edition 版を RDF データストアとして選定した．Virtuoso に 4.1 で取得したデータセットをアップロードし，ユーザ嗜好 Linked Data，ニュース記事 Linked Data をそれぞれ構築した．Virtuoso は RDF フォーマット以外にも Turtle，N3 といったフォーマットで表現される Linked Data をアップロードし，利用することができる．また，SPARQL とは RDF クエリ言語の一つである．SPARQL エンドポイント（検索対象とするグラフの URI）を指定し，SPARQL クエリを入力することで Linked Data の情報を容易に検索することができる．SPARQL クエリはデータベース言語である SQL のクエリに似たような構文をしているが，検索対象は Linked Data のようなグラフデータである．本

[1]http://virtuoso.openlinksw.com/

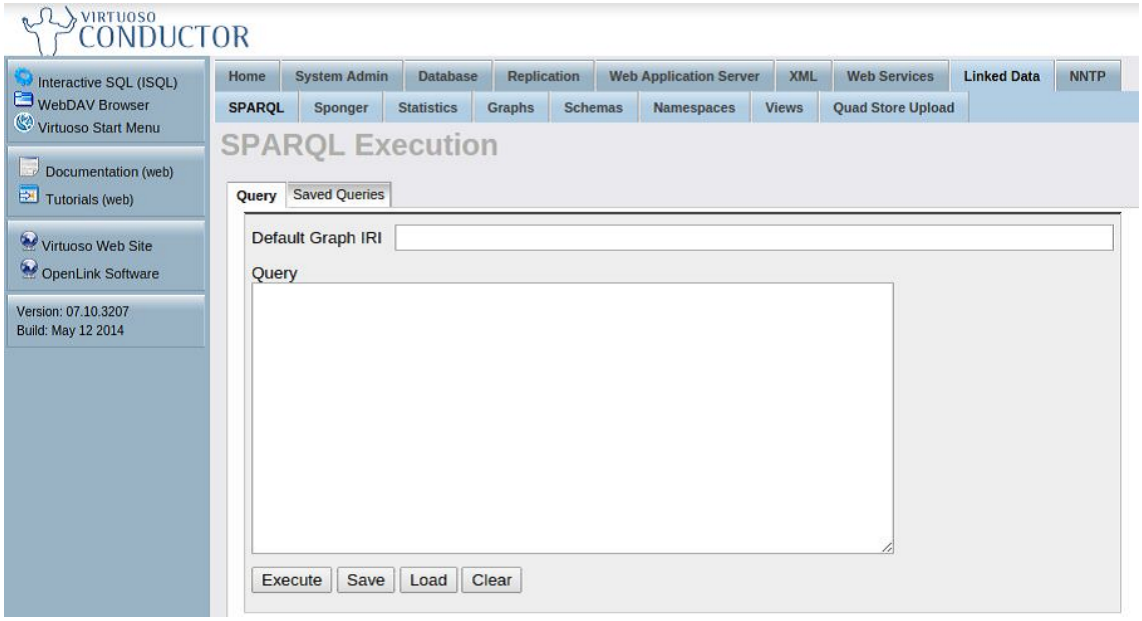


図 4.1: RDF ストア Virutoso

研究でも同様にユーザ嗜好 Linked Data，ニュース記事 Linked Data 間の類似部分グラフを SPARQL クエリを用いて検索している．Virtuoso で構築した Linked Data は，SPARQL エンドポイントを外部に公開することにより，Linked Open Data となる．本実験で利用しているデータセットはローカルでのみ扱い，Linked Open Data 化はしていない．

4.3　実験概要

4.1 のデータセットを用いてユーザ嗜好 Linked Data, ニュース記事 Linked Data 間で検索することができたユニークな類似部分グラフは 978 件であった．この類似部分グラフ 978 件はユーザ嗜好 Linked Data を構成するニュース 142 件と一致するニュース記事 Linked Data を構成するニュース記事 578 件の一致により作成された．そのため，ユーザ嗜好 Linked Data の部分グラフと類似する部分グラフに紐づく 578 件のニュース記事をユーザに推薦することが可能である．また，計算時間は 3577 秒であった．以上から多数のニュース記事がユーザに推薦することができる結果となったが，類似部分グラフの構成によっては関連度が著しく低く，ユーザにとって面白くない記事を推薦してしまう可能性がある．ユーザが興味を持つニュース記事から関連度の低いニュース記事を推薦する可能性がある．

本実験では，ユーザの興味 · 関心に関連しないニュース記事を多く推薦すると思われる類似部分グラフを制限するために条件を設ける．類似部分グラフを検索する際のユーザ嗜好 Linked Data，ニュース記事のそれぞれのトリプルが持つノードの特徴に制限を設ける条件付けをした．予備実験により筆者らがユーザの興味 · 関心に関連しないニュース記事を推薦すると判断したトリプルの特徴は以下の場合である．

- 2Linked Data 間で一致したトリプルの Property が Date の場合
- ユーザ嗜好 Linked Data，ニュース記事 Linked Data のどちらかのトリプルが持つノードが時制を示す語句のみの場合
- ユーザ嗜好 Linked Data，ニュース記事 Linked Data のどちらかのトリプルが持つ 2 ノードが 2 文字以下の場合
- ユーザ嗜好 Linked Data，ニュース記事 Linked Data のどちらかのトリプルが持つノードが “その”，“これ” などの指示語の場合
- ユーザ嗜好 Linked Data，ニュース記事 Linked Data のどちらかのトリプルが持つノードがカギ括弧を表す語句（[SYSTEM:KAKKO1]）のみの場合

今回はユーザが興味をもった記事 1 件に対して，1 件のニュース記事を推薦することとした．ユーザが興味をもったニュース記事からは複数の類似部分グラフを検索することが出来る場合があり，複数件のニュース記事が推薦出来ることがある．その場合は類似部分グラフを作成する際の一致トリプルの一致度により，記事推薦の優先度を決める．記事推薦の優先度が一番大きい 1 件をユーザに推薦することとする．以下に優先度ルールを決定する．以下ルールに基づいた上で，複数件のニュース記事が推薦できることがある場合には，複数件のニュース記事の中からランダムでニュース記事をユーザに推薦することとする．

1. ユーザ嗜好 Linked Data，ニュース記事 Linked Data のトリプルの一致が 2 ノードとも完全一致する場合
2. ユーザ嗜好 Linked Data，ニュース記事 Linked Data のトリプルの一致のうち，1 ノードが完全一致し，1 ノードが WordNet による一致の場合

3. ユーザ嗜好 Linked Data，ニュース記事 Linked Data のトリプルの一致のうち，1 ノードが完全一致し，1 ノードが Jaccard 係数による一致の場合

4. ユーザ嗜好 Linked Data，ニュース記事 Linked Data のトリプルの一致のうち，2 ノードが WordNet により一致する場合

5. ユーザ嗜好 Linked Data，ニュース記事 Linked Data のトリプルの一致のうち，1 ノードが WordNet により一致し，1 ノードが Jaccard 係数による一致の場合

6. ユーザ嗜好 Linked Data，ニュース記事 Linked Data のトリプルの一致が 2 ノードとも Jaccard 係数による一致の場合

以上の条件により，ユーザ嗜好 Linked Data と類似する部分グラフを持つニュース記事 Linked Data から推薦することができるニュース記事は 978 件中 62 件となった．今回はこの 62 件の記事を各被験者に掲示する形で提案手法の評価を行う．62 件の記事中，提案手法とベースライン手法が同じニュース記事を 3 件推薦した．

4.4　実験手順

評価実験ではまず，各被験者には掲示する 62 件の記事から被験者が興味 · 関心を示すニュース記事を最大 5 件選択させる．ユーザが選択したニュース記事の文章を用いた類似部分グラフの検索により取得することができる記事 5 件を推薦する．次に被験者が推薦されたニュース記事の内容に対して関連度，面白さの評価を行う．本研究では “面白さ” の定義を「興味があり，意外と感じることができる」とした．以上より，今回は関連度，興味度，意外度の 3 指標で被験者に評価をさせる．評価指標の詳細を以下に記す．

- 関連度 : 推薦されたニュース記事の内容と興味をもったニュース記事の内容は関連しているか
- 興味度 : 推薦されたニュース記事の内容に対して興味を持つことができるか
- 意外度 : 推薦されたニュース記事の内容はユーザにとって意外な情報であるか

各指標につき，そう思う (4 点)· ややそう思う (3 点)· ややそう思わない (2 点)· 思わない (1 点) の 4 点満点の評価で解答させた．被験者は 20 人とした．このうち 13 人は電気通信大学の学生である．また，ベースライン手法として文章内の重要度の組み合わせからニュース記事を推薦する手法を用いて比較実験を行った．特徴としてユーザが興味を示す記事文章内の名詞から構成される Bag-of-Words モデルを用いた．本提案手法と同様に各被験者が興味 · 関心を示した 5 件の各記事文章からそれぞれ名詞を抽出した．単語の重要度を出現頻度から計算する tf-idf を用いた．提案手法と同じ 20 人の被験者を対象とした．算出した tf-idf スコア上位 3 語 (名詞) を含むニュース記事をニュース記事 Linked Data と同様のデータセットから検索した．同様に各被験者に興味 · 関心を示す記事と対応する 5 件のニュース記事を推薦し，評価をした．

4.5　実験結果

掲示した 62 件の記事中，ユーザが興味 · 関心を示し選択した記事数は平均して一人あたり 4.4 件であった．提案手法とベースライン手法の評価結果の平均値を表 4.3 に示す．興味度，意外度については提案手法がベースライン手法を上回る結果となった．特に興味度はベースライン手法と比較し，差があるため語句間の意味構造を考慮してニュース記事検索を行うことで，ユーザが興味を示す記事を多く推薦することができた．今回，本論では部分グラフうち 1 ノードを変数とする類似部分グラフの検索を行い，これに紐づくニュース記事検索を行うことで，関連度を維持しながら，ユーザが面白いと感じるニュース記事を推薦できると想定した．しかし，関連度においては提案手法がベースライン手法を下回る結果となった．ベースライン手法のほうがより関連度の高いニュース記事を検索する精度が高いことがわかる．ベースライン手法は記事に含まれる tf-idf スコア上位語 (名詞) を 3 語利用している．今回ベースライン手法には本手法と同様に Wikipedia の見出し語辞書を用いている．提案手法では語句間の意味構造を考慮しているため，名詞の他に動詞も利用している．ニュース記事を検索する際の特徴として名詞のみを利用するほうがニュース記事のトピックを表現することができるといえる．このため，ベースライン手法の関連度の結果が高くなったといえる．しかし，関連度に関しては平均値が下がるが意外度に関しては提案手法が上回った．ベースライン手法は関連度の高いニュース記事を推薦すること

表 4.3: 比較実験の結果

	関連度	興味度	意外度
提案手法	3.06	3.30	2.93
ベースライン手法	3.22	3.03	2.79

ができる手法ではあるが，関連度が高すぎるが故に必ずしもユーザが面白いと感じることができる記事を推薦することができるわけではないことがわかる．以上から提案手法ではベースライン手法と比較して関連度が低いがユーザが面白いと感じることができるニュース記事を推薦できることがわかった．

4.5.1　結果分析

被験者 20 名に対して行った実験結果の具体的な分析を行う．

被験者が面白い記事と判断したニュース記事件数

本研究では面白さの定義を「興味があり，意外と感じることができる」としている．ユーザが推薦された記事に対して面白いと感じることができたニュース記事の件数がどの程度であるかは重要な指標である．ここでは興味度，意外度の 2 指標に対し，3 点以上 (そう思う (4 点)· ややそう思う (3 点)) の評価がされた記事を “面白い記事”，2 指標が 4 点以上の評価がされた記事を “特に面白い記事” とする．被験者 1 人あたり面白いと感じた記事の推薦件数を表 4.4 に示す．提案手法では被験者が面白いと感じることができる記事を 1 被験者あたり平均 4.4 件中 2.55 件推薦できていることに対し，ベースラインは 4.4 件中 2.40 件であり，さほど差は見られなかった．しかし，特に面白い記事とユーザが感じる記事を提案手法では 1 人あたり 1.15 件の割合で推薦できている．そのため，ベースライン手法と比較し，関連度の軸を意図してずらすことで特に面白い記事を推薦する手法として提案手法は優れていることを示した．

表 4.4: 被験者に対する面白い記事の平均推薦数

	面白いと感じた記事数	特に面白いと感じた記事数
提案手法	2.55	1.15
ベースライン手法	2.40	0.65

被験者が関連度が低いと判断したニュース記事の例

提案手法で関連度が低いニュース記事を推薦した部分グラフについて分析する．3 人の被験者がニュース記事について興味を示したが，推薦されたニュース記事の関連度の平均が 2 であったニュース記事の例を述べる．対象ニュース記事と紐づくニュース記事 Linked Data の部分グラフは「**平井伸治知事 (Subject)** + 明らかにした (Activity) + 会見 (Location)」であった．ユーザ嗜好 Linked Data の部分グラフ「**下村博文文部科学相 (Subject)** + 明らかにした (Activity) + 会見 (Location)」の類似部分グラフである．

上記例はユーザ嗜好 Linked Data 内の「明らかにした (Activity) - Property:Location - 会見 (Location)」というトリプルを用いて,「○○が会見で明らかにした」という語句間の意味構造に基づきユーザにニュース記事を推薦している．この語句間の意味構造は誰が，または何を会見で明らかにしたかが自明でないため，関連度が低かったといえる．また，ユーザにとって既知の情報である “会見”，“明らかにした” の組み合わせ，または語句の意味のみでは推薦した記事の関連性をユーザに示すことが難しかったといえる．しかし，この記事に対し，興味をもち，意外と感じた被験者は 3 人中 2 人いた．この結果から，一見関係のないようなニュース記事が推薦されたとしても，本手法ではユーザが興味を示し，意外度の高いニュース記事を推薦することができる．

被験者が関連度が高いと判断したニュース記事の例

提案手法で関連度が高いニュース記事を推薦した部分グラフについて分析する．6 人の被験者がニュース記事について興味を示し，且つ推薦されたニュース記事との関連度が平均 3.16 であったニュース記事の例を述べる．記事を検索したニュース記事 Linked Data の部分グラフは「サントリービール (Subject) + 売り出す (Activity) + **レモン風味の発泡酒 [SYS-**

TEM:KAKKO1](Object)」であった．ユーザ嗜好 Linked Data の部分グラフ「サントリービール (Subject) + 売り出す (Activity) + **レモン風味の [SYSTEM:KAKKO1](Object)**」の類似部分グラフである．

上記例はユーザ嗜好 Linked Data の「サントリービール (Subject) → Property:Activity → 売り出す (Activity)」というトリプルを用いて,「サントリービールが○○を売り出す」という語句間の意味構造に基づきユーザにニュース記事を推薦している．このニュース記事はユーザ嗜好 Linked Data の「売り出す (Activity) - Property:Object - レモン風味の [SYSTEM:KAKKO1](Object)」というトリプルを用いた類似部分グラフ検索により同様に推薦することができる．このため,「サントリービール (Subject) + 売り出す (Activity) + レモン風味の [SYSTEM:KAKKO1](Object)」という部分グラフを用いて検索していることと同義である．本論では 2 ノードが一致し，1 ノードが異なる類似部分グラフ集合を推薦に用いたが，より関連度の高いニュース記事をユーザに推薦しようとする場合，は 3 ノード以上が 2 つの Linked Data 間で一致する類似部分グラフを用いることで実現できるといえる．

第5章　まとめと今後の課題

本論文では，文章内の語句間の意味構造を Linked Data により表現し，ユーザの興味·関心と類似する部分グラフの検索を行うことでインターネット上のニュース記事をユーザに推薦する手法を提案した．本手法は，Bag-of-Words モデルではなく，語句と語句間の意味構造を特徴として用いている．そのため，ユーザの興味をより具体化し，ユーザが面白いと感じるニュース記事をより多く推薦することができた．しかし，ベースライン手法のほうが関連度が高くなる結果となった．これは Wikipedia の見出し語辞書を用いた上で，tf-idf スコア上位の名詞が推薦するニュース記事のトピックを強く反映させているからである．

また，よりユーザが面白いと感じるニュース記事をユーザに推薦するために今後の課題として以下を予定している．

面白い記事が得られる部分グラフのパターンの検出

本論では共通するプロパティを持つ 2 ノード，異なる 1 ノードを持つ類似部分グラフを用いてニュース記事検索を行っている．しかし，現在は部分グラフの語句ノードの組み合わせを考慮していない．語句の組み合わせが強いもの，語句そのものが強いものなどを数値化することで部分グラフの優先度を決める．ユーザが興味を持つ記事集合から tf-idf スコアが高い語句を選出し，これを実現する．この際に語句に対し，Entity Linking を行うことでより幅広い組み合わせに対応する．

また，本手法は文構造の出現位置を考慮せずにニュース記事の推薦を行っている．ユーザの興味は文の先頭から順に強いものであると考える．ニュース記事の冒頭文のみに注目した類似部分グラフ検索を行う．

さらに，一致トリプルの個数や Property の組み合わせを考慮し，よりユーザの興味を具体化し，ユーザが面白いと感じるニュース記事を推薦するシステムを構築する．

文構造ラベル推測精度向上

文構造ラベル推測精度が向上することは，文構造を高精度に抽出するためにも必要なことである．現状，Subject，Location, Time のラベル推測精度が低く，うち Location，Time ヒューリスティックルールによりこれを補っている．Location においては地名辞書を拡充することでこラベル推測精度向上を狙う．また，CRF は入力に対するラベル推測確率を出力する．そのため，尤もらしくないラベル推測確率が出力された形態素に対してのラベル付与について考慮する必要もある．しかし，CRF による自動ラベリングの精度には限界があると考える．確かなヒューリスティックルールを多く検討し，適用することでラベル推測精度向上を実現する．

Entity Linking の拡充

現在，Entity Linking には Jaccard 係数と日本語 WordNet を利用している．日本語 WordNet は単純な語の同義語には強いものの，長い語や Wikipedia の見出しのような語には対応していない．このため，Wikipedia のカテゴリリンク情報を利用した Entity Linking を提案する．Wikipedia は見出し語以外にも見出し語と対応したページ情報や，カテゴリリンク情報も提供している．この際に距離が近いほど語が類似するなどの制約を設けたい．ただし Wikipedia のカテゴリリンク情報はユニーク数でもかなりの数がある．また “日本の法人”，“〜の一覧” など，カテゴリリンク情報を用いても名寄せにならない場合がある．このため，各カテゴリの尤もらしさを計算しスコア付けした上で Entity Linking に適したカテゴリリンク情報を用いることが必要である．そして，カテゴリリンク情報自体を名寄せし，カテゴリリンク情報 Linked Data を構築することで，より簡潔に Entity Linking を行える仕組みづくりをする．これらを組み合わせて Entity Linking の拡充を狙う．

Linked Data のスキーマの見直し

現在の本研究で扱う Linked Data のスキーマは Activity を中心としたスター型である (参照 図 3.2，図 3.3)．このため，Activity が存在しない文構造，Activity ラベル推測ができなかった · 抽出できなかった文構造を用いての部分グラフ検索を行うことができない．よ

り多くの類似部分グラフを検索し，多くの候補から面白い記事を推薦できるような Linked Data のスキーマの設定を検討する．

ニュース記事推薦手法の提案

本研究ではユーザが興味 · 関心を示す記事 1 件の部分グラフを用いて 1 件のニュース記事を検索している．ここからより関連度の高いニュース記事を検索するために，ユーザが興味 · 関心を示す複数の記事から得られた複数の部分グラフが存在するニュース記事の推薦を提案する．今回扱ったニュース記事はそれほど文章が長くなく，必要性が薄いと思われるが，Web サイトや学術論文などを対象として該当ページを推薦するときに効果を発揮するのではないかと考える．

また，提案手法では 2 トリプルがリンクした 3 ノード 2 プロパティで構成される類似部分グラフを用いてニュース記事を検索しているが，検索ができるニュース記事が限られてしまい，推薦機会損失につながっている可能性があると考えられる．ここで A → B → C という部分グラフを A → B，B → C の 2 トリプルへと分解し，2 トリプルが出現するニュース記事を検索する手法を提案する．“クリミア美人検事” で例えば，記事中の「美人→検事総長」の出現する文と「検事総長→クリミア」の出現する文とを繋げれば,「美人-検事総長-クリミア」という 3 語句が出現するニュース記事を探し出すことが可能である．

ユーザの興味の学習

本研究での実験では 1 つの記事の文構造を用いて，類似部分グラフを持つ記事を 1 件推薦している．しかし，冒頭で説明した Gunosy などのキュレーションシステムではユーザの閲覧履歴から興味を学習し，適した記事を推薦している．ユーザが本当に面白いと感じる記事を推薦するためには反復的にユーザの評価を取り，ニュース記事を掲示することを繰り返す必要がある．各ユーザにとって面白いと感じる特徴 (部分グラフの構造，語句の意味など)，つまらないと感じる特徴を学習し，各ユーザに適した記事をそれぞれ推薦することを目的とする．またこの際に，万人にとって面白いと感じる特徴，つまらないと感じる特徴を収集し，コールドスタート問題に対応したい．

更に，Webアプリケーション作成を通して，語句と語句間の意味的構造を用いたより高精度なニュース記事推薦システムの実現を目指す.

謝辞

本研究を行うにあたり，ご多忙の中，有益なコメントと適切なアドバイスを下さった，大須賀昭彦教授，植野真臣教授，田原康之准教授に心より感謝いたします．川村隆浩客員准教授，清雄一助教にはご多忙の中，週１回のゼミを初めとして熱心な研究指導を賜り，貴重な勉学の機会を与えてくださったことに厚く御礼申し上げます．大須賀・田原研究室の皆様，国立情報学研究所 ・東京大学の本位田研究室皆様，早稲深澤研究室の皆様に感謝の意を表します．最後に，温かい励ましをいつも送り続けてくれた両親と家族に心から感謝いたします．

参考文献

[1] JCAST ニュース　ビジネス& メディアウォッチ , 「萌え化」クリミアの美人検事総長　日本で大騒ぎに英 BBC もびっくり , JCAST ,http://www.j-cast.com/2014/03/22199873.html?p=all, 参照 Aug.27, 2013

[2] 早川 豪, 岡部 誠, 尾内 理紀夫: “Twitter を利用したソーシャルニュース記事推薦システム”, 情報処理学会研究報告. データベース・システム研究会報告 2011-DBS-153(16), 1-4, 2011.

[3] Won-Jo Lee, Kyo-Joong Oh, Chae-Gyun Lim, and Ho-Jin Choi: User profile extraction from Twitter for personalized news recommendation, Proceedings of the 16th Advanced Communication Technology, pp.779-783, 2014.

[4] IJntema, Wouter and Goossen, Frank and Frasincar, Flavius and Hogenboom, Frederik: Ontology-based News Recommendation, Proceedings of the 2010 EDBT/ICDT Workshops, pp.16:1-16:6, 2010.

[5] 松林 優一郎, 岡崎 直観, 辻井 潤一也: “自動意味役割付与における意味役割の汎化”, 自然言語処理 = Journal of natural language processing 17(4), pp.59-89, 2010.

[6] 吉野 幸一郎， 森信介， 河原達也: “述語項の類似度に基づく情報推薦を行う音声対話システム”, 情報処理学会研究報告. SLP 音声言語情報処理 2011-SLP-87(11), pp.1-6, 2011.

[7] 今村 賢治， 東中 竜一郎， 泉 朋子: “ゼロ代名詞照応付き述語項構造解析の対話への適応”, 言語処理学会 第 20 回年次大会 発表論文集, pp.709-712, 2014.

[8] M. Capelle, F. Hogenboom, A. Hogenboom: Semantic News Recommendation Using WordNet and Bing Similarities, Proceedings of the 28th Annual ACM Symposium on Applied Computing, pp.296-302, 2013.

[9] ヨコハマ・アート・LOD　,　公益財団法人　横浜市芸術文化振興財団 ,http://yan.yafjp.org/lod, 参照 Jan.26, 2015

[10] DATA CITY Sabae , データシティ鯖江 ,http://data.city.sabae.lg.jp/, 参照 Jan.26, 2015

[11] Kira Radinsky, Sagie Davidovich, and Shaul Markovitch: Learning causality for news events prediction. Proceedings of the 15th international conference on World Wide Web, pp. 909-918, 2012.

[12] Ohsawa, Shohei and Matsuo, Yutaka: Like Prediction: Modeling Like Counts by Bridging Facebook Pages with Linked Data. Proceedings of the 22Nd International Conference on World Wide Web Companion, pp. 541-548, 2013.

[13] Elahi, Najeeb and Karlsen, Randi and Holsbø, Einar J.: Personalized Photo Recommendation By Leveraging User Modeling On Social Network. Proceedings of International Conference on Information Integration and Web-based Applications, pp. 68:68–68:71, 2013.

[14] H. Khrouf, R. Troncy: Hybrid event recommendation using linked data and user diversity, Proceedings of the 7th ACM conference on Recommender systems, pp.185-192, 2013.

[15] Roberto Mirizzi, Tommaso Di Noia, Azzurra Ragone, Vito Claudio Ostuni, Eugenio Di Sciascio: Movie Recommendations with Linked Data, IIR, volume 835 of CEUR Workshop Proceedings, page 101-112, CEUR-WS.org, 2012.

[16] Alexandre Passant: dbrec: music recommendations using DBpedia, ISWC'10 Proceedings of the 9th International Semantic Web Conference on The Semantic Web - Volume Part II, pp.209-224, 2010.

[17] Meymandpour, Rouzbeh and Davis, Joseph G.: Recommendations Using Linked Data, Proceedings of the 5th Ph.D. Workshop on Information and Knowledge, pp.75-82, 2012.

[18] T.M. Nguyen, T. Kawamura, Y. Tahara, and A. Ohsuga: Self-supervised capturing of users´ activities from weblogs, International Journal of Intelligent Information and Database Systems, Vol.6, No.1, pp.61-76, 2012.

[19] 越川 兼地，川村 隆浩，中川 博之，田原 康之，大須賀 昭彦: "CRF を用いたメディア情報の抽出と LinkedData 化 - ソーシャルメディアとマスメディアの比較事例 -"，合同エージェントワークショップ＆シンポジウム（JAWS2012）, 2012.

[20] Lafferty, John D. and McCallum, Andrew and Pereira, Fernando C. N. : Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, ICML '01 Proceedings of the Eighteenth International Conference on Machine Learning, pp.282-289, 2001.

[21] 高村大地, 奥村学 : 自然言語処理シリーズ 1 言語処理のための機械学習入門, pp.147-161, 2010.

[22] Zhu, Guangyu and Bethea, Timothy J. and Krishna, Vika: Extracting Relevant Named Entities for Automated Expense Reimbursement, Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.1004-1012, 2007

[23] Takuhiro Kaneko, Masamichi Shimosaka, Shigeyuki Odashima, Rui Fukui and Tomomasa Sato: Consistent collective activity recognition with fully connected CRFs, Proceedings of the 21st International Conference on Pattern Recognition, pp.2792 - 2795, 2012

[24] R. Bunnescu, M. Pasca: Using Encyclopedic Knowledge for Named Entity Disambiguation. Proceesings of the 11th Conference of the European Chapter of the Association for Computational Linguistics , pp.9-16, 2006.

[25] J. Hoffart, M. A. Yosef, I. Bordino, H. Fürstenau, M. Pinkal, M. Spaniol, B. Taneva, S. Thater, and G. Weikum: Robust disambiguation of named entities in text. Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 782-792, 2011.

研究業績

論文誌

1. 横尾 亮平，川村 隆浩，清雄一，田原 康之，大須賀 昭彦 : 語句間の意味的リレーションに基づくキュレーションエージェント，電子情報通信学会論文誌 ソフトウェアエージェントとその応用 特集号

査読付き国際会議

1. **Ryohei Yoko**, Takahiro Kawamura, Yuichi Sei, Yasuyuki Tahara and Akihiko Ohsuga: News Recommendation based on Semantic Relations between Events， Proceedings of the 4th Joint International Semantic Technology Conference (JIST 2014), 2014. (ポスター発表論文) **Best Poster Award**

査読付き国内シンポジウム・ワークショップ

1. 横尾 亮平，川村 隆浩，清雄一，田原 康之，大須賀 昭彦 : 語句間の意味的リレーションに基づくキュレーションエージェント ，合同エージェント＆シンポジウム 2014 (JAWS2014),2014. (ポスター発表論文)

付 録A　被験者に掲示したアンケート用紙

A.1　アンケート用紙 (1)

被験者の興味のある記事を 5 件選んでもらい，それぞれについて興味度を 4 段階で評価させ，3 点以上のものを被験者が興味のある記事だと同定し，その評価を実験結果として扱ったアンケート用紙 (1) を図 A.1 に示す.

A.2　アンケート用紙 (2)

被験者が興味をもったニュース記事 ID を入力させ，それに対応するニュース記事の内容について各指標 4 段階で評価させたアンケート用紙 (2) を図 A.2 に示す.

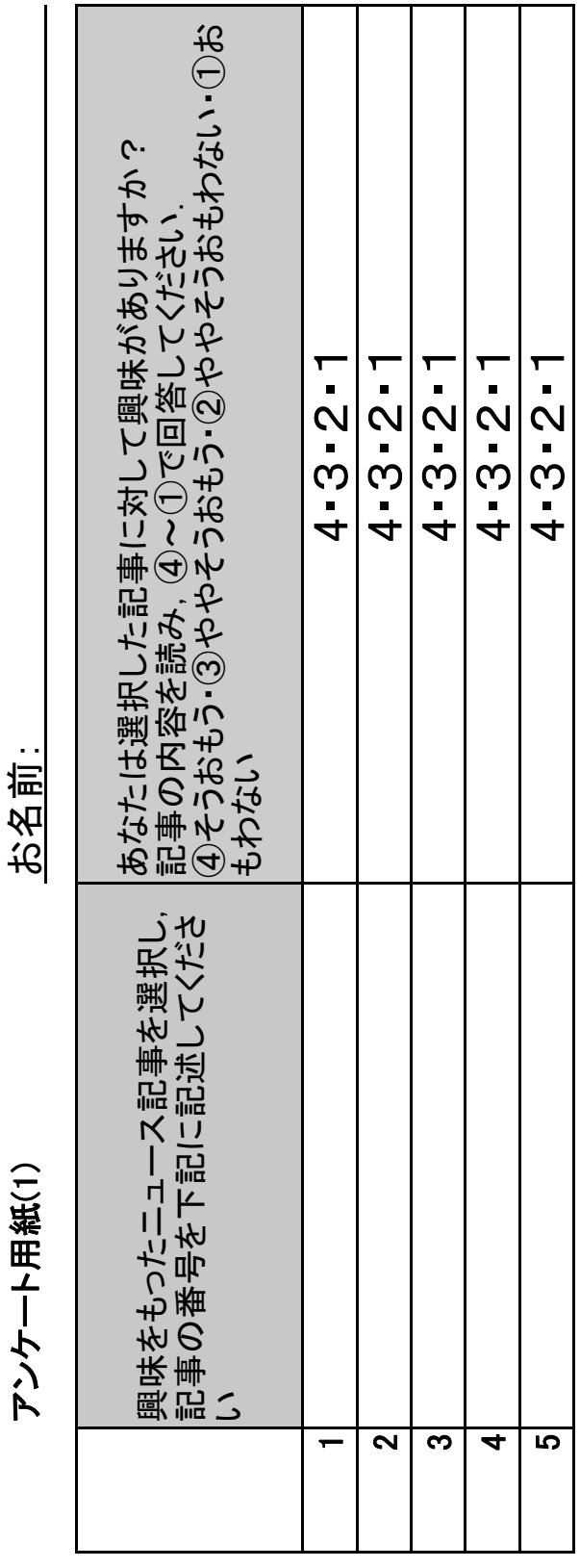


図 A.1: アンケート用紙 (1)

アンケート用紙(2)

お名前:

推薦記事群Aに対して

	興味をもったニュース記事の番号を下記に記述してください	あなたが興味を持ったニュース記事と"関連している"と思いますか？ ④そうおもう・③ややそうおもう・②ややそうおもわない・①おもわない	推薦されたニュース記事の内容に対して"興味"を持てたと思いますか？ 記事の内容を読み，④～①で回答してください． ④そうおもう・③ややそうおもう・②ややそうおもわない・①おもわない	推薦されたニュース記事の内容は，あなたにとって，"意外な情報"であったと思えますか？ 記事の内容を読み，④～①で回答してください． ④そうおもう・③ややそうおもう・②ややそうおもわない・①おもわない
1		4・3・2・1	4・3・2・1	4・3・2・1
2		4・3・2・1	4・3・2・1	4・3・2・1
3		4・3・2・1	4・3・2・1	4・3・2・1
4		4・3・2・1	4・3・2・1	4・3・2・1
5		4・3・2・1	4・3・2・1	4・3・2・1

推薦記事群Bに対して

	興味をもったニュース記事の番号を下記に記述してください	あなたが興味を持ったニュース記事と"関連している"と思いますか？ ④そうおもう・③ややそうおもう・②ややそうおもわない・①おもわない	推薦されたニュース記事の内容に対して"興味"を持てたと思いますか？ 記事の内容を読み，④～①で回答してください． ④そうおもう・③ややそうおもう・②ややそうおもわない・①おもわない	推薦されたニュース記事の内容は，あなたにとって，"意外な情報"であったと思えますか？ 記事の内容を読み，④～①で回答してください． ④そうおもう・③ややそうおもう・②ややそうおもわない・①おもわない
1		4・3・2・1	4・3・2・1	4・3・2・1
2		4・3・2・1	4・3・2・1	4・3・2・1
3		4・3・2・1	4・3・2・1	4・3・2・1
4		4・3・2・1	4・3・2・1	4・3・2・1
5		4・3・2・1	4・3・2・1	4・3・2・1

図 A.2: アンケート用紙 (2)