博士論文

医学文書における意味を考慮した単語重み付け手法の開発

松尾 亮輔

主指導教員　Ho Tu Bao

北陸先端科学技術大学院大学

知識科学研究科

平成 29 年 6 月

Abstract

Term weighting where a term is given a numerical weight regarding its importance, is fundamental to analyze text data. As term weighting can transform documents into the computable forms in a vector space, it enables to execute various document analysis such as text classification, clustering, information retrieval and so on. The traditional measure used in term weighting is TFIDF, derived from the term frequency and the inverse document frequency. TFIDF is simple and effective, and it forms a popular base for advanced algorithms in spite of its age. However, the term importance captured by term weighting methods using TFIDF and its variants does not relate to the term meanings but only to the frequencies. These methods are not suitable for applications that require considering the term meanings.

In medical domain, therefore, semantic term weighting (STW) methods have been developed aiming at assigning weights to document terms based on their meanings by exploiting ontologies and class information of terms. However, these methods are developed for a certain task in medicine. There is no framework of STW that can correspond to any task. Moreover, there is no framework to put effectively term weighting into practice. In order to exploit the computable forms of documents by term weighting for secondary use, we need to consider the nature of documents, the target of analysis and the adequate terms' meanings. To this end, we apply the idea of frame semantics to term weighting. The frame semantics is a research program in empirical semantics that emphasizes the continuities between language and experience. It assigns term meanings based on the frame that characterize a situation. The benefit of the exploitation of the frame semantics is to keep an assumption that the determination of term importance is not unique but diverse depending on the nature of documents and the target of analysis. Moreover, it considers the encyclopedic semantics of terms that is adequate to put semantic term weighting into practice.

The objective of this thesis is to develop STW methods for medical document analysis considering the idea of the frame semantics. We especially focus on two important targets: information retrieval on medical documents and prediction of patients' conditions on EMRs such as mortality prediction. To this end, we propose two frameworks: a framework of STW using a proposed procedure to apply the idea of the frame semantics to term weighting and a common framework of STW based on a proposed medical knowledge representation. The key idea is to hierarchically divide the terms into categories by exploiting ontologies and class information of terms based on medical knowledge and machine learning techniques. The terms in each category are reasonably considered to have the same medical importance (except the case that a term's weight is the continuous value) regarding a certain aspect of terms' meanings. The categories containing terms are exploited to represent various aspects of terms' meanings on computer. Based on the proposed frameworks, we developed STW methods for the two targets. As the proposed STW methods were verified by the experimental evaluations, we attained the objectives by using the proposed frameworks.

As term weighting transforms documents into computable forms, the proposed STW method considering the idea of the frame semantics can apply to various applications as secondary use when keeping various aspects of terms' meanings in medicine.

Key words: Semantic term weighting, Medical documents, Ontology, Class information, Frame semantics

目 次

図 目 次

表 目 次

第1章　緒論

1.1　研究背景

1.1.1　単語重み付け

現在、世の中にはテキストデータが膨大に存在し、そのデータ量も日々増え続けているため、テキストデータの分析は様々な分野において重要である。テキストマイニングとは、溢れるテキストデータの問題をデータマイニングや機械学習、自然言語処理、情報検索、知識マネジメントの技術により解決するための手法である [7]。テキストデータを用いて様々な文書分析を行うには、通常、文書を計算可能な形式に変換する必要がある。

単語の重み付けは、単語の重要度という観点から単語に対して重みを与えることで単語同士を区別する手法 [41] である。この手法は、文書をベクトル空間上で計算可能な形式に変換できるため、文書分類やクラスタリング、センチメントアナリシス、情報検索といった様々な文書分析の基盤である。本論文において重み付けの対象となる単語（term）とは、ある特定の領域で使われている領域の概念が具現化された名称であり、また、領域概念を特定するために用いられる科学的コミュニケーション手段のことである [18, 43, 37]。生物医学領域における単語は、遺伝子やタンパク質、遺伝子産物、生体、薬、化合物などの名称である [18]。本論文で扱う単語はある 1 語だけでなく、ある語の集合で構成されている用語も含んでいる。例えば、医学単語において、悪性新生物を意味する malignant neoplasm は、malignant と neoplasm の 2 語で構成されている単語である。本論文ではこのような単語が重み付けの対象となる。

伝統的な重み付け手法は、Term Frequency and Inverse Document Frequency (TFIDF)[41]

であり、単語頻度と逆文書頻度の 2 つの統計的要素により求められる。

$$tf \cdot idf\left(t_i, d\right) = tf\left(t_i, d\right) \times idf\left(t_i, d\right) = \frac{n_i}{\sum_k n_k} \times \frac{|D|}{|\{d : t_i \in d\}|}$$

文書 d 内の単語 t_i の TFIDF の重みは $tf \cdot idf\left(t_i, d\right)$ となり、$tf\left(t_i, d\right)$ と $idf\left(t_i, d\right)$ の組み合わせにより求められる。ここで n_i は文書 d における単語 t_i の出現回数で、$\sum_k n_k$ は文書 d 内のすべての単語の出現回数の合計である。$|D|$ は総文書数で、$|\{d : t_i \in d\}|$ は単語 t_i が含まれている文書数である。TFIDF の意味合いは、 1 文書内で多く出現し、かつ、多くの文書で出現しない単語を重要語とし、高い重みを付与する。TFIDF は新しい手法ではないが、その有効性かつ扱いやすさから、先進アルゴリズムの基盤としてよく用いられている [33]。しかしながら、TFIDF は単語の頻度情報による重み付けであるため、単語がある領域においてどのような意味を備えているかという単語の意味が考慮されていない。文書分析によっては単語の意味を考慮して重み付けをする必要があるため、そのような目的に対しては頻度情報を考慮するだけでは不十分である。

1.1.2 医学文書分析とその課題

医学分野における文書は以下の図 1.1 のように医学論文と電子カルテに分けられる。医学論文は、通常、疾病や治療に関することが書かれており、原著や症例報告、総説といったタイプの文書がある。電子カルテは、患者についての診断や治療に関することが書かれ、退院時要約や看護記録といった文書タイプが存在する。

このような医学文書を分析する意義に関して、生物医学研究の達成すべき目標は、知識を発見し、発見した知識を診断や予防、治療に活用することとされる [6]。生物医学テキストマイニングにおける目標は、必要な情報をより効果的に見分けること、利用可能な膨大な情報によって隠れた関係を明らかにすること、そして文献とデータベースのフリーテキスト内に存在する膨大な生物医学の知識に対して、アルゴリズムや統計、データマネジメントの手法を適用することによって、情報過多による負担を研究者からコンピュータへ

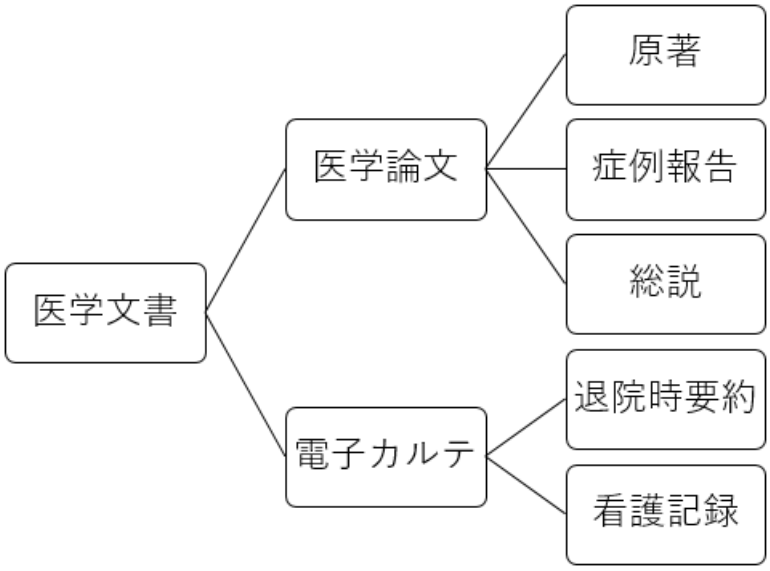


図 1.1: 医学文書の種類

移すことである [6]。具体的に生物医学テキストマイニングは、薬や遺伝子の名前といった特定の種類の名前を捉える固有表現抽出や、ある特徴が文書または文書の一部にあるかどうかを自動的に判定する文書分類、同義語（類義語）及び省略語の抽出、遺伝子やたんぱく質、薬といったタイプのエンティティ間の関係抽出、明示的な関係による推論からこれまで認識されていない関係を明らかにする仮説生成といったタスクに分けられる [6]。

医学論文を用いた分析では、利用可能性及びコストの理由から、生命科学に関する文献のオンラインデータベースである MEDLINE 内の論文データを用いたテキストマイニングの研究が多くなされている [37]。公開されている膨大な論文データの活用例として、情報検索が挙げられる。

一方で、近年の電子カルテの普及により、医学における多くの重要な問題を解決するための手法開発の可能性が開かれていることから [3, 52, 35]、電子カルテの分析は新しく重要な課題である。電子カルテを用いた分析により、看護教育支援 [71, 59] や診断支援 [65, 66, 51, 38]、サーベイランス [24, 32, 26, 55, 49] といった様々な応用がなされている。診断支援研究では、患者状態の予測のための手法開発が例として挙げられる。

図 1.2 は電子カルテを用いたコンテストである I2B2 と NTCIR の実施状況である。英語の電子カルテを用いた公開タスクの実施は 2006 年から開始され、日本語の電子カルテの

図 1.2: 電子カルテを用いた公開タスクの実施状況

場合は 2012 年から開始されている。したがって、電子カルテの活用は新しい課題であることがいえる。

医学文書を対象とした単語の重み付け手法に関する研究では、単語の頻度情報のみを考慮した TFIDF を用いた重み付けだけでなく、TFIDF を拡張して単語の意味を考慮した重み付け手法がこれまで多く開発されている。しかしながら、それらは情報検索や分類、センチメント分析といったある特定のタスクに対する手法であり、どのタスクにも対応が可能な意味を考慮した単語の重み付けのフレームワークはこれまでみられない。文書分析によって意味の捉え方が変化すると考えられるため、適切な医学知識を組み込みながら、様々な意味の視点を計算機でどのように表現するかが理論的な課題であると考えられる。

ビジネスにおける実務的な課題は、意味を考慮した単語の重み付けという基礎的な研究を効果的に実践へ結びつけることであると考えられる。しかしながら、そのようなフレームワークはこれまで存在していない。単語の意味的側面を考慮して変換した文書を、様々な分析のために二次利用するには、どのような文書を用いているのか、どのような分析に適用するのか、どのような意味を捉えるのかを考慮する必要があると考えられる。

1.2 研究の着眼点と目的

計算機に自然言語の意味を理解させることは、単語の重み付けという様々な文書分析の基盤となる手法において重要である。本論文では、単語の意味に基づいて計算機に単語の重要度を捉えさせる際に、まず人の認知プロセスに着目し、人の場合は単語の意味をどう捉え、単語の重要度を決定するかを考える。そうすることで、人が想起する状況によって単語の意味は様々な視点から捉えられることがみえてくる。これはフレーム意味論の考え方で説明できると考えられる。

フレーム意味論とは、言語と経験の関連に重点を置く経験的意味論であり、フレームの概念を用いて状況を特徴づけるあるフレームを考えることで、そのフレームがある単語の意味を割り当てるとする [8] 。フレーム意味論は人間の認知プロセスに着目している。単語の意味を概念的に考えると、主体がどのように単語の意味を捉えているか [20] が認知言語学において重要であり、その捉え方にあたるフレームを用いることで語の意味は異なる捉え方を可能にする点がフレーム意味論の特徴である。また、フレームとは背景知識であり、それは多くの場合、人々の日常生活を通じて形成された経験的知識 [70, 64] 、またはそれと似た意味である人間が様々な経験を通して身に付けた百科事典的な知識 [69] のことである。したがって、語の意味を理解するのに百科事典的知識を参照していることもフレーム意味論の特徴である。 語は百科事典的意味と、それとは対称的な基本的意味を有している [69] 。ここで百科事典的意味とは、「ある語が指し示す対象（の典型的なもの、代表的なもの）がもつもろもろの性質・特徴、さらには、その対象と関連をもつ（たとえば、その対象から連想される）様々な事柄」[68] であり、基本的意味とは、「語が指し示す対象のすべてに該当する意味であり、かつ、類義語との弁別的特徴を含むもの」と定義されている [69] 。野球の四番打者の例 [69] を挙げると、四番打者の百科事典的意味は「そのチームの最も優れた打者」や「長打力を有する」となり、基本的意味は「打順が 4 番である打者」となる。このように、基本的意味は語が指し示す対象のすべてに該当する意味であることから抽象的である。一方で、百科事典的意味は性質や特徴という具体的な意味と

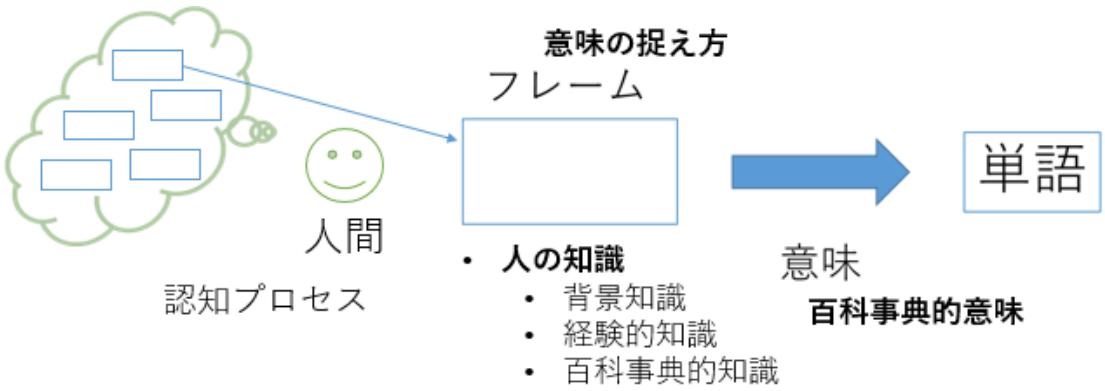


図 1.3: フレーム意味論の特徴のまとめ

なることから、意味を考慮した重み付け手法を実践に応用するのに適しているのは百科事典的意味であると考えられる。図 1.3 は、フレーム意味論の特徴をまとめたものである。

本論文では、捉え方によって異なる意味を考慮し、かつ、百科事典的意味の観点から語の意味を理解するフレーム意味論の考え方を単語の重み付けという計算手法に適用する。ある単語が重要か重要でないかは、フレーム意味論の考え方と同様に人間の認知プロセスを考慮すると、同じ語でも主体の視点の違いによって重要度の程度は変化し得る。

例えば、図 1.4 に示すように、医学単語である癌を例に挙げると、癌の基本的意味は三省堂 Web Dictionary によると悪性の腫瘍 であり、百科事典的意味は、例えば、死因ランキング 2 位に該当する危険な疾病である、という患者状態の重症度という視点や医療費の程度といった様々な視点による意味が考えられる。

コンピュータがこれらの視点から単語の意味を捉えるには百科事典的知識を組み込むこと、あるいは百科事典的意味を機械学習により捉えることが求められるが、単語の重要度の決定の仕方は一意に決まらず多様であることが人間の認知プロセスの観点から単語の重み付けを照射することでみえてくる。

本論文では、単語の意味を考慮して重みを与える際に、状況を特徴づけるフレームが単語の意味を割り当てるとするフレーム意味論の考え方を用いて、単語の重要度の決定の仕方は多様であり、想起する状況によって変化するという主な前提を保持し、ある状況ごとで百科事典的意味という具体的な応用に適した意味の観点から語の意味を計算機に理解させる。そこで、本論文の目的は、フレーム意味論の考え方を用いて、設定する文書分析

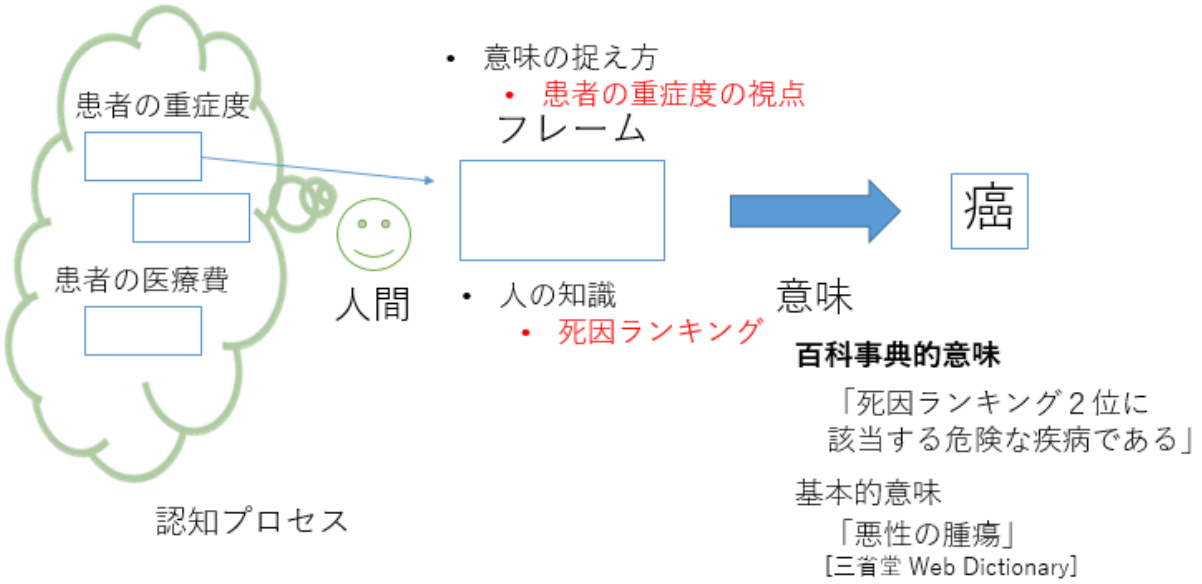


図 1.4: 医学単語を用いた百科事典的意味の例

のある目的に適した意味を考慮した医学単語の重み付け手法を開発することである。本論文では、文書分析の目的を情報検索と予測とし、特に医学における重要な課題でかつ研究が盛んな医学文書を用いた情報検索と電子カルテを用いた患者状態の予測と設定する。図 1.5 は本論文の目的と、具体的な 2 つの文書分析の目的を示している。

本論文の目的を達成するため、2 つのフレームワークを提案する。まず、フレーム意味論の考え方を単語の重み付け手法へ適用するための手順を示し、その提案手順に基づいた重み付けフレームワークを提案する。このフレームワークにより、フレーム意味論の考え方を取り入れながら、オントロジーと単語のクラス（カテゴリー）情報を活用して、意味を考慮した単語の重み付け手法を開発する。次に、医学知識表現に基づいた、医学文書分析に共通する意味を考慮した単語の重み付けフレームワークを提案する。医学知識表現は、医学文書内の単語を階層で区分けし、その区分けしたカテゴリーに対して重みを与え、そのカテゴリー内に属する単語は連続値で重みを付与できる場合を除いて、全てにそのカテゴリーの重みを同じように付与する。意味の視点は、それらのカテゴリー内の単語集合により計算機で表現する。本論文で対象とする情報検索と予測の 2 つの文書分析について、医学文書の情報検索では、単語の共起の偏りという観点から医学単語の特定性という意味を機械学習による単語のトピック情報を活用しながら捉える。電子カルテを用いた

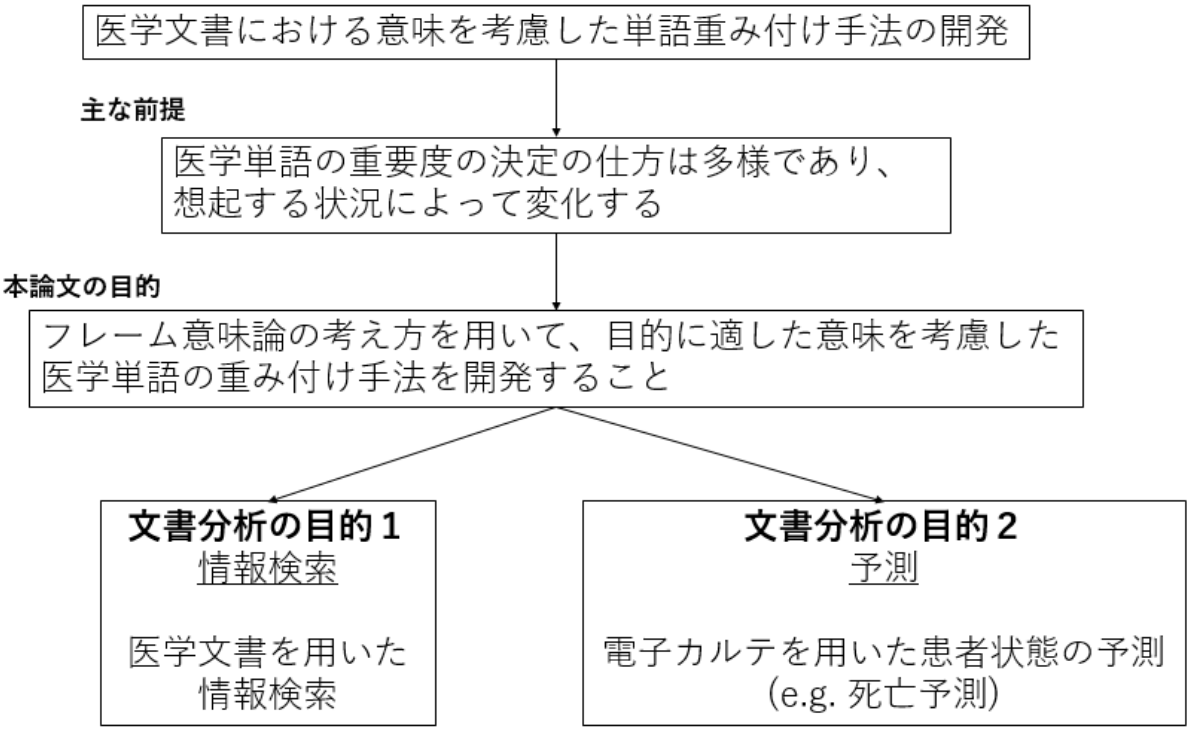


図 1.5: 本論文の目的と 2 つの文書分析の目的

患者状態の予測では、死因ランキングという百科事典的知識を活用することで、患者状態の重症度の視点から意味を捉える。これらの意味を百科事典的知識や機械学習により捉えた医学知識表現を組み込んだ重み付け手法は、医学論文の情報検索のデータセットである TREC 2014 Clinical Decision Support Track [42] と、電子カルテデータが含まれている MIMIC II データベース [39] を用いた実験による評価がなされ、提案手法の有効性が示された。

1.3 本論文の貢献

本論文の貢献は以下の 5 つである。

1. 単語の重み付けという計算手法にフレーム意味論の考え方を適用することで、単語の意味は様々な視点で捉えられることを前提として保持し、具体的な応用に適した百科事典的意味の観点から計算機に単語の意味を理解させることを可能にした。それによって、実践を重視した目的志向の意味を考慮した単語の重み付け手法の開発

が可能となり、実務的な課題である単語の重み付けという基礎的手法をどのように効果的に実践に結びつけるかに対する解決策を示した。

2. 医学知識（特にランキング）や機械学習を用いながらオントロジーや単語のクラス情報を活用して、医学文書内の単語を階層的に区分けをし、それぞれのカテゴリー内の単語集合を意味の視点として計算機に表現させた。連続値を付与できる場合を除いて、カテゴリーに対して重みを付与することで、カテゴリー内の単語に同じ重みを付与させるという文書分析に依存しない共通する重み付けフレームワークを提案した。単語の重み付けにおいて、計算機に人が想起する単語の様々な意味の視点を理解させる新たな医学知識表現により、様々な意味の視点を計算機でどのように表現するかという理論的な課題の解決策を示した。

3. 医学単語の特定性を考慮した単語の重み付け手法を開発した。情報検索の実験でその有効性が示されたことから、提案手法は医学論文を用いた情報検索の目的に適した意味を考慮した単語の重み付け手法といえる。理論的貢献は、機械学習手法により得られる単語のトピック情報を活用して、共起の偏りの視点から単語の重みを連続値で捉え、特徴的な単語と一般的な単語という共起の偏り度合いを捉えた点であると考えられる。このアイデアに基づいた単語の重み付けにおける医学知識表現の提案により、様々な意味の視点を計算機で表現するという理論的な課題の解決策を示した。

4. 患者の重症度を考慮した単語の重み付け手法を開発した。死亡予測の実験でその有効性が示されたことから、提案手法は電子カルテを用いた患者状態の予測の目的に適した意味を考慮した単語の重み付け手法といえる。理論的貢献は、百科事典的知識である死因ランキングのようなランキングを活用することで、単語集合で構成されるカテゴリーごとに対応するランク情報を活用して、カテゴリー間の優劣を求めている点であると考えられる。このアイデアに基づいた単語の重み付けにおける医学知識表現の提案により、様々な意味の視点を計算機で表現するという理論的な課

題の解決策を示した。

5. 単語の重み付けは、文書を計算可能な形式に変換するという文書分析において基盤となる手法である。したがって、提案するフレームワークを用いて開発された意味を考慮した単語の重み付け手法は、様々な視点からの単語の意味を保持した上で、データの二次利用として様々な医学文書分析への適用を可能にすると考えられる。

1.4 本論文の構成

本論文では、第2章で意味を考慮した単語の重み付け手法の関連研究と、設定した文書分析ごとの課題を述べる。第3章では、フレーム意味論の考え方を単語の重み付けへ適用する手順とその提案手順に基づくフレームワーク、文書分析に共通する重み付けフレームワーク、提案手法の本論文内の位置づけを述べる。第4章と第5章では、情報検索と予測の文書分析の目的に適した提案手法を述べ、その有効性を実験による評価を示しながら述べる。最後に第6章で結論として本論文のまとめと、提案手法の特徴と課題、効果を述べる。

第2章　関連研究

意味を考慮した単語の重み付け手法は、オントロジーと単語のクラス（カテゴリー）情報の活用により様々な手法がこれまで開発されている。

オントロジーは概念化の明示的な規約と定義され、単語間の概念関係を階層構造でシステマティックに表現する [9]。また、オントロジーは「人間が対象世界をどのように見ているかという根元的な問題意識をもって物事をその成り立ちから解きあかし、それをコンピュータと人間が理解を共有できるように書き記したもの」と定義され、基礎となる概念は対象とする世界に存在する概念とそれらの間に成立する関係に関する記述を意味する概念化であるとしている [67]。例えば、図 2.1[15] から、医学単語では、Cardiovascular Disease is a Vascular Disease や Cardiovascular Disease is a Heart Disease のように、心臓血管疾患は血管疾患の一種であり、心臓疾患の一種でもあるといった単語間の概念を is_a のような形でオントロジーは表現できる。

領域によらないオントロジーとしては WordNet[47] があり、医学オントロジーでは Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT)[12] や Medical Subject Headings (MeSH)[30]、Unified Medical Language System (UMLS)[4] が国外で開発されている。UMLS は生物医学概念のリポジトリであるメタシソーラスや 135 の概念カテゴリーを含むセマンティックネットワーク、語彙資源で構成されている [4]。国内では、画像診断所見オントロジー [61] や疾患や解剖学のオントロジー [62, 63] の開発がされている。

単語のクラス（カテゴリー）情報は、単語がある基準により区分けされた場合のそのクラスまたはカテゴリーの情報のことである。例えば、図 2.2 のように単語を領域ごとに分けることができる。医学単語の場合は医学のカテゴリーに分けられ、医学単語はさらに疾病や症状、薬といったカテゴリーに分けることができる。疾病に関する単語はさらに、ある

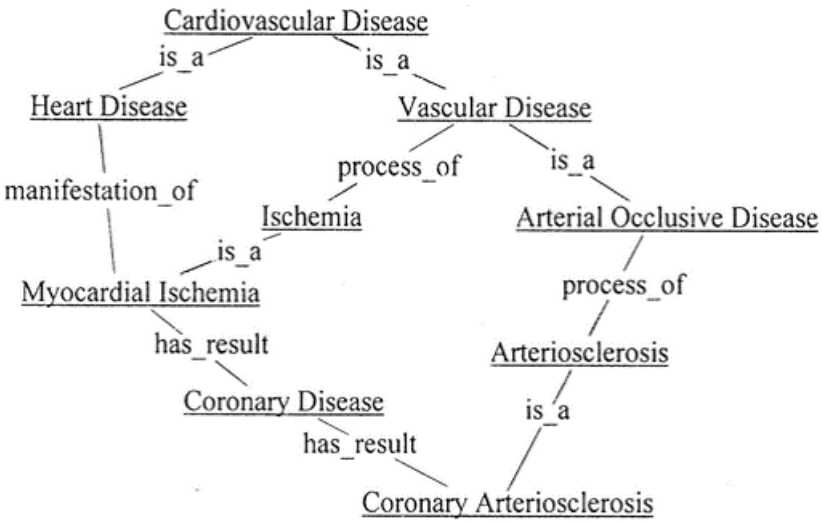


図 2.1: オントロジーの例

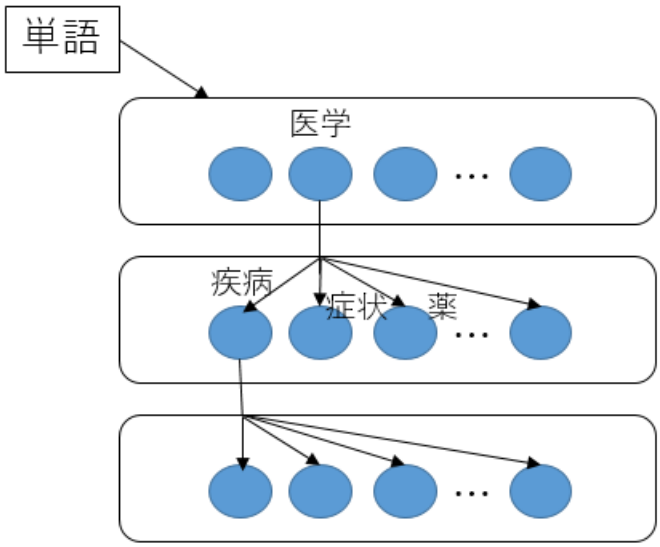


図 2.2: 単語のクラス情報の例

基準により区別ができると考えられる。このように、ある単語がいずれかのクラスに属するのがクラス情報の基本的な考えである。例えば、International Statistical Classification of Diseases and Related Health Problems (ICD) は、診断用の医学単語の英数字コードを備えており、疾病や症状、初見に関する医学単語を分類している [50]。

オントロジーの活用により TFIDF を改良して意味を考慮した単語の重み付け手法が多く開発されている。オントロジーを活用することで、単語の出現頻度だけでなく、概念間の距離や概念がオントロジーに存在するかどうか、概念が文書に発生する確率を求めることで単語の概念に基づいた重み付けが提案されている [46, 45]。

また、WordNet オントロジーから得られる語義、同義語、上位語、下位語の 4 つの概念情報を活用して、単語の一般性及び特定性を考慮した重み付けが提案されている [54, 40]。語義数、同義語数、下位語数が多く上位語数が少ない場合に一般的な単語とみなし低い重みを与え、その反対の場合（語義、同義語、下位語数が少なく上位語数が多い）に特徴的な単語とみなし高い重みを与えている。このような重み付けによる重みを Inverse Document Frequency (IDF) のかわりに Term Frequency (TF) と組み合わせている。

概念の類似度に基づいて単語の意味的関係を考慮して重み付けをする手法も提案されている [56, 57, 48, 14]。この手法はある単語と同じベクトル内にある他の単語との意味的な類似度が高い場合にその単語の頻度をもとに得られる重みを上昇させている。

クラス情報を活用した単語の重み付け手法では、文書のポジティブクラスとネガティブクラスという 2 つのクラスに基づいて単語の統計的側面から重み付けを行い、テキスト分類やセンチメント分析に適用している [19, 17, 23]。

オントロジーとクラス情報を組み合わせた単語の重み付け手法では、WordNet のオントロジーとそのオントロジー内のカテゴリーを活用し、WordNet により認識された単語と WordNet 内のカテゴリーの意味的な類似度を用いた重み付けをしている [21]。また、医学カテゴリーが WordNet 内の 1 つのカテゴリーであることから、このハイブリッド手法は医学文書への適用が可能となっている。

医学オントロジーの活用により、医学文書を用いた意味を考慮した単語の重み付け手法では、MeSH により単語間の意味的関係に着目した方法 [56, 57] が提案されている。医学オントロジーとクラス情報を活用した重み付け手法では、医学オントロジーである UMLS を活用することで、クエリー内の単語を UMLS concept や UMLS synonym のようにカテゴリー分けを行い、そのカテゴリー情報をもとに IDF の重みを上昇させている [53]。また、UMLS から得られる単語の意味タイプに関して、TREC 2004 Genomics Ad Hoc Retrieval Task における主な意味タイプを Amino Acid、Peptide、Protein と設定し、その主な意味タイプを持つ単語の重みを TFIDF の重みから増加させている研究もある [58]。この研究では、UMLS で MeSH に基づいた主な意味タイプの単語の同義語の認識によりクエ

リー拡張を行い、その同義語の重みも上昇させている。

表 2.1 は意味を考慮した単語の重み付けの先行研究をオントロジーや単語のクラス情報の利用、医学論文への適用、文書分析でまとめたものである。医学論文への適用に記号の三角を示している理由は、該当する先行研究が用いている WordNet のオントロジーのカテゴリーの 1 つが医学カテゴリーであるためである。

これまで述べた医学文書に適用した意味を考慮した単語の重み付け手法は、医学論文を用いている。一方、電子カルテといった医学文書に対しては TFIDF による単語の重み付けが適用されているが [10, 28]、単語の意味を考慮して TFIDF を拡張する手法はこれまでみられない。TFIDF を用いずに固有表現や意味的な叙述関係といった意味的特徴を考慮して電子カルテに診断コードを割り当てる手法は開発されている [16]。

2.1 文書分析ごとの課題

2.1.1 情報検索

本論文では 2 つの課題を設定する。課題 1 は、UMLS を活用して診断に関する情報検索のために単語タイプを識別し、その文書分析における重要語の重みを上昇させることである。なぜなら、タスクの重要語に高い重みを付与することは、情報検索の精度向上に結び付くと考えられるためである。課題 2 は、共起の偏りに基づいて医学単語の一般性及び特定性を考慮する重み付けをすることである。なぜなら、高血圧のような様々なタイプの医学単語と共起することが想定される単語がクエリーにあり、かつ、その単語に高い重みが与えられると、クエリーと関係のない文書が多く検索されると考えられるためである。

先行研究では、医学論文を用いて、医学オントロジーの UMLS により得られる医学単語のタイプを活用した研究がある [53, 58]。また、一般的な文書に対して単語の一般性及び特定性を考慮した研究では、WordNet のオントロジーを活用した手法 [54, 40] や文書のポジティブクラスとネガティブクラスを活用した手法 [19, 17, 23] がある。

しかしながら、UMLS により得られる医学単語のタイプだけでなく、情報検索における

表 2.1: 先行研究のまとめ

	オントロジー	単語の クラス情報	医学論文への適用	文書分析
(Tar, 2011) (Sureka, 2012)	○			クラスター分析
(Zakos, 2006) (Sakre, 2009)	○			情報検索
(Varelas, 2005) (Jing, 2006) (Zhang, 2007) (Zhang, 2008)	○			情報検索，クラスター分析
(Lan, 2006) (Martineau, 2009) (Ko, 2015)		○		分類，センチメント分析
(Luo, 2011)	○	○	△	分類
(Zhang, 2007) (Zhang, 2008)	○		○	クラスター分析
(Yu, 2009)	○	○	○	情報検索
(Zhu, 2006)	○	○	○	情報検索

医学単語の識別力を意味する医学単語の特定性（単語特定性）も考慮して、医学単語に重み付けする手法はこれまでみられない。そのため、本論文では、医学単語の共起の偏りという観点から、医学単語の一般性及び特定性を考慮し、医学重要度と単語特定性という2つの要素の組み合わせによる4つの重み付け手法を提案する。そして、評価セットが利用可能なTREC 2014 Clinical Decision Support Track (TREC 2014 CDS Track)[42]の英文の医学論文データを用いて、診断に関する情報検索の実験により、4つの重み付け手法の分析することで、医学単語の特定性を捉えることの有効性を示す。この提案手法は第4章の研究1で記述する。

2.1.2 予測（死亡予測）

課題は、患者状態の重症度の観点から単語の意味を捉え、重症度の高い重要語の重みを上昇させることと設定する。なぜなら、患者が亡くなるか亡くならないかという患者のリスクの予測精度の向上に寄与すると考えられるためである。切に捉えた上で単語に対して重み付けする必要があると考えられる。

死亡予測のための手法では、sequential organ failure assessment (SOFA) や simplified acute physiology score (SAPS) といったスコアリングや、スコアリングを用いずにアルゴリズムを開発する研究 [34, 13, 36, 11] がみられる。TFIDF による重み付けを大腸癌の予測モデルの構築に適用している研究がある [10]。しかしながら、TFIDF を拡張して、患者状態の重症度をもとに単語の意味を考慮した重み付け手法はこれまでみられない。そのため、本論文では、患者状態の重症度という視点から医学単語の意味的な重要度を捉える手法を提案する。そして、英文の電子カルテが含まれる MIMIC II データベース [39] を用いた死亡予測の実験により、患者状態の重症度を捉えることの有効性を示す。この提案手法は第5章の研究2で記述する。

第3章　提案手法のフレームワーク

本章では、まず、フレーム意味論の考え方を単語の重み付けへ適用する手順を述べ、その手順に基づく重み付けのフレームワークを示す。このフレームワークでは、フレーム意味論の考え方を取り入れながら、関連研究同様、意味を考慮した単語の重み付け手法の開発において、オントロジーと単語のクラス情報を活用する方法を述べる。そして、医学文書分析に共通する意味を考慮した単語の重み付けフレームワークを述べる。このフレームワークは、意味の視点を階層上の単語集合により表現し、文書分析に共通する部分と依存する部分の両方を含む新たな医学知識表現に基づいている。ここでは、提案する医学知識表現の基本的な考え方や性質、構築方法を述べ、医学知識表現を組み込んだ意味を考慮した単語の重み付けを述べる。最後に提案手法の本論文内の位置づけを述べる。

3.1　フレーム意味論の考え方を単語の重み付けへ適用するための手順

本論文で提案する意味を考慮した単語の重み付け手法は、状況を特徴づけるフレームが単語の意味を割り当てるとするフレーム意味論の考え方を適用している。図 3.1 はフレーム意味論の考え方を単語の重み付け手法へ適用するための 3 つのステップである。以下にそれぞれのステップについて述べる。

ステップ 1 では、単語の意味を考慮しながら重み付けをする時に、どのような視点から単語の重要度を捉えるかを設定する。その際、設定する視点による重み付けをどのような実践の場に適用するかという文書分析の目的も併せて考慮する。既に目的を設定している場合は、その目的に適した単語の重要度の視点を設定する。このステップでは、特に百科

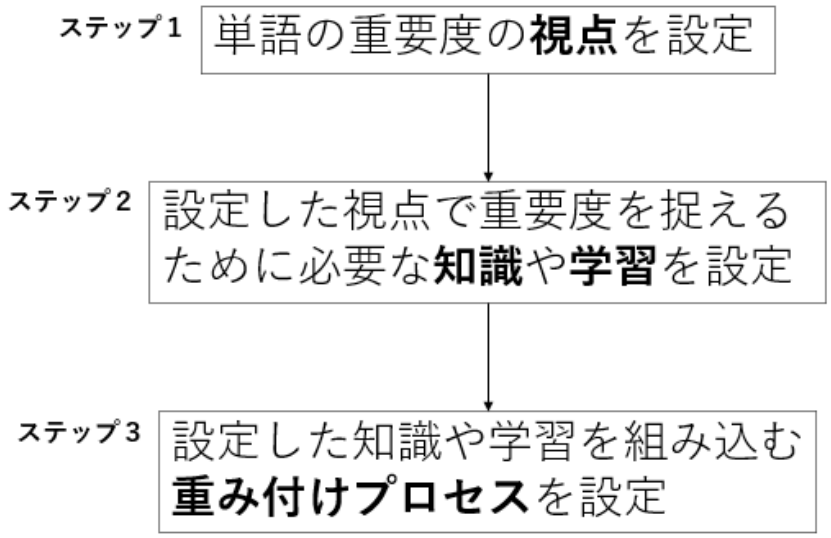


図 3.1: フレーム意味論の考え方を単語の重み付けに適用する手順

事典的意味という具体的な応用に適した意味を考慮することが重要な点である。ステップ 2 では、設定した視点から計算機が単語の重要度を捉えるために必要な知識や学習を設定する。まず、設定した視点で重要度を捉えるのに適した既存の知識を選定する。適した知識が存在しない場合は、データからの学習により設定した視点における重要度を捉える。ステップ 3 では、設定した知識や学習を組み込む重み付けプロセスを設定する。知識や学習を設定するだけでなく、それらをどのように重み計算の過程に組み込むかが重要である。これらの 3 つのステップによりフレーム意味論の考え方を適用した単語の重み付けが実現される。

3.2 提案手順に基づく重み付けフレームワーク

ここでは、3.1 で示した提案手順に基づき、フレーム意味論の考え方を取り入れながらオントロジーと単語のクラス情報の活用による文書分析の目的ごとの重み付けのフレームワークを示す。図 3.2 は提案手順に基づく重み付けフレームワークである。

医学文書の情報検索に適した重み付け手法の開発では、単語の重要度の視点を情報検索の識別力を意味する医学単語の特定性とする。その視点から単語の重要度を捉えるために、主に機械学習による単語のトピック情報を活用して医学単語の特定性及び一般性を捉

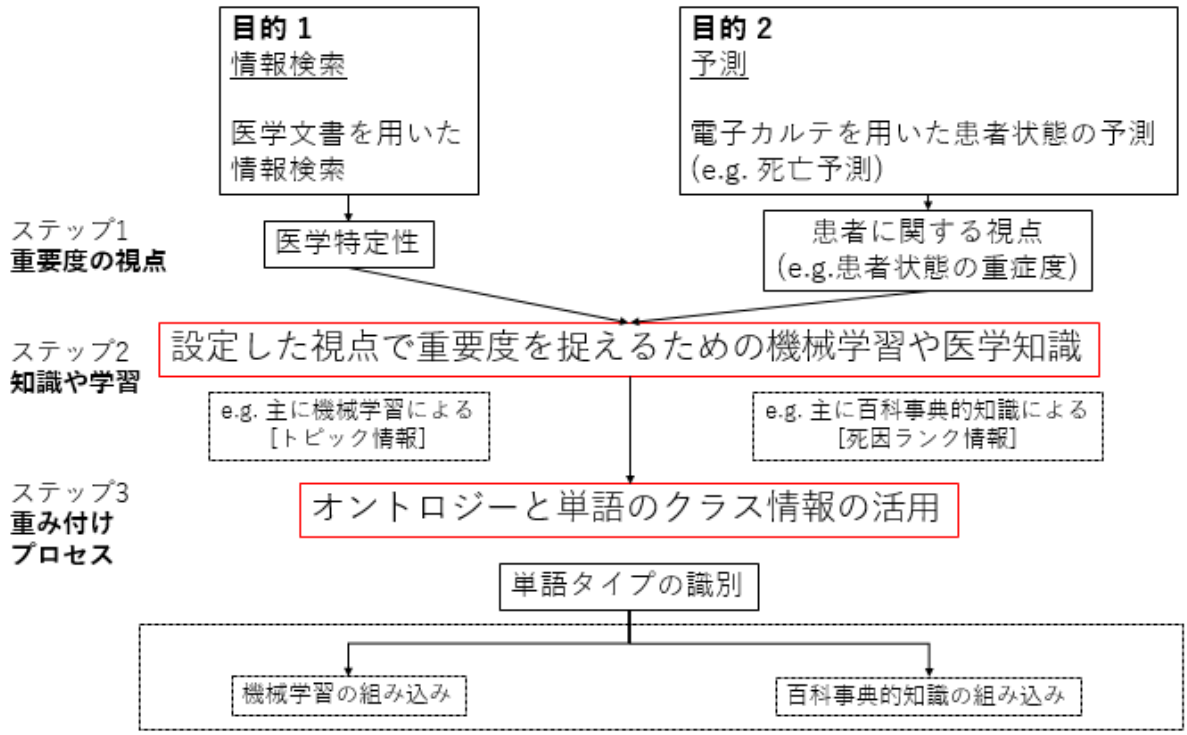


図 3.2: 提案手順に基づく重み付けフレームワーク

える。電子カルテを用いた患者状態の予測に適した重み付け手法の開発では、単語の重要度の視点を患者状態の重症度とする。その視点から単語の重要度を捉えるために、主に医学の百科事典的知識による死因ランク情報を活用して、患者が亡くなるか亡くならないかという予測のために医学単語が持つ患者へのリスク度合いを捉える。設定した知識や学習を組み込むための重み付けプロセスは、特にオントロジーと単語のクラス情報を活用する。その際、2 つの目的に共通する重み付けプロセスとして、医学文書内の単語のタイプを識別する。その後、目的ごとで機械学習や百科事典的知識を組み込むことにより、それぞれの目的に適した意味を考慮した単語の重要度を捉える重み付け手法を開発する。

3.3　文書分析に共通する重み付けフレームワーク

文書分析に共通する重み付けフレームワークは、階層上で意味の視点を単語集合で表現し、文書分析に共通する部分と依存する部分を含む医学知識表現に基づく。以降で提案する医学知識表現の基本的な考え方や性質、構築方法を述べる。最後に医学知識表現を組み

込んだ意味を考慮した単語の重み付けについて述べる。

3.3.1 医学知識表現の基本的な考え方

膨大な医学単語それぞれに対して、ある意味の視点から異なる重みを付与することは困難であると考えられる。そこで、医学文書内の単語をある視点により階層的にカテゴリー化をして、そのカテゴリーに対して重みを与え、連続値で重みを付与できる場合を除いて、カテゴリー内の全ての単語は該当するカテゴリーの重みが一意に与えられるものとする。ここでは、カテゴリー内の単語集合を意味の視点として計算機に表現させる。階層的に区分けすることで、階層の浅い単語集合は様々な文書分析に共通する部分となり、階層の深い単語集合はある特定の文書分析に依存する部分となる。つまり、文書分析に共通する部分の単語集合はより抽象的で、文書分析に依存する部分の単語集合はより具体的であるといえる。このように、提案する医学知識表現は、文書分析に共通する部分と依存する部分の両方を含んでいる。

3.3.2 医学知識表現の性質と構築方法

本論文で提案する医学知識表現の基本的な性質は、階層ごとの単語集合が is-a 関係である。例えば、医学単語は医学文書内の単語の一種となる。また、論理否定（¬）を活用した区分けをしていることも特徴である。例えば、医学文書内の単語を非医学単語と医学単語のように分類する。

本論文では，文書分析に共通する部分において ICD-10 を活用している．ICD は、もともと死亡原因を分類するためのもので、死亡率のデータ処理に適用されている [50]。ICD の第 10 版である ICD-10 は、患者の重症度を捉えるのに適しているだけでなく、疾病や症状、所見に関する医学単語を扱うため、疫学や健康管理といった多くの目的で用いられている。したがって、ICD-10 は文書分析に共通する部分に適した重要な要素であると考えられる。ただし、目的に応じて MeSH などの他の医学単語のコードを用いることも考

えられる。

医学文書内の単語のタイプを識別し、単語をカテゴリーごとに分けるために、オントロジーと単語のクラス情報を活用する。文書分析に共通する部分では、医学文書内の単語を医学単語でない一般的な単語（非医学単語）、医学単語であるがICD-10コードを持たない単語（非ICD-10単語）、医学単語の中でもICD-10コードを持つ単語（ICD-10単語）の3つのタイプのいずれかに区分けする。医学文書内の単語を3つの単語タイプに区分する方法は2つのステップにより構成される。ステップ1では、医学文書内の単語が医学単語かどうかを識別する。そのため、本論文ではUMLSの医学オントロジーを活用する。ここでは、医学単語を識別するためにUMLSのConcept Unique Identifiers (CUI) を活用する。CUIはUMLSのメタシソーラス上の大文字のCと7つの数字で表現される単語の概念コードである。医学文書内の単語のCUIコードを取得するため、MetaMapというマッピングツールを活用する[1]。マッピングをする際は、本論文の対象となる文書分析（診断に関する医学論文の情報検索と電子カルテの死亡予測）を考慮して、疾病や症状、所見及び薬に関する医学単語に焦点を当てる。最終的に医学文書内の単語がCUIコードを保有している場合は医学単語、そうでない場合は非医学単語とする。ステップ2では、医学単語がICD-10を保持しているかどうかを識別する。そのために医学単語のCUIコードを活用してICD-10コードのマッピングを試みる。UMLSのツールではそのマッピングができないため、本論文ではCUIコードをICD-10へマッピングできるBioPortal[29]という医学オントロジーを用いる。最終的に、ICD-10コードを持つ医学単語をICD-10単語とし、そうでない単語は非ICD-10単語とする。このように、2つのステップにより医学文書内の単語に分けられ、いずれかの単語集合に属すことになる。

図3.3は、医学文書に共通する部分において区分けされる3つの単語タイプとその識別のための主な医学知識を示している。

単語のタイプを識別するためには、特に単語の概念を取り扱うオントロジーの活用が有効であると考えられる。図3.4で示すように、オントロジーを辞書的役割と他の知識へのマッピングのために活用することで、単語のタイプが識別できる。そして、これらの単語

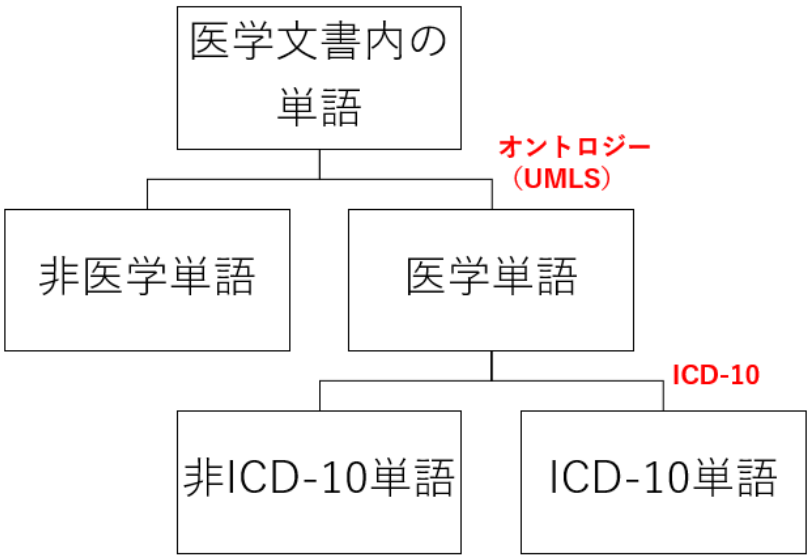


図 3.3: 医学文書分析に共通する部分の医学知識表現

のカテゴリー情報をもとに計算機に学習をさせることや、百科事典的知識を組み込むことで、ある目的に適した百科事典的意味を捉えるのに役立つと考えられる。

本論文で扱う情報検索と予測の 2 つの文書分析に依存する部分は、図 3.5 のように階層的に単語を分類する。電子カルテを用いた患者状態の予測の 1 つである死亡予測では、死因ランキングという百科事典的知識を活用して、ICD-10 単語を ICD-10 ランク単語と ICD-10 ランク外単語に分け、ICD-10 ランク単語の場合は 15 の死因ランクを含む死因ランキング内の該当するランクに分ける。伝統的な重み付け手法の TFIDF は、単語に対して統計的な要素を用いてランク付けが可能であるが、提案する方法は、百科事典的知識である死因ランキングといったランキングを活用することで、単語集合で構成されるカテゴリーごとに対応するランク情報により、カテゴリー間の優劣を求めている点が特徴である。医学文書を用いた情報検索では、ICD-10 単語の場合、Latent Dirichlet Allocation（LDA）という機械学習手法により得られる単語のトピック情報を活用して、単語の共起の偏りという観点から特徴的な単語と一般的な単語に区別する。図 3.5 では単語の共起の偏り度合いの意味合いを示すため、特徴的な単語と一般的な単語の 2 つのカテゴリーに区分けしている。伝統的な重み付け手法の TFIDF は、単語の頻度情報に基づいて、単語に連続値で重みを付与できるが、提案する方法は、機械学習を用いることで、単語の共起の

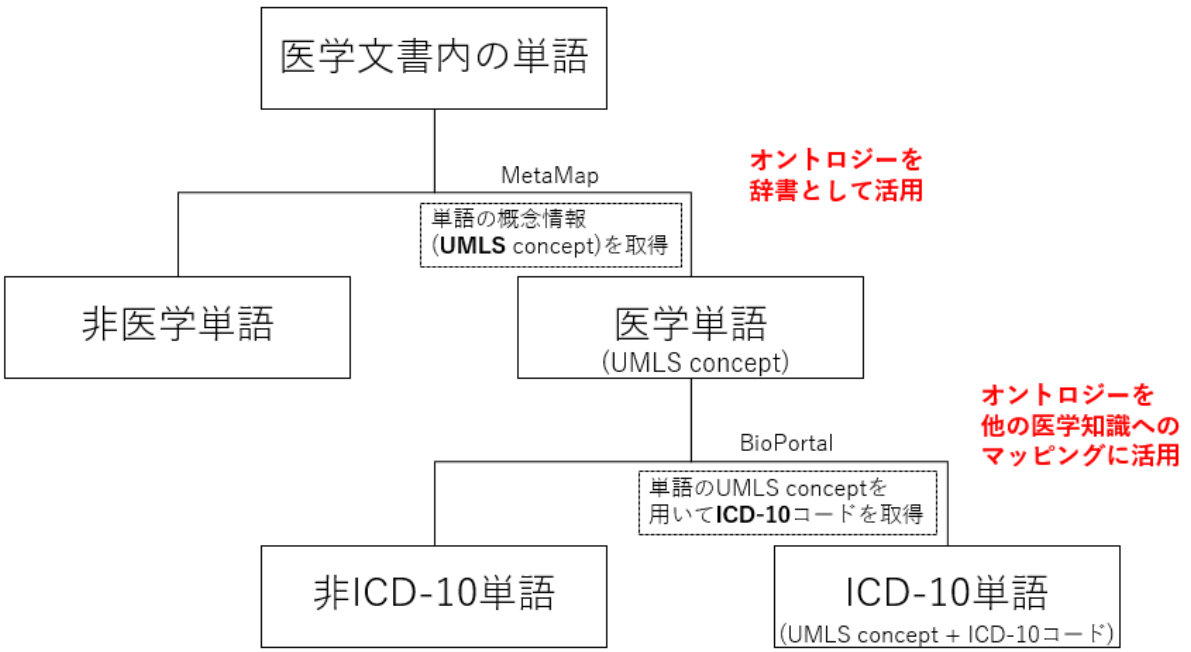


図 3.4: オントロジーの活用による単語タイプの識別

偏りの観点から単語の優劣を連続値の重みによりつけられる点が特徴である。

本論文で扱う情報検索と予測の 2 つの文書分析に依存する部分の構築方法は、それぞれ、第 4 章研究 1（情報検索のための意味を考慮した単語の重み付け）と第 5 章研究 2（予測のための意味を考慮した単語の重み付け）で詳しく述べる。図 3.5 内の赤文字は、上記の文書分析に依存する部分を構築するために重要な知識や学習である。

3.3.3 医学知識表現を組み込んだ意味を考慮した単語の重み付け

提案する医学知識表現を組み込んだ意味を考慮した単語の重み付けは以下の特徴を持つ。

3.3.3.1 階層ごとの単語集合が is-a 関係

意味を考慮した単語の重み付けにおいて、ある意味の視点から単語に重要度を与え、単語間に差異をもたらすために、提案する知識表現の構成要素である is-a 関係で示された単語集合の階層の粒度を活用する。それにより、提案する重み付け手法は、浅い階層の単語集合よりも深い階層の具体的な単語集合に対してより高い重みを付与する。例えば、文

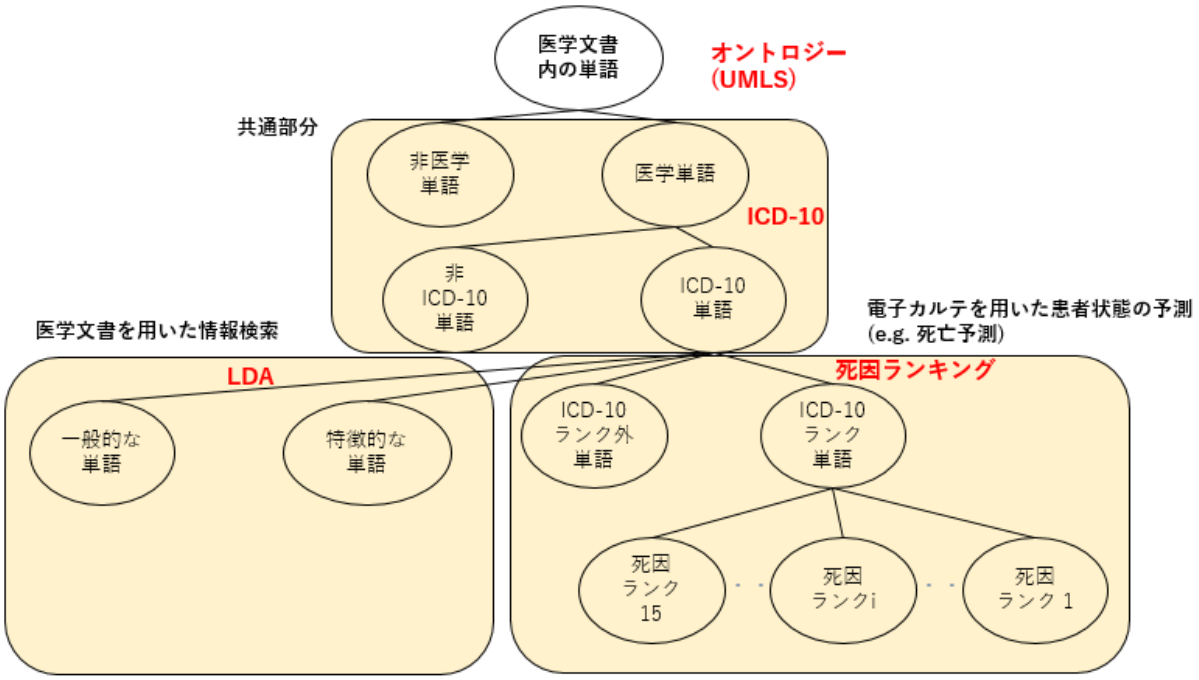


図 3.5: 文書分析に依存する部分を含んだ医学知識表現

書分析に共通する部分では、ICD-10 単語に医学単語よりも高い重みを与える。文書分析に依存する部分では、例えば、ICD-10 ランク単語に ICD-10 単語よりも高い重みを与える。このように、階層の深い具体的な単語集合に対してより高い重みを与える。

3.3.3.2　論理否定を活用した区分け

医学文書内の単語が A でないという場合よりも A であるという場合により高い重みを付与する。ここで、A に該当する単語は、医学文書分析において重要語である。例えば、医学単語である場合と医学単語でない場合（非医学単語）や、ICD-10 単語である場合と ICD-10 単語でない場合（非 ICD-10 単語）では、医学文書の分析で重要語と考えられる医学単語と ICD-10 単語に対して、それぞれ非医学単語と非 ICD-10 単語よりも高い重みを与える。

3.3.3.3　ランキングで表された医学知識の活用

ランキングで表された知識を活用することで、単語集合で構成されるカテゴリーごとに対応するランク情報に基づいてカテゴリー間の優劣をつけて重みを与える。例えば、15

の死因ランクを含む死因ランキングにおいて、死因ランク 1 位に該当するカテゴリーの重みが 15 の死因ランクのカテゴリー間で最も高くなる。ランキングにおいて、ランクが高ければ高くなるほど、対応するカテゴリーの重みが高くなる。

3.3.3.4　機械学習の活用による連続値の重みの付与

機械学習手法を活用することで、ある意味の視点から単語の重みを連続値で捉える。例えば、LDA により得られる単語のトピック情報を活用して、単語の共起の偏りという観点から単語の重みを連続値で求め、特徴的な単語と一般的な単語という共起の偏り度合いを捉え、特徴的な単語であればあるほど高い重みが与えられる。

このように、医学知識表現を組み込んだ意味を考慮した単語の重み付けにより、計算機に人が想起する単語の様々な意味の視点を理解させた上で単語の重み付けが可能になると考えられる。

本論文では、単語自身が持つ医学的な重要度を医学重要度とする。文書分析で共通する部分の単語集合の医学重要度の重みは、医学知識表現の性質に基づいて、非医学単語、非 ICD-10 単語、ICD-10 単語の順に増加する。このように、深い階層の単語集合が浅い階層の単語集合よりも高い重みとなっている。また、論理否定を活用した区分けでは、例えば、非医学単語よりも医学単語の中の非 ICD-10 単語に高い重みを与えている。非 ICD-10 単語よりも ICD-10 単語に高い重みを与える。このように、階層上で、A でないという場合よりも A であるという場合により高い重みが与えられる。

このような重み付けをする理由は、どの文書分析においても、基本的には医学単語でない単語よりも医学単語が重要語であると考えられるためである。また、医学単語の中でも特に ICD-10 コードを持つ医学単語に医学重要度としてより高い重みを与える。なぜなら、ICD-10 コードを持つ医学単語は、疾病や症状、所見に関する単語であることから、医学文書の情報検索や患者の情報を含む電子カルテを用いた患者状態の予測において重要な要素だと考えられるためである。文書分析に依存する部分の単語集合の医学重要度の重みは、情報検索の分析では、ICD-10 単語の場合に、そのカテゴリーの重みと連続値の重み

の組み合わせにより求められる。予測の分析では、ICD-10 ランク単語の場合に、ICD-10 単語のカテゴリーの重みよりも高い死因ランキングにより該当するカテゴリーの重みを用いる。

3.4 提案手法の本論文内の位置づけ

図 3.6 は提案手法の本論文内の位置づけを示している。以降の第 4 章では、研究 1 として医学文書の中でも医学論文を用いた情報検索を目的に、医学特定性という重要度の視点から意味を考慮した単語の重み付け手法を述べる。第 5 章では、研究 2 として電子カルテを用いた患者状態の予測の 1 つである死亡予測を目的に、患者状態の重症度という重要度の視点による意味を考慮した単語の重み付け手法を述べる。本論文では患者状態の重症度を医学単語の独立型と依存型の重みから求め、独立型の重みを考慮した重み付けを研究 2 の主な手法とする。独立型の重みを考慮した重み付けは文書内の文脈によらず、患者状態の重症度の観点から一意に重みを付与する。一方、依存型の重みを考慮した重み付けは文書ごとで異なる患者状態の重症度の文脈により重みを可変する。本論文では、独立型だけではなく依存型の重みも考慮した手法を改良手法 1 とし、研究 2 で述べる。最後に、医学単語の独立型の重みを考慮した重み付けを改良した手法を改良手法 2 とし、研究 2 で述べる。

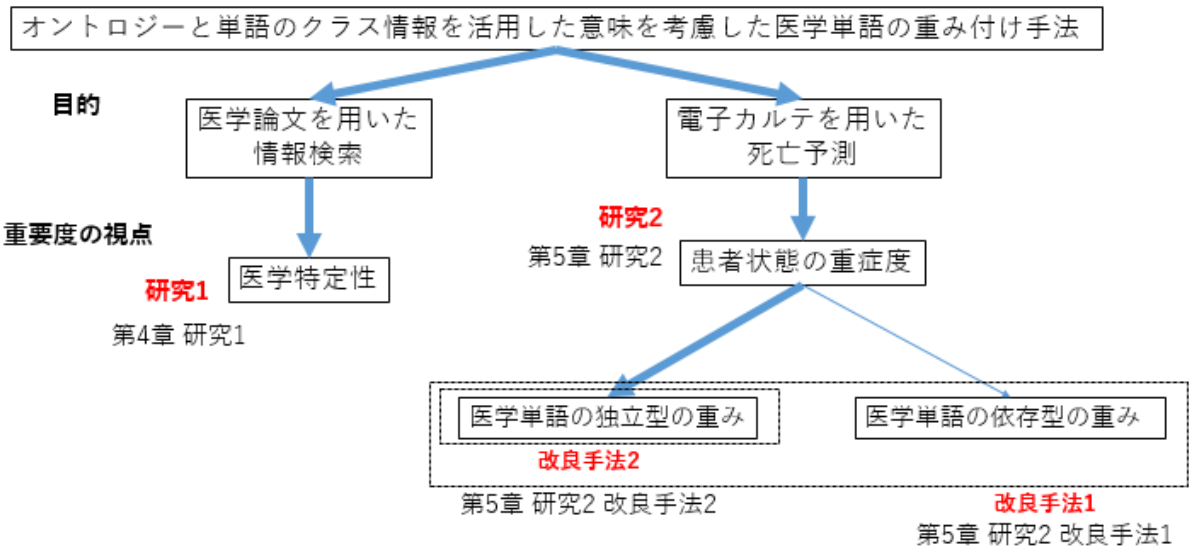


図 3.6: 提案手法の本論文内の位置づけ

第4章　研究1（情報検索のための意味を考慮した単語の重み付け）

4.1　目的

本研究は、単語のトピック情報と単語のタイプといったクラス（カテゴリー）情報と医学オントロジーの組み合わせから、単語特定性という重み要素を提案する。そして、医学重要度による重みとの組み合わせによる意味を考慮した4つの医学単語の重み付け手法を分析する。

4.2　提案手法

4.2.1　提案手法の基本的な考え方

本研究では、単語の共起の偏りを単語特定性とする。医学文書内の医学単語の一般性及び特定性について考えてみると、例えば”高血圧”という医学単語は様々なタイプの医学単語と共起することが想定される。このような単語がクエリーに含まれており、かつ高い重みが与えられてしまうと、そのクエリーと関係のない文書が検索で返されてしまう恐れがあると考えられる。そこで、様々なタイプの単語と共起する傾向にある医学単語の一般性及び特定性を捉えるため、本研究ではLatent Dirichlet Allocation (LDA)[2]から得られる医学単語のトピック情報を活用する。なぜなら、そのトピック情報はある医学単語のそれぞれのトピックに対する出現確率で構成されているため、それら複数のトピックに対する出現確率を活用することで、様々なタイプの単語と共起するような医学単語の共起の偏りを捉えられるからである。しかしながら、トピックに対する出現確率のみでは共起の偏

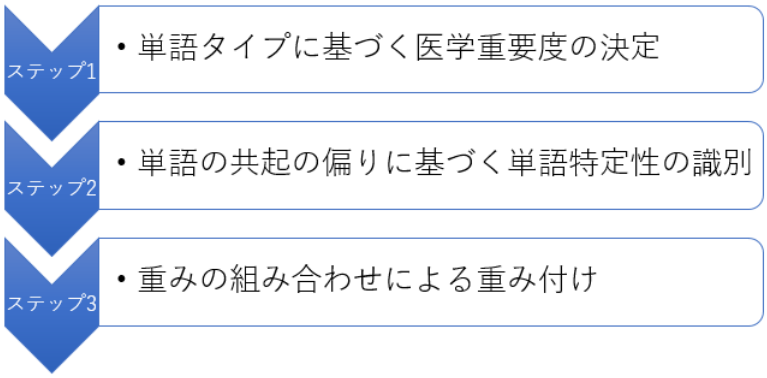


図 4.1: 研究 1 の提案手法の 3 つのステップ

りの程度まで捉えられないため、本研究ではエントロピーの不確実性の概念を活用して単語の共起の偏りの度合いをトピックの出現確率情報から捉える。

共起の偏りの程度は情報検索のタスクにおける医学単語の識別力を示しており、例えば他の医学単語と偏った共起の仕方をする医学単語がクエリーにあった場合、その単語の情報検索のタスクにおける識別力が高いとされ、共起の偏りの程度に基づきその単語の重みを高くする。本研究ではこのような共起の偏りの程度に基づいた重みを医学単語の特定性（単語特定性）とし、医学文書の情報検索の精度を高めるため、医学重要度に加えて医学単語に重み付けする際の要素として用いる。

医学重要度と単語特定性の組み合わせによる 4 つの重み付け手法は、ベースラインによる重みと提案する重み要素が組み合わされる。4 つの手法のうち、2 つの手法は医学重要度のみを用いた場合で、扱う単語タイプ数の違いにより 2 つの手法に分けられる。もう 2 つの手法は、上記の 2 手法それぞれと単語特定性を組み合わせた場合である。

4 つの重み付け手法を導くプロセスは、図 4.1 のように 3 つのステップで構成される。始めに、単語タイプに基づいて医学重要度を決定する。次に、単語の共起の偏りに基づいて医学単語の単語特定性を識別する。最後に、医学重要度と単語特定性の 2 つの要素の組み合わせにより意味を考慮した 4 つの単語の重み付け手法を導く。それぞれのステップの詳細は以下となる。

4.2.2　単語タイプに基づく医学重要度の決定

ステップ 1 では、本論文の共通する重み付けフレームワークにより、単語タイプを非医学単語、非 ICD-10 単語、ICD-10 単語という 3 つの単語タイプを識別している。ここでは、それぞれをタイプ 1、タイプ 2、タイプ 3 とし、医学重要度は 0、0.04、0.08 と設定している。

4.2.3　単語の共起の偏りに基づく単語特定性の識別

ステップ 2 では、タイプ 2 及びタイプ 3 の医学単語に対象に、共起の偏りという観点から単語の特定性を計算する。様々な種類の単語と共起する単語に高い重みを与えると、クエリーと無関係な文書が検索されると考えられるからである。そこで、本研究では、LDA から得られるトピックの出現確率情報を用いて、ある特定のトピックに出現する確率が高い単語（ケース 1）には、複数のトピックで高い出現率を持つ単語（ケース 2）よりも高い重みを与える。例えば、高血圧、皮膚病変、スキャンの医学単語において、皮膚病変が最も共起の偏りが高く、スキャンが最も共起の偏りが低い。この場合、皮膚病変とスキャンはそれぞれケース 1 とケース 2 に該当する。したがって、皮膚病変に対してより高い重みを与えるのが単語特定性の特徴である。

本研究では、単語特定性の程度を求める際に、エントロピーの不確実性の概念を活用する。なぜなら、上記のケース 1 とケース 2 の状況はエントロピーの視点で捉えられるからである。図 4.2 はケース 1 とケース 2 に該当する単語の性質をまとめたものである。図 4.2 内の表の x 軸はトピック番号で y 軸はトピックの出現確率である。

図 4.2 で示すように、ケース 1 は単語がどのトピックに出現するかの不確実性が低く、エントロピー値も低くなる。この場合、共起の偏りの程度が高いことから、単語特定性が高くなる。このような単語は情報検索における識別力が高いと考えられるため、情報検索で重要語といえる。ケース 2 の場合は不確実性が高く、エントロピー値も高くなる。この場合、共起の偏りの程度が低いことから、単語特定性が低くなる。このような単語は情報

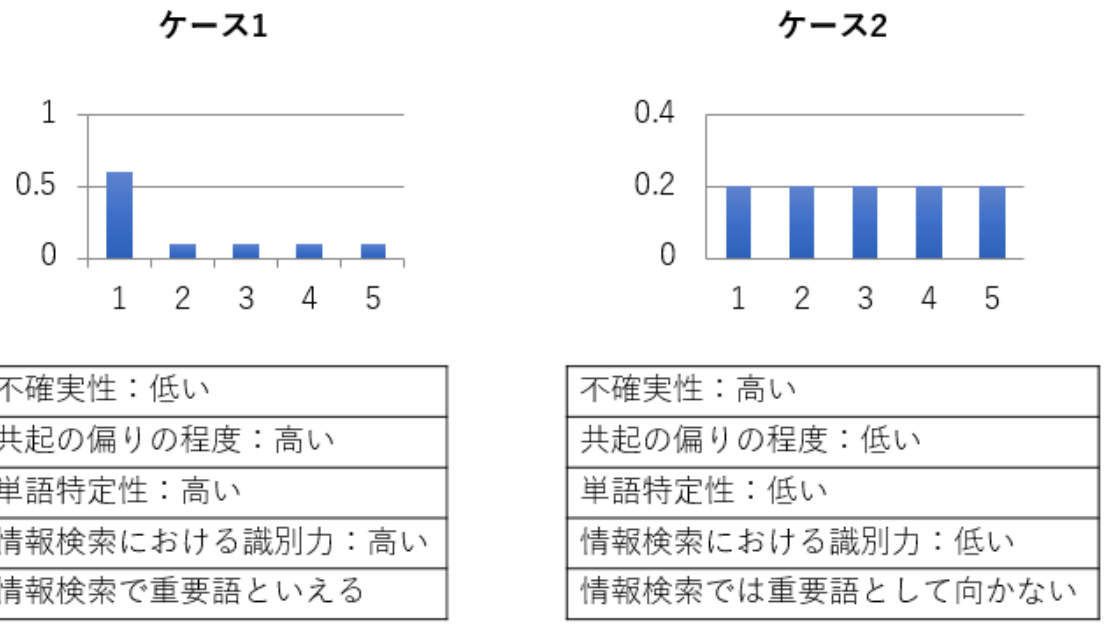


図 4.2: ケース 1 とケース 2 の単語の性質

検索における識別力が低いと考えられるため、情報検索では重要語として向かないと考えられる。

上記の状況では、不確実性が低く、エントロピー値が低い場合に単語特定性が高くなることから、1 からエントロピー値（Min-Max 正規化後）を引いた値を特定性の程度とする。

$$H_t = -\sum_{k=1}^{n} p_t^k \times log_2 p_t^k$$

$$TS_t = 1 - H_t$$

n はトピックの数であり、本研究では 20 とする。p_t^k はトピック k に対する単語 t の出現確率を単語 t の全トピックの出現確率の合計で割った値である。H_t は単語 t のエントロピー値である。TS_t は単語 t の特定性であり、1 からエントロピー値 H_t を引いて求める。ここで用いるエントロピー値は Min-Max 正規化による正規化後の値である。

しかしながら、TS_t は文書頻度が低い単語に対して高くなる傾向がある。したがって、文書頻度（Min-Max 正規化後）の対数をその値に掛けたものが最終的な単語特定性 $\acute{TS}_t$ となる。

$$\acute{TS}_t = TS_t \times log_2 c_t$$

c_t は Min-Max 正規化による正規化後の単語 t の文書頻度である。

LDA の実行時は文書頻度が 10 以上の医学単語を対象とし、イテレーション回数は 1500 とする。

4.2.4 重みの組み合わせによる重み付け

ステップ 3 では、医学重要度の重み MI と単語特定性の重み TS を組み合わせて、意味を考慮した 4 つの重み付け手法を導く。提案手法はベースラインによる重みと提案する重み要素が組み合わされる。

MI と TS はそれぞれ以下の式 (4)、式 (5) によりベースラインの重みと組み合わされる。ここで、ベースラインは BL と表記される。

$$(\alpha_1 \times BL) + (\alpha_2 \times MI) \, (4)$$

$$(\alpha_1 \times BL) + \left\{\alpha_2 \times \left(\tfrac{TS}{2}\right)\right\} (5)$$

α_1 と α_2 はそれぞれベースラインと MI 及び TS の重みの係数であり、それぞれ 1、0.05 とする。ベースラインによる重みと医学重要度の重み及び単語特定性の重みの調整のため、これらの係数を用いる。

本研究では、上記のベースラインとの組み合わせ方をもとに、4 つの重み付け手法を以下のように提案する。手法 1、2、3、4 はそれぞれ以下の式 (6)、(7)、(8)、(9) により重み付けがされる。

$$w_{jt} = (\alpha_1 \times BL_{jt}) + (\alpha_2 \times MI_t^1) \, (6)$$

$$w_{jt} = (\alpha_1 \times BL_{jt}) + (\alpha_2 \times MI_t^2) \, (7)$$

$$w_{jt} = \begin{cases} (\alpha_1 \times BL_{jt}) + (\alpha_2 \times MI_t^1) & , c_t < 10 \\ (\alpha_1 \times BL) + \left\{\alpha_2 \times \left(\tfrac{TS}{2}\right)\right\} & , c_t \geqq 10 \end{cases} (8)$$

$$w_{jt} = \begin{cases} (\alpha_1 \times BL_{jt}) + (\alpha_2 \times MI_t^2) & , c_t < 10 \\ (\alpha_1 \times BL) + \left\{\alpha_2 \times \left(\frac{TS}{2}\right)\right\} & , c_t \geqq 10 \end{cases} \quad (9)$$

w_{jt} は文書 j の単語 t に対する重みである。BL_{jt} は文書 j の単語 t に対するベースラインにより求められる重みである。MI_t^1 はタイプ 1 またはタイプ 2 の単語を対象とした場合の医学重要度の重みで、0 または 0.04 の値をとる。MI_t^2 はタイプ 1、タイプ 2 またはタイプ 3 の単語を対象とした場合の重みで、0、0.04 または 0.08 の値をとる。

手法 1（式 (6)）は医学重要度のみを考慮した手法である。医学重要度の重みを求める際は、タイプ 1 またはタイプ 2 の 2 つの単語タイプを扱う。したがって、手法 1 は医学単語（タイプ 2）に対して医学重要度として一意に重みを増大させる。

手法 2（式 (7)）は医学重要度のみを考慮した手法であるが、医学重要度の重みを求める際は、タイプ 1、タイプ 2 またはタイプ 3 の 3 つの単語タイプを扱う。したがって、手法 2 は医学単語の場合、診断に関連する疾病や症状、所見を表す ICD-10 コードを持つタイプ 3 の単語に対して ICD-10 コードを持たないタイプ 2 の単語よりも高い重みを付与する。

手法 3（式 (8)）は医学重要度と単語特定性の重みを組み合わせた場合の手法である。単語の文書頻度が 10 以上の場合には単語特定性を考慮した式 (5) を用い、文書頻度が 10 未満の場合は医学重要度を考慮した式 (4) を用いて単語の重みを求める。手法 3 はタイプ 1、タイプ 2 の単語を対象として医学重要度の重みを求める。

手法 4（式 (9)）は医学重要度と単語特定性の重みを組み合わせた手法であるが、医学重要度の重みを求める際は、タイプ 1、タイプ 2 またはタイプ 3 の単語を対象とする。本研究では、医学重要度で扱う単語をタイプ数で区別したことによる差異をみるために、手法 3 と手法 4 に分ける。

このような意味合いを持つ 4 つの重み付け手法が、医学文書の情報検索においてどのような効果をそれぞれがもたらすのかを実験分析する。

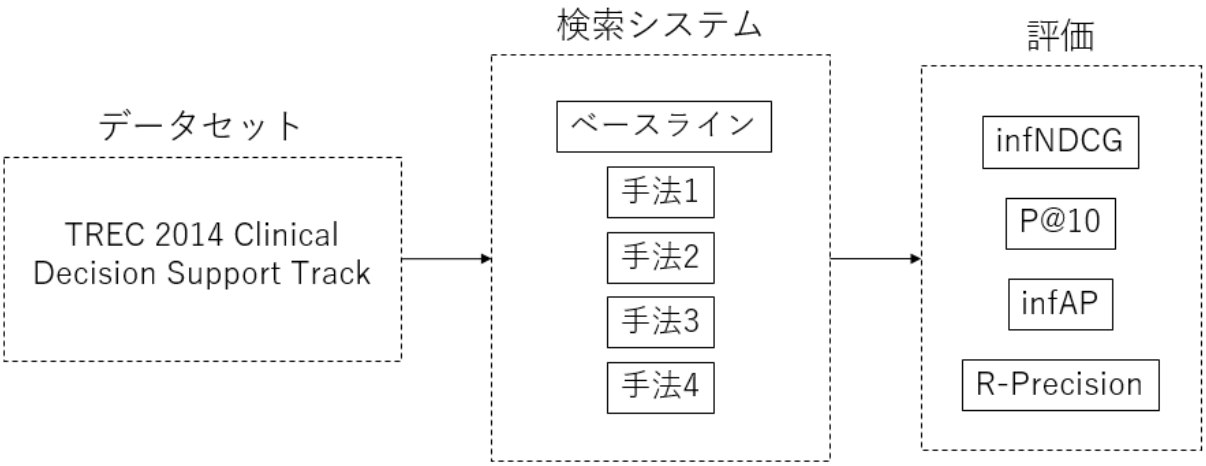


図 4.3: 情報検索の実験概要

4.3 実験による評価

4.3.1 実験設定

図 4.3 は本研究の実験の概要である。データセットは医学論文の情報検索のタスクである TREC 2014 CDS Track を用いる。このデータセットは Diagnosis、Test、Treatment という 3 タイプのクエリーに関して合計 30 のクエリーが含まれている。それぞれのクエリーは Description と Summary の 2 種類の文で記述されている。本研究では、特に Summary 文で記述された diagnosis に関する 10 のクエリーと、その関連度のラベル付けがされた 11839 文書（重複なし）を扱う。

クエリー内の医学単語の句を考慮するため、UMLS を用いてクエリー拡張がされる。

クエリーと検索対象の文章は、前処理としてストップワードの除去とチャンキングが施される。また、否定検出を MetaMap により行い、否定語に係る単語を取り除く。

本研究ではベースラインとして TFIDF と BM25 を用いる。BM25 は、TFIDF から文書長を考慮した手法である。文書単語頻度と文書長の影響を調整するパラメータ k_1 と b の適切な値は、k_1 の場合 1.2 から 2.0 の間で、b は 0.75 とされる [22]。よって、b の値を 0.75 とし、本研究ではパラメータの影響を抑えるために k_1 の値を 1.2 と設定する。

これら 4 つの評価指標を用いて 4 つの重み付け手法を Baseline として用いる TFIDF 及

び BM25 と比較する。

クエリーと検索対象文書の関連度は、コサイン類似度で算出される。4 つの重み付け手法は、関連度の高い順で上位 1000 文書を検索結果として、4 つの評価指標により検索性能が評価される。

4.3.2 評価指標

TREC 2014 CDS Track 指定の評価指標は infNDCG、P@10、infAP、R-Precision の 4 つである。それぞれの指標の定義とその意味合いを以下に示す。ただし、TREC 2014 CDS Track では上位 1000 文書の検索結果を対象としている。

infNDCG: NDCG (Normalized Discounted Cumulative Gain) は、検索対象文書の関連度の程度を考慮した指標である。関連度のより高い文書が検索の上位に現れるほど、検索性能が高いとする。

P@10: P@10 は検索結果の上位 10 文書までの精度を求める指標である。検索システムのトップ 10 の検索結果における正確性を意味する。

infAP: AP (Average Precision) は平均精度で、関連する文書が現れた位置における精度の平均値である。関連する文書が上位で返されるほど検索の精度が高いとする。

R-Precision: R-Precision はクエリーに対して関連する文書数が R の場合に、検索結果の上位 R 件までの精度を求める指標である。検索結果の上位から関連文書数までの検索性能の正確性を意味する。

4.3.3 実験結果

表 4.1 と表 4.2 はそれぞれベースラインが TFIDF と BM25 の場合の結果である。TFIDF と 4 つの重み付け手法の 5 つの手法それぞれに対する 4 つの評価指標の値を示している。表内の値は 10 のクエリーの結果の平均である。

表 4.1: ベースラインが TFIDF の場合の結果

	infNDCG	P@10	infAP	R-Precision
TFIDF	0.4194	0.23	0.123	0.1693
手法 1	0.4362	0.23	0.1304	0.1822
手法 2	0.4401	**0.24**	0.132	0.1802
手法 3	**0.458**	0.23	**0.1373**	0.1973
手法 4	0.4579	0.23	0.1372	**0.1974**

表 4.2: ベースラインが BM25 の場合の結果

	infNDCG	P@10	infAP	R-Precision
BM25	0.4216	0.23	0.1264	0.1712
手法 1	0.442	0.27	0.1346	0.2012
手法 2	0.4536	0.28	0.1354	0.2031
手法 3	**0.4705**	**0.34**	**0.1371**	**0.2221**
手法 4	0.4704	**0.34**	**0.1371**	0.2208

表 4.1 をみると、ベースラインとして用いた TFIDF は、P@10 以外の全ての評価指標で 4 つの重み付け手法よりも低い精度であった。医学重要度と単語特定性を組み合わせた手法 3 と 4 は、P@10 以外の評価指標全てで他の重み付け手法（TFIDF、手法 1、手法 2）よりも良い結果となった。手法 1 と 2 を比較すると、infNDCG と P@10、infAP の場合で手法 2 の方が手法 1 よりも良く、R-Precision の場合は手法 1 の方が手法 2 よりも良かった。手法 3 と 4 を比較すると、P@10 の評価指標が同じで、それ以外の評価指標もほとんど差がみられなかった。評価指標が infNDCG の場合は手法 3 が最も高い精度で、P@10 の場合は手法 2 が最も高く、infAP の場合は手法 3 が最も高く、R-Precision の場合は手法 4 が最も高い精度を示した。

表 4.2 をみると、ベースラインの BM25 は全ての評価指標で 4 つの重み付け手法よりも低い精度であった。手法 1 と 2 を比較すると、全ての評価指標で手法 2 の方が手法 1 よりも良い結果であった。一方、手法 3 と 4 では、P@10 と infAP の評価指標が同じで、infNDCG の評価指標もほとんど差がみられなかった。R-Precision の場合は、手法 3 が手法 4 よりも高い精度であった。

医学重要度と単語特定性の 2 つの要素を組み合わせた手法 3 と 4 では、2 つのベースラインでの実験でどちらも TREC 2014 CDS Track の主要な評価指標である infNDCG で最

も高い精度を示した。

ベースラインごとの比較では、infAP の評価指標の場合は差がほとんどみられなかったが、それ以外の全ての評価指標では BM25 が良い精度を示す傾向にあった。

4.4 考察

本研究の実験結果からわかったことは以下の 2 つである。

1: 医学重要度と単語特定性を考慮した 2 つの手法は、全ての評価指標で高い精度を示したことから、その有効性が示された。しかし、医学重要度で扱う単語タイプ数の違いは明確に現れなかった。

2: 医学重要度と単語特定性を考慮した 2 つの手法は、特に、BM25 を組み合わせた場合に検索結果の上位 10 までに関連文書を高い精度で返すことができる。

医学重要度という要素に関して、手法 1 は一般単語（非医学単語）よりも医学単語が情報検索において重要であると考え、医学単語の重みを一律で高くした手法であった。手法 2 は、さらに診断において重要な要素の一部だと考えられる疾病や症状、所見を取り扱う ICD-10 コードを持つ医学単語により高い重みを与えた。この手法 2 は手法 1 と比較して全体的に良い結果であった。したがって、医学重要度として医学単語の重みを ICD-10 コードの有無により 2 段階で変化させることが、医学論文を用いた診断に関連する情報検索で有効であると考えられる。手法 3 と 4 と比較して、手法 2 の有効性は、ベースラインを TFIDF とした時に、検索結果の上位 10 の文書にクエリーと関連した文書を高い精度で返すことができるという点であった。

医学重要度だけでなく、共起の偏りという観点から医学単語の特定性も考慮した手法 3 と 4 では、どちらも TREC 2014 CDS Track における主要な評価指標 infNDCG において、2 つのベースラインで高い方の精度と比べて良い結果であった。infNDCG は検索対象文書の関連度の程度を考慮するため、手法 3 と 4 は関連度がより高い文章を上位の結果で返す

ことができると考えられる。また、手法3と4は評価指標全てにおいて、2つのベースラインで高い方の精度よりも良い結果を示した。したがって、診断に関連する医学論文検索において、医学重要度と単語特定性を考慮した手法は有効性が高いと考えられる。特に、ベースラインをBM25とした時に、P@10の評価指標の差が、手法3及び手法4とベースラインとの間で百分率で11%であった。したがって、手法3と4は、特に関連文書を検索結果の上位10までに高い精度で返すのに有効であると考えられる。しかし、医学重要度で扱う単語のタイプ数の違いで分けた手法3と4を比較するとほとんど差がみられなかった。よって、医学重要度と単語特定性の組み合わせ方についてさらなる実験が必要だと考えられる。

4.5 結語

本研究は医学オントロジーのUMLSと単語のタイプやトピックといったクラス情報の組み合わせにより、新たに医学重要度と単語特定性という要素を求め、それら2つの要素の組み合わせからなる4つの重み付け手法を提案した。

医学重要度は診断に関する情報検索における単語の医学的な重要度のことである。本研究ではUMLSとICD-10を活用して、単語を3タイプに分け、タイプの違いに基づいて医学重要度を求めた。さらに、医学単語の共起の偏りという観点から、単語のトピック情報をもとにエントロピー計算で単語特定性を求めた。

本研究ではTREC 2014 CDS Trackの診断に関するクエリーを用いて医学論文検索の実験を行った。結果の分析から、それぞれの手法の特色が捉えられた。医学重要度と単語特定性の2つの要素を考慮した手法は、全ての評価指標でTFIDFとBM25のベースラインよりも高い精度を示した。特に、TREC 2014 CDS Trackの主要な評価指標であるinfNDCGで高い精度を示した。さらに、P@10の評価指標において、BM25のベースラインとの差が百分率で11%あった。このような結果から、提案手法は診断用の医学論文検索システムに適用できると考えられる。特に関連論文を検索結果の上位10までに高い精度

で返すのに有効であると考えられる。

しかしながら、医学重要度と単語特定性の 2 つの要素を考慮した手法において、医学重要度で扱う単語のタイプ数の違いによる 2 つの手法の差がほとんどみられなかった。したがって、今後の課題として、医学論文を用いた診断に関する情報検索における重要な単語は何かを捉えるために結果のさらなる分析を行う。また、医学重要度と単語特定性の組み合わせ方や、重み係数 α_1、α_2 の影響度に関して、さらなる実験をしながら再考する。LDA のトピックの数に関して最適なトピック数を探索することも今後の課題とする。

本章では、医学単語の特定性として共起の偏りの視点による百科事典的意味を学習により捉え、その意味を考慮した単語の重み付け手法を開発し、情報検索の実験でその有効性を示した。したがって、本論文の目的であるフレーム意味論の考え方を用いて、目的に適した意味を考慮した医学単語の重み付け手法の開発が達成されたと考えられる。特に、理論的貢献では、機械学習手法の LDA により得られる単語のトピック情報を活用して、共起の偏りの視点から単語の重みを連続値で捉え、特徴的な単語と一般的な単語という共起の偏り度合いを捉えた点であると考えられる。それによって、様々な意味の視点を計算機でどのように表現するかという理論的な課題に対して、単語の重み付けで計算機に人が想起する単語の様々な意味の視点を理解させる新たな医学知識表現を提案し、その課題の解決策を示した。

第5章　研究2（予測のための意味を考慮した単語の重み付け）

5.1　目的

本研究は医学オントロジーであるUMLSを活用しながら、患者の重症度という観点から医学単語の重要度を捉える。そのため、本研究ではUMLSから得られるマッピング情報とICD-10の階層構造、死因ランキングによる医学単語間の線形関係を組み合わせることで、患者の重症度を考慮した医学単語の重み付け手法を提案する。

5.2　提案手法

提案手法は電子カルテ内の単語に対して意味的な重要度を付与する。それぞれの単語はTFIDFによる重みと医学重要度の重みの組み合わせによる重みが付与される。本研究では、電子カルテ内の単語の医学重要度を患者状態の重症度とする。実現可能性の観点から電子カルテ内の全ての単語に対して、全て異なる医学重要度の重みを付与することは困難である。したがって、本研究のキーアイデアは、単語を階層的にカテゴリー分けをし、同じカテゴリー内の単語は同じ患者状態の重症度の重みを持つとする。また、本研究では特に疾病の重症度に関連のある単語に着目する。そのため、死因ランキング[27]の医学知識を用いて、電子カルテ内の単語を疾病の重症度に関する医学重要度をもとに18のカテゴリーに分ける。そして、ランキング内のランク情報を活用しながら、カテゴリーごとで医学重要度の程度を決定し、最後に医学重要度の重みとTFIDFの重みを組み合わせることで提案手法の重みを求める。図5.2のように、提案手法は3つのステップにより構成され

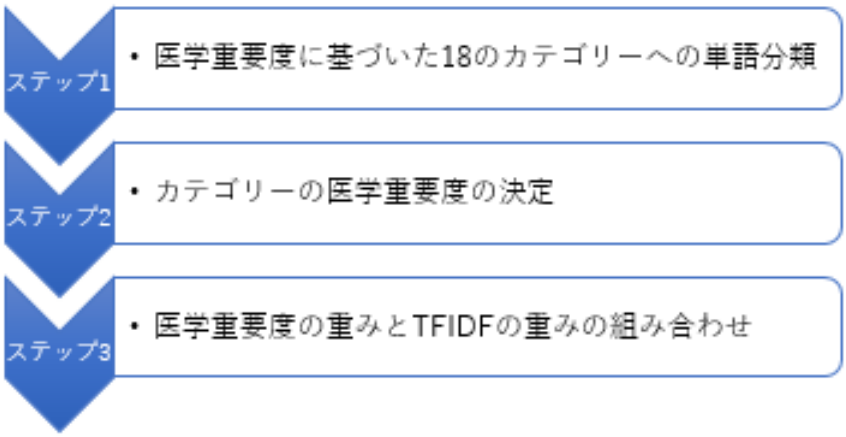


図 5.1: 研究 2 の提案手法の 3 つのステップ

る。以下に 3 つのステップの詳細をそれぞれ述べる。

5.2.1 医学重要度に基づいた 18 のカテゴリーへの単語分類

本論文の共通する重み付けフレームワークにより識別した 3 つの単語タイプ（非医学単語、非 ICD-10 単語、ICD-10 単語）を土台に、死因ランキング内のランク単語を識別する。死因ランキングは 15 の死因名とそれに対応する ICD-10 コードが含まれている。したがって、ICD-10 単語に対して、ICD-10 コードを参照することで、死因ランキング内のランクへ紐づけることができる。それにより、ICD-10 単語に対して、重症度を考慮した医学重要度の程度を実数で付与できるようになる。ここで、ランクに対応する ICD-10 単語は ICD-10 ランク単語とし、対応しない場合は ICD-10 ランク外単語とする。

本研究では、非医学単語、非 ICD-10 単語、ICD-10 ランク外単語はそれぞれ C_1、C_2、C_3 のカテゴリーへ属すこととする。ICD-10 ランク単語は対応するランクに基づいて C_4、C_5、...、C_{18} のいずれかのカテゴリーに基本的に属す。カテゴリー番号が高ければ高いほど、患者状態の重症度に関する医学重要度の重みは高くなる。したがって、C_{18} に属する単語が最も高い医学重要度を保有し、それらは死因ランキング内のランク 1 に該当する。以下はカテゴリー間の包含関係を示している。図 5.2 は 18 のカテゴリーを階層関係で表したものである。

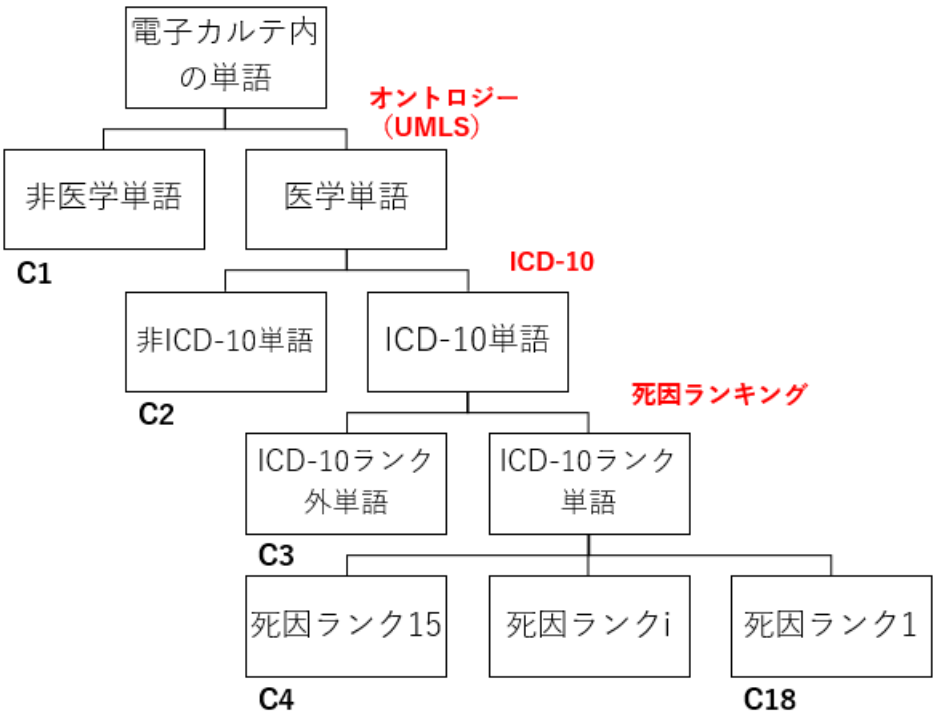


図 5.2: 医学重要度に基づいた 18 のカテゴリー分類

$$\{\text{医学単語}\} \cup \{\text{非医学単語}\} = \{\text{電子カルテ内の単語}\}$$

$$\{ICD-10\text{単語}\} \cup \{\text{非}\,ICD-10\text{単語}\} = \{\text{医学単語}\}$$

$$\{ICD-10\text{ランク単語}\} \cup \{ICD-10\text{ランク外単語}\} = \{ICD-10\text{単語}\}$$

以下は電子カルテ内の 3 つのセンテンスの例である。図 5.3 は、その電子カルテ内の単語を医学重要度に基づいた 18 のカテゴリーへ分類するプロセスを示している。

1. Mrs. [**Known patient lastname 4483**] is an 81 year old female with congestive heart failure. She has been medically managed but has gradually experienced worsening symptoms of dyspnea on exertion and paroxysmal nocturnal dyspnea.

2. She did have that one episode of shortness of breath which was most likely due to acute pulmonary edema.

3. As the patient has risk factors of diabetes mellitus, hypertension, and hypercholesterolemia and possible old inferior myocardial infarction on electrocardiogram it was felt that ischemia was the likely cause of her conduction system abnormalities.

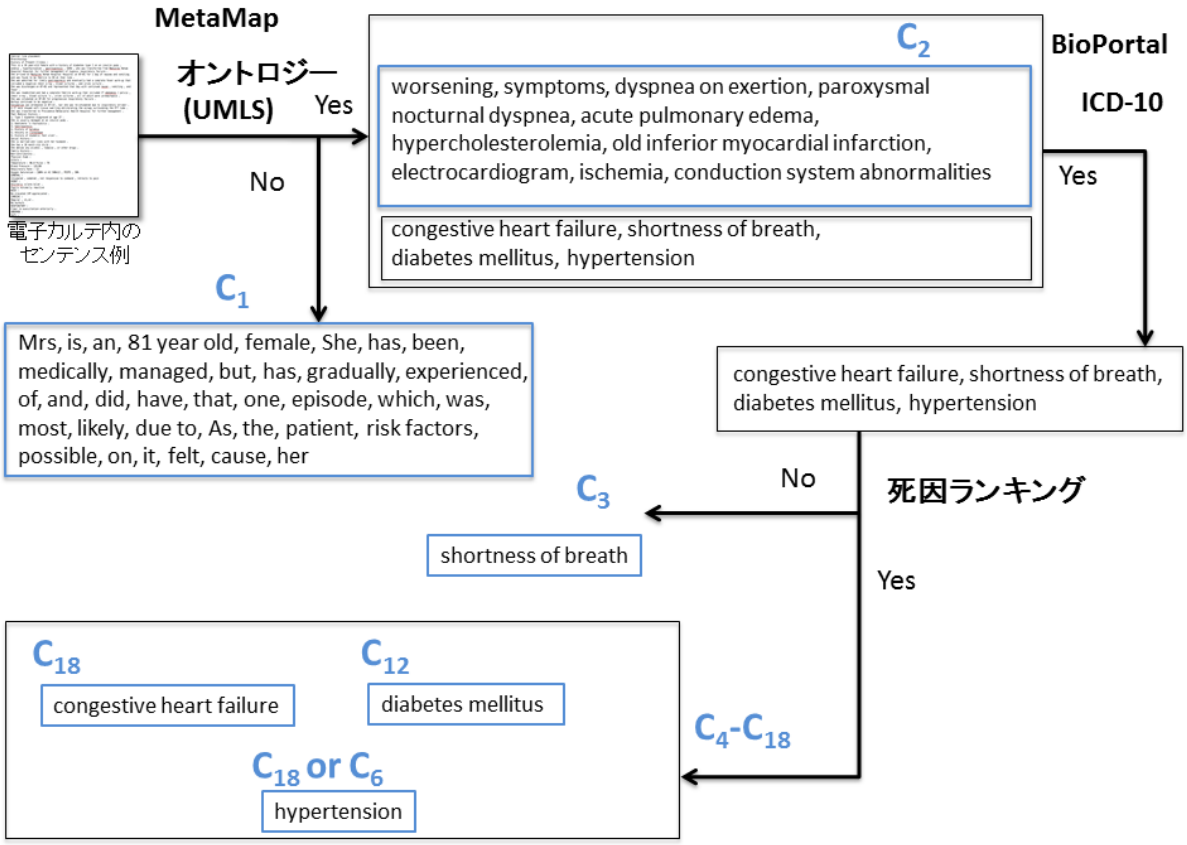


図 5.3: 電子カルテ内の単語の分類例

例えば、female や episode の単語は MetaMap により非医学単語としてカテゴリー C_1 へ分類される。一方、paroxysmal nocturnal dyspnea や hypercholesterolemia の単語は、医学単語ではあるが ICD-10 に対応していないため、非 ICD-10 単語としてカテゴリー C_2 へ分類される。shortness of breath の単語は ICD-10 コードを持つ医学単語であるが、死因ランキング内のランクに対応していないため、ICD-10 ランク外単語としてカテゴリー C_3 へ分類される。congestive heart failure の単語は I50 の ICD-10 コードに対応し、そのコードは死因ランキング内のランク 1 に対応していることから、ICD-10 ランク単語としてカテゴリー C_{18} へ分類される。diabetes mellitus は、E10-E14.9 の ICD-10 コードを持ち、そのコードは死因ランキング内のランク 7 に該当するため、ICD-10 ランク単語としてカテゴリー C_{12} へ分類される。一方で、hypertension は、I10-I15.9 の ICD-10 コードを持ち、そのコードはランク 1（カテゴリー C_{18}）とランク 13（カテゴリー C_6）のどちらにも該当し得る。ICD-10 ランク単語の場合、hypertension のように特定のカテゴリーに分類できない単語も存在する。

5.2.2 カテゴリーの医学重要度の決定

本ステップでは、死因ランキングのランクの活用により医学重要度に基づいて分類した 18 のカテゴリー内の単語に対して、重症度を考慮した医学重要度の程度を適切に与える。ここで、$w(C_i)$ はカテゴリー C_i の医学重要度の重みの値とする。$w(C_i)$ の値は以下の線形関係の制約に従う。

$$w(C_1) < w(C_2) < ... < w(C_{18})$$

上記の制約に基づいて $w(C_i)$ の値は以下のようにカテゴリー間で同じ間隔で割り当てられる。

$$w(C_i) = w(C_{i+1}) - \Delta$$

ここで、Δ は 18 のカテゴリー間における医学重要度の程度の差を変化させるパラメーターである。本研究では、パラメータ Δ の値は 0.04、0.03、0.02、0.01 の 4 つのパターン

を用いる。カテゴリーC_1は医学単語が含まれていないため、その医学重要度の重み$w(C_1)$はゼロとする。表5.1はそれぞれのカテゴリーの医学重要度の重みと対応する死因名及びICD-10コードである。パラメータΔによってそれらのカテゴリーの重みは4つのパターンで表記される。

I10-I15.9のICD-10コードを持つhypertensionのように、複数のランクに該当する単語に対しては、該当する複数のランクの重みの平均値を付与する。

5.2.2.1 ICD-10ランク単語の医学重要度の伝搬

ICD-10は階層構造であるため、その特徴を活用して死因ランキングのランクに対応するICD-10ランク単語の医学重要度の重みをその下位に位置するICD-10ランク単語へ伝搬させる。例えば、死因ランキング内のランク2はICD-10コードC00-C97の範囲に該当するが、その際、ICD-10コードC00-C97に対応する単語（例えばC00のコードを持つ単語）だけでなく、それらのICD-10コードの下位に属する単語（例えばC00.0やC00.9のコードを持つ単語）に対しても医学重要度として同じ重みを付与する。このように、ICD-10の階層構造から階層内の単語の意味的な類似性を考慮することにより、より多くの医学単語に対して患者の重症度を考慮した重み付けをすることが可能になる。

5.2.2.2 医学重要度におけるICD-10単語間の関係性の例

例えば、coronary artery disease、secondary hypertension、peripheral vascular diseaseは、I25.1とI15、I73.9のICD-10コードにそれぞれ対応している。これらのICD-10コードは全て循環器系の疾患（I00-I99）のクラスに属しているが、提案手法により、上記3つのICD-10コードは患者の重症度という観点から以下のように区別される。

$$I73.9\left(w\left(C_3\right)\right) < I15\left(w\left(C_6\right)\right) < I25.1\left(w\left(C_{18}\right)\right)$$

I73.9のICD-10コードを持つ単語は、ICD-10ランク外単語となるためカテゴリーC_3に属し、I15とI25.1のICD-10コードを持つ単語は、ICD-10ランク単語としてそれぞれC_6

表 5.1: 死因ランキングに基づいた患者の重症度を考慮した医学重要度

ランク	死因名	ICD-10 コード	重み (Δ=0.04)	重み (Δ=0.03)	重み (Δ=0.02)	重み (Δ=0.01)	カテゴリー
1	Disease of heart	I00-I09, I11, I13, I20-I51	0.7	0.7	0.7	0.7	C_{18}
2	Malignant neoplasms	C00-C97	0.66	0.67	0.68	0.69	C_{17}
3	Chronic lower respiratory diseases	J40-J47	0.62	0.64	0.66	0.68	C_{16}
4	Cerebrovascular diseases	I60-I69	0.58	0.61	0.64	0.67	C_{15}
5	Accidents (unintentional injuries)	V01-X59, Y85-Y86	0.54	0.58	0.62	0.66	C_{14}
6	Alzheimer's disease	G30	0.5	0.55	0.6	0.65	C_{13}
7	Diabetes mellitus	E10-E14	0.46	0.52	0.58	0.64	C_{12}
8	Nephritis, nephritic syndrome and nephrosis	N00-N07, N17-N19, N25-N27	0.42	0.49	0.56	0.63	C_{11}
9	In uenza and pneumonia	J09-J18	0.38	0.46	0.54	0.62	C_{10}
10	Intentional self-harm (suicide)	U03, X60-X84, Y87.0	0.34	0.43	0.52	0.61	C_9
11	Septicemia	A40-A41	0.3	0.4	0.5	0.6	C_8
12	Chronic liver disease and cirrhosis	K70, K73-K74	0.26	0.37	0.48	0.59	C_7
13	Essential hypertension and hypertensive renal disease	I10, I12, I15	0.22	0.34	0.46	0.58	C_6
14	Parkinson's disease	G20-G21	0.18	0.31	0.44	0.57	C_5
15	Pneumonitis due to solids and liquids	J69	0.14	0.28	0.42	0.56	C_4
16	ICD-10 ランク外単語	なし	0.1	0.25	0.4	0.55	C_3
17	非 ICD-10 単語	なし	0.06	0.22	0.38	0.54	C_2
18	非医学単語	なし	0	0	0	0	C_1

表 5.2: 係数の 7 つの組み合わせ

ケース（α_1,α_2）	α_1	α_2
ケース 1（1, 1）	1	1
ケース 2（1, 0.75）	1	0.75
ケース 3（0.75, 1）	0.75	1
ケース 4（1, 0.5）	1	0.5
ケース 5（0.5, 1）	0.5	1
ケース 6（1, 0.25）	1	0.25
ケース 7（0.25, 1）	0.25	1

と C_{18} に属す。死因ランキング内のランク 1 に該当する C_{18} に属す I25.1 のコードを持つ単語が、患者の重症度という観点から上記 3 つの ICD-10 コードの中で最も高い重要度を持つことになる。このように、提案手法は死因ランキングと ICD-10 を組み合わせることで、ICD-10 単語間を患者の重症度に基づく医学重要度の観点から区別することが可能になる。

5.2.3 医学重要度の重みと TFIDF の重みの組み合わせ

医学重要度の重み w_m は TFIDF の重み w_0 と以下のように組み合わされる。

$$w = \alpha_1 \times w_0 + \alpha_2 \times w_m$$

ここで、α_1 と α_2 は TFIDF の重みと医学重要度の重みの係数であり、2 つの重みのバランスを取るために用いる。本研究では、表 5.2 のように 7 つのケースの係数の組み合わせを用いる。

医学重要度の重みと TFIDF の重みの組み合わせは、ストップワードの除去やチャンキング、否定語に係る単語の除去といった前処理の後に行う。

5.3 実験による評価

5.3.1 実験設定

本研究の提案手法の有効性を検証するため、TFIDF による重み付けをベースラインとして、提案手法による意味を考慮した重み付け手法を比較する。また、提案手法における適切な 2 つのパラメータ（Δ と α）の値を解明することを実験による評価の目的とする。

本研究では、患者の重症度に基づく提案手法の重み付けの評価のために死亡予測の実験を行う。実験に用いるデータは MIMIC II データベース [39] から 60 歳以上の患者の退院時要約の電子カルテで合計 13,026 文書を用いる。患者は病院で亡くなった患者と亡くならなかった患者の 2 つのグループに属す。これは、MIMIC II データベースの expire_flg のカラムから識別する。死亡予測ではこの患者の 2 つのグループのラベルを用いる。それぞれのラベルに対応する文書数は 2,158、10,868 である。

予測のための入力データでは 20 以上の文書で出現した単語のみを扱い、その単語数は 11,858 である。表 5.3 は 18 のカテゴリー内の単語数と単語の文書頻度を示している。

本実験において、パラメータ Δ の変化による 4 パターンの提案手法は、TFIDF + MED (Δ=0.04)、TFIDF + MED (Δ=0.03)、TFIDF + MED (Δ=0.02)、TFIDF + MED (Δ=0.01) と表記される。これらはパラメータ α を変えながらベースラインと比較される。

本実験では、SVM (rbf)、SVM (linear)、Naive Bayes、Random forest、Decision tree の 5 つの分類器を用いる。これらは単語間の相関を考慮する L2 正則化による特徴選択後に実行される。分類器及び L2 正則化のパラメータはデフォルトとする。これらは Scikit-learn[31] のライブラリを用いて実行する。

評価はラベルごとの比率を保つ 5-fold stratified cross validation により 500 の電子カルテを用いて F1 スコアを求めることにより行う。F1 スコアは以下の式で計算される。

$$F1\ スコア = \frac{2 \times precision \times recall}{precision + recall}$$

表 5.3: カテゴリーごとの単語数と単語の文書頻度

カテゴリー	単語の文書頻度の合計	単語の文書頻度の平均	単語の文書頻度の割合	単語数
C_1	4222157	324.133	0.840461229	102068
C_2	675397	51.8499	0.134444312	12418
C_3	74904	5.7503	0.014910366	1507
C_4	0	0	0	0
C_5	1238	0.095	0.000246436	13
C_6	10304	0.791	0.002051111	24
C_7	389	0.03	7.74342E-05	17
C_8	1340	0.103	0.00026674	19
C_9	106	0.01	2.11003E-05	1
C_{10}	2633	0.202	0.000524124	39
C_{11}	2571	0.197	0.000511782	20
C_{12}	4699	0.361	0.000935381	37
C_{13}	211	0.02	4.20016E-05	12
C_{14}	0	0	0	0
C_{15}	3053	0.234	0.000607729	30
C_{16}	3317	0.255	0.000660281	23
C_{17}	780	0.06	0.000155267	2
C_{18}	20520	1.5753	0.004084705	239

本実験では、データセットをランダムで選択しながら、5-fold stratified cross validation を 100 回繰り返すことによる平均の F1 スコアにより手法ごとの比較を行う。毎回の cross validation における訓練セットとテストセットの割合はそれぞれ 80%と 20%である。

5.3.2 実験結果

5 つの表 5.4 - 5.8 が分類器ごと結果である。それぞれの表は、ベースラインとパラメータ Δ の変化による 4 パターンの提案手法の 5 つの手法の F1 スコアの値が、パラメータ α の 7 つの組み合わせごとで示されている。

表 5.4: SVM (rbf) による結果

ケース (α_1, α_2)	TFIDF	TFIDF + MED ($\Delta = 0.04$)	TFIDF + MED ($\Delta = 0.03$)	TFIDF + MED ($\Delta = 0.02$)	TFIDF + MED ($\Delta = 0.01$)
(1, 0.25)	0.7849	**0.6754**	**0.7838**	**0.8148**	**0.8197**
(1, 0.5)	0.7849	0.6464	0.7737	0.8076	0.8148
(1, 0.75)	0.7849	0.6357	0.7721	0.8057	0.8137
(1, 1)	0.7849	0.6298	0.7692	0.8051	0.814
(0.75, 1)	0.7849	0.6294	0.7681	0.8045	0.8132
(0.5, 1)	0.7849	0.63	0.7674	0.804	0.8122
(0.25, 1)	0.7849	0.633	0.7667	0.8033	0.8118

表 5.5: SVM (linear) による結果

ケース (α_1, α_2)	TFIDF	TFIDF + MED ($\Delta = 0.04$)	TFIDF + MED ($\Delta = 0.03$)	TFIDF + MED ($\Delta = 0.02$)	TFIDF + MED ($\Delta = 0.01$)
(1, 0.25)	0.785	0.7044	0.8468	**0.8633**	**0.8649**
(1, 0.5)	0.785	0.7293	**0.8605**	0.8572	0.8498
(1, 0.75)	0.785	0.7724	0.8543	0.8473	0.8384
(1, 1)	0.785	**0.8011**	0.848	0.8385	0.8312
(0.75, 1)	0.785	0.7911	0.8463	0.8376	0.8297
(0.5, 1)	0.785	0.7829	0.8451	0.8366	0.8286
(0.25, 1)	0.785	0.7753	0.8429	0.8355	0.8277

表 5.6: Naive Bayes による結果

ケース (α_1, α_2)	TFIDF	TFIDF + MED ($\Delta = 0.04$)	TFIDF + MED ($\Delta = 0.03$)	TFIDF + MED ($\Delta = 0.02$)	TFIDF + MED ($\Delta = 0.01$)
(1, 0.25)	0.8799	**0.8057**	**0.8431**	0.85	**0.8528**
(1, 0.5)	0.8799	0.7802	0.8423	**0.8512**	0.8522
(1, 0.75)	0.8799	0.7722	0.8419	0.8498	0.8513
(1, 1)	0.8799	0.7711	0.8406	0.8479	0.8496
(0.75, 1)	0.8799	0.7633	0.8387	0.8471	0.8489
(0.5, 1)	0.8799	0.7543	0.8377	0.8468	0.8486
(0.25, 1)	0.8799	0.7461	0.8367	0.8463	0.8481

表 5.7: Random forest による結果

ケース (α_1, α_2)	TFIDF	TFIDF + MED ($\Delta = 0.04$)	TFIDF + MED ($\Delta = 0.03$)	TFIDF + MED ($\Delta = 0.02$)	TFIDF + MED ($\Delta = 0.01$)
(1, 0.25)	0.8476	0.8508	0.8556	0.8598	0.8596
(1, 0.5)	0.8476	0.8527	0.8602	**0.8616**	0.8592
(1, 0.75)	0.8476	0.8554	0.8586	0.8614	**0.8596**
(1, 1)	0.8476	0.856	0.86	0.8582	0.8556
(0.75, 1)	0.8476	0.8584	0.8611	0.8569	0.8557
(0.5, 1)	0.8476	0.8613	**0.8624**	0.8547	0.8541
(0.25, 1)	0.8476	**0.8654**	0.8585	0.8489	0.8401

表 5.8: Decision tree による結果

ケース (α_1, α_2)	TFIDF	TFIDF + MED ($\Delta = 0.04$)	TFIDF + MED ($\Delta = 0.03$)	TFIDF + MED ($\Delta = 0.02$)	TFIDF + MED ($\Delta = 0.01$)
(1, 0.25)	0.8266	0.8275	0.8304	0.8309	0.8313
(1, 0.5)	0.8266	0.8292	0.8308	**0.8325**	0.8306
(1, 0.75)	0.8266	0.8296	0.8305	0.8314	0.8306
(1, 1)	0.8266	0.8307	0.8325	0.8317	0.829
(0.75, 1)	0.8266	0.8289	0.8316	0.8310	**0.8314**
(0.5, 1)	0.8266	0.8317	0.8342	0.8322	0.8307
(0.25, 1)	0.8266	**0.8349**	**0.8347**	0.83	0.8293

実験結果によるとパラメータを変化させた提案手法は、概して TFIDF による重み付け手法よりも F1 スコアの値が高かった。特に、Naive Bayes 以外の全ての分類器を用いた場合で TFIDF による手法よりも高いスコアであった。

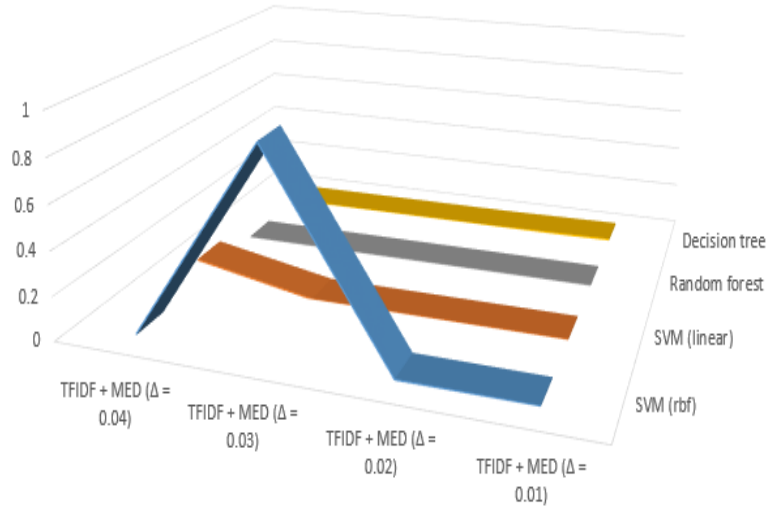
パラメータ Δ については、概して Δ の値が小さければ小さいほど予測性能が良かった。SVM (rbf) と SVM (linear)、Naive Bayes の分類器を用いた場合において、パラメータ Δ が最小値の時に提案手法の F1 スコアが最も高かった。TFIDF + MED (Δ=0.04) と TFIDF + MED (Δ=0.01) を比較するとスコアに大きな差がみられた。例えば、SVM (rbf) を用いてパラメータ α の最も良い組み合わせの場合でその差はおよそ 14%であった。一方、分類器が Random forest と Decision tree の場合はスコアの差がみられなかった。

パラメータ α については、α_1 と α_2 の値を変化させることは予測性能に大きな変化をもたらさなかった。しかしながら、α_1 の値が 1 で α_2 の値が 0.25 または 0.5 の時で、かつ、パラメータ Δ の値が小さい場合で、概して他の組み合わせのケースと比べて良い結果を示した。

5.4 考察

死亡予測の実験結果から、提案する患者の重症度を考慮した単語の重み付け手法が概して TFIDF による重み付けよりも良い結果を示していた。したがって、提案手法は死亡予測のタスクに適した手法であることが示唆される。

パラメータ Δ において、その値が小さいことは医学重要度の重みを強調していることを意味するため、パラメータ Δ による医学重要度の重みの影響を際立たせることが高い予測性能に寄与していたと考えられる。そして、提案手法は重み係数であるパラメータ α によらず、パラメータ Δ に依存する傾向が高いことがわかった。



	TFIDF + MED (Δ = 0.04)	TFIDF + MED (Δ = 0.03)	TFIDF + MED (Δ = 0.02)	TFIDF + MED (Δ = 0.01)
SVM (rbf)	7.26382E-17	0.907985255	0.003407563	0.000703102
SVM (linear)	0.106082852	6.31888E-12	1.82058E-12	9.38799E-13
Random forest	5.33081E-22	3.0162E-15	1.16808E-12	1.8016E-11
Decision tree	2.91101E-06	0.000132074	0.000259545	0.007487979

図 5.4: 統計検定の結果

5.4.1 統計検定

提案手法が TFIDF による手法よりも有意で良い結果をもたらしたかどうかを調べるために T 検定を行う。その際、全ての重み付け手法の F1 スコアは同じ実験設定のもとで導かれているものとする。

図 5.4 は TFIDF による手法とそれぞれのパラメータ Δ ごとで最も良い α の組み合わせのケースを用いた場合の提案手法との p 値である。ここでは、2 つのパラメータを変更しながら提案手法が TFIDF による重み付けと比べて高いスコアを出した際の 4 つの分類器が示されている。図 5.4 内の上のグラフの x 軸はパラメータ Δ の 4 つのパターンごとの手法で、y 軸は 4 つの分類器、z 軸は p 値で目盛りの最小値が 0 で最大値が 1 である。これらの軸を表で示した結果が図 5.4 内の下の表となる。

検定結果をみると、概して提案手法が良い場合の p 値が 0.05 未満であったことから、提案手法の高い予測性能が TFIDF による重み付けよりも有意であることがわかった。一方、パラメータ Δ が 0.03 で SVM (rbf) が分類器として用いられた場合は、F1 スコアにほとん

ど差がみられなかったことから、

その条件における p 値は 0.9 以上であった。提案手法の F1 スコアが TFIDF による重み付けよりも良かった場合でも、パラメータ Δ が 0.04 で SVM (linear) が分類器として用いられた場合の p 値が約 0.1 であった。このような検定結果から、TFIDF による重み付けと比べて提案手法による改善があまりみられない場合は、提案手法による性能が有意であるとはいえず、TFIDF による性能と同程度のものであることが示唆される。

5.5 結論

本研究は患者の重症度に基づいて電子カルテに対する新たな意味を考慮した単語の重み付け手法を提案した。患者の重症度の観点から単語の重要度を捉えるために、医学オントロジーである UMLS の活用による単語のマッピング情報や、ICD-10 を活用した医学単語の階層関係、死因ランキングの活用による医学単語の線形関係といった様々な種類の単語の関係性を効果的に組み合わせた。

提案する意味を考慮した単語の重み付け手法の有効性の検証、及び提案手法で用いた 2 つのパラメータ（Δ と α）の適切な値を解明するために死亡予測の実験を行った。実験結果から、パラメータを変化させた提案手法は概して TFIDF による手法よりも良いスコアを出した。T 検定の結果から、TFIDF による重み付けと比較して提案手法の高い予測性能は有意であることから、統計検定からも提案手法の有効性が示された。

提案手法は UMLS や ICD-10、死因ランキングに基づいて開発されているため、その提案手法により導かれた結果は医学知識に基づいている。

実験結果から、パラメータ Δ の値が小さければ小さいほど、ほとんどの条件で死亡予測の性能が優れていることがわかった。小さい値の意味は医学重要度の影響を強めていることから、提案する意味を考慮した重み付け手法の重要性が示されたと考えられる。パラメータ α（α_1 と α_2）の値の変化は、予測性能に重要な影響をもたらさなかったが、α_1 の値が 1 で α_2 の値が 0.25 または 0.5 の場合で、かつ、パラメータ Δ の値が小さい時に他の

組み合わせのケースよりも概して良いスコアが出されていた。このようなことから、提案手法はパラメータ α によらず、パラメータ Δ に依存することがわかった。

提案手法は特定の疾病や症状のみを対象としていないため、15 の死因ランクと限られているが、統一的に患者の重症度の観点から単語の医学重要度を捉えることができる。この提案手法は電子カルテを用いた患者のリスク予測（例えば、入院日数の予測）や患者の重症度に基づいた類似症例検索に応用できると考えられる。提案手法の重み付けは電子カルテの文書をベクトル空間上で簡易に計算可能な形式に変換できることから、電子カルテが普及する中で、データの二次利用の観点から患者の重症度という意味を保持したまま、異なるシステム間での文書データの共有や統合、さらには計算可能なデータを用いた患者の重症度に基づく様々な文書分析を可能にする点が医学分野における貢献であると考えられる。

提案手法は死因ランキング内で該当するランクが同じ単語に対して、全てに同じ重みを付与する。例えば、癌に関する単語の場合、癌の種類の違いによらず、重症度の観点から癌に関する全ての単語に同じ重みを付与している。したがって、今後の課題は、同じ死因ランク内の単語間を重症度の観点から区別し、重み付けに反映させることである。また、患者状態の重症度の観点からより適切に単語に重みを付与するために、単語間の時間関係を考慮することも今後の課題とする。

本章では、患者の重症度の視点による百科事典的意味を既存の百科事典的知識である死因ランキングをもとに捉え、その意味を考慮した単語の重み付け手法を開発し、死亡予測の実験でその有効性を示した。したがって、本論文の目的であるフレーム意味論の考え方を用いて、目的に適した意味を考慮した医学単語の重み付け手法の開発が達成されたと考えられる。特に、理論的貢献では、百科事典的知識である死因ランキングといったランキングを活用することで、単語集合で構成されるカテゴリーごとに対応するランク情報を活用してカテゴリー間の優劣を求めている点であると考えられる。このアイデアによって、様々な意味の視点を計算機で表現するという理論的な課題に対して、単語の重み付けにおいて計算機に人が想起する単語の様々な意味の視点を理解させる新たな医学知識表現を提

案し、その課題の解決策を示した。

5.6　改良手法1（患者状態の重症度を考慮した独立型と依存型の重み結合による医学単語の重み付け手法）

5.6.1　目的

これまで提案してきた研究2の意味を考慮した重み付け手法は、患者状態の重症度の観点から重みを与える際に出現する文書によらず、それぞれの単語に対して重みを固定していた。患者状態の重症度をより正確に捉えるためには、患者の重症度を医学単語単体のみで捉えるだけではなく、単語が出現する文書の文脈により変動し得る患者状態の重症度の影響も考慮する必要があると考えられる。そこで、本研究では研究2の手法から得られる医学単語の独立型の重みに加えて、文書の文脈から患者の状態を考慮した依存型の重みを捉え、TFIDF の重みと独立型及び依存型の重みを組み合わせた手法を提案する。

5.6.2　改良方法

本研究は、依存型の重みにおける文脈内の患者状態をチャールソン併存疾患指数に基づいた危険な疾患の組み合わせとする。依存型の重みは、疾患の組み合わせという患者状態への依存度と、疾患の組み合わせの危険度をもとに計算する。依存度は Word2Vec により、ある医学単語と危険な疾患の組み合わせ内の医学単語との類似度から求め、組み合わせの危険度はチャールソン併存疾患指数から得られる患者のリスクスコアで捉える。

患者状態の重症度を考慮した重みを医学重要度として、その独立型と依存型の重みを TFDIF と組み合わせる提案手法は図 5.5 の 4 つのステップで構成される。

ステップ1の単語タイプの識別は、本論文で示している共通の重み付けフレームワークにより行われる。

ステップ2の独立型の医学重要度の重み計算は、研究2のように4つのパターンでパラメータ Δ を変化させながら重みを求める。

ステップ3の依存型の医学重要度の重み計算は、死因ランキングを活用した独立型の重

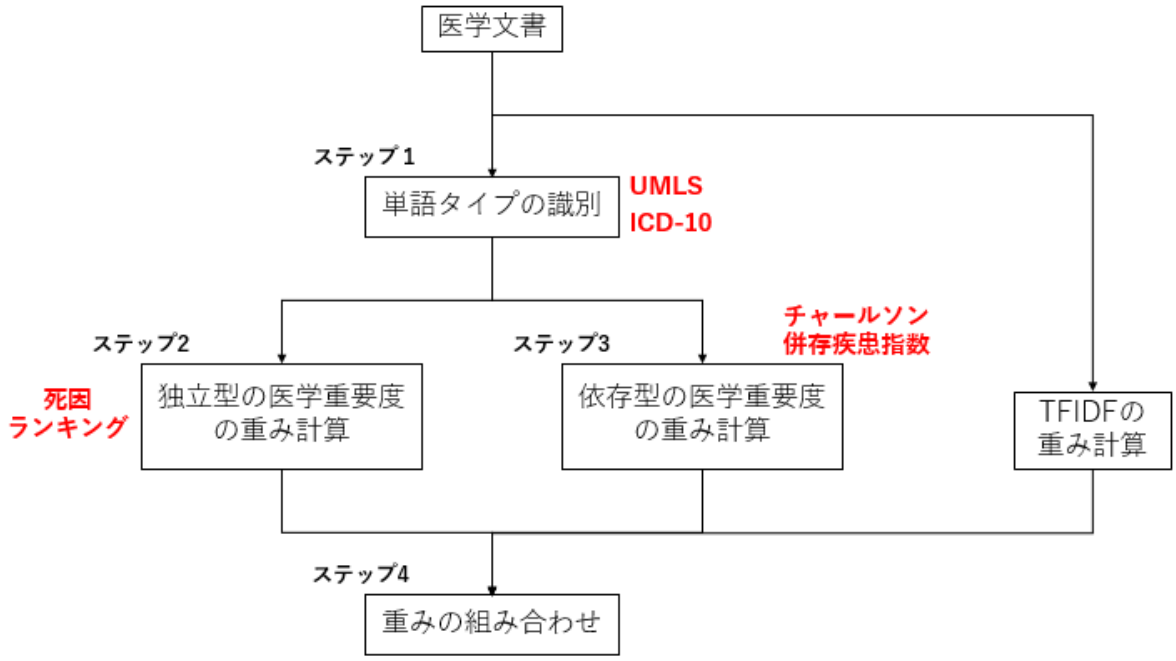


図 5.5: 改良手法 1 のフレームワーク

みに加えて、個々の患者の状態に依存する医学重要度の重みを求める。本研究では、患者状態を危険な疾患の組み合わせとする。なぜなら、高齢患者の場合、複数の疾患を持つことが多く、その疾患の組み合わせによって患者状態に与える悪影響の度合いが変わると考えられるからである。疾患の組み合わせはチャールソン併存疾患指数 [5] を用いて捉える。この指数は、疾患名とそれに対応するリスクスコアが付与されている。さらに、疾患名が図 5.6 のように ICD-10 コードとも対応していることから [44]、これまでに識別した ICD-10 単語を活用することで、疾患の組み合わせを捉える。

2 以上の疾患が組み合わされているパターンを抽出した後、その中から組み合わせが実際に危険かどうかの基準を設けるために、文書ごとで患者が亡くなったか（ラベル 1）亡くならなかったか（ラベル 0）を示すラベルを用いる。なぜなら、チャールソン併存疾患指数のスコアが高い患者でも、亡くなっていない場合があるからである。本研究ではラベル 1 の割合 R^1 が 0.7 以上の組み合わせを危険な疾患の組み合わせとする。R^1 は以下の式で求められる。

$$R^1 = \frac{F^1}{F^1 + F^0}$$

ここで、F^1 と F^0 はそれぞれラベル 1 とラベル 0 の文書内にある疾患の組み合わせが出

Condition	Weights	Codes	
		ICD-9-CM	ICD-10-AM
Acute myocardial infarction	1	410, 412	I21, I22, I252
Congestive heart failure	1	428	I50
Peripheral vascular disease	1	441, 4439, 7854, V434	I71, I790, I739, R02, Z958, Z959
Cerebral vascular accident	1	430–438	I60, I61, I62, I63, I65, I66,G450, G451, G452, G458, G459, G46, I64, G454, I670, I671, I672, I674, I675, I676, I677 I678, I679, I681, I682, I688, I69
Dementia	1	290	F00, F01, F02, F051
Pulmonary disease	1	490, 491, 492, 493, 494, 495, 496, 500, 501, 502, 503, 504, 505	J40, J41, J42, J44, J43, J45, J46, J47, J67, J44, J60, J61, J62, J63, J66, J64, J65
Connective tissue disorder	1	7100, 7101, 7104, 7140, 7141, 7142, 71481(now 5171), 725	M32, M34, M332, M053, M058, M059, M060, M063, M069, M050, M052, M051, M353
Peptic ulcer	1	531, 532, 533, 534	K25, K26, K27, K28
Liver disease	1	5712, 5714, 5715, 5716	K702, K703, K73, K717, K740, K742, K746, K743, K744, K745
Diabetes	1	25002501, 2502, 2503, 2507	E109, E119, E139, E149, E101, E111, E131, E141, E105, E115, E135, E145
Diabetes complications	2	2504, 2505, 2506	E102, E112, E132, E142 E103, E113, E133, E143 E104, E114, E134, E144
Paraplegia	2	342, 3441	G81 G041, G820, G821, G822
Renal disease	2	582, 5830, 5831, 5832, 5833, 5835, 5836, 5837, 5834, 585586588	N03, N052, N053, N054, N055, N056, N072, N073, N074, N01, N18, N19, N25
Cancer	2	14, 15, 16, 18, 170, 171, 172, 174, 175, 176, 179, 190, 191, 192, 193, 194, 1950, 1951, 1952, 1953, 1954, 1955, 1958, 200, 201, 202, 203, 204, 205, 206, 207, 208	C0, C1, C2, C3, C40, C41, C43, C45, C46, C47, C48, C49, C5, C6, C70, C71, C72, C73, C74, C75, C76, C80, C81, C82, C83, C84, C85, C883, C887, C889, C900, C901, C91, C92, C93, C940, C941, C942, C943, C9451, C947, C95, C96
Metastatic cancer	3	196, 197, 198, 1990, 1991	C77, C78, C79, C80
Severe liver disease	3	5722, 5723, 5724, 5728	K729, K766, K767, K721
HIV	6	042, 043, 044	B20, B21, B22, B23, B24

図 5.6: ICD-10 コードに対応するチャールソン併存疾患指数

表 5.9: 危険な疾患の組み合わせの例

文書頻度	ラベル 1 の割合	チャールソン併存疾患指数のスコア	危険な疾患の組み合わせ要素
15	0.8	5	K72.9, N19
5	0.8	4	I64, I73.9, N19
4	0.75	6	I73.9, K72.9, N19

現した頻度である。表 5.9 は、高い出現頻度の危険な疾患の組み合わせの例である。表 5.9 には文書頻度、ラベル 1 の割合、チャールソン併存疾患指数のスコア、危険な疾患の組み合わせの要素の ICD-10 コードが示されている。最も高い出現頻度の危険な疾患の組み合わせは、K72.9（肝不全、詳細不明）と N19（詳細不明の腎不全）で構成され、チャールソン併存疾患指数のスコアは 5 である。

危険な疾患の組み合わせを患者状態として、依存型の医学重要度の重みを求める際、患者状態への意味関係の観点から医学単語（ICD-10 単語を含む）を 3 タイプに分類する。なぜなら、チャールソン併存疾患指数が対象とする医学単語が限られているため、患者状態となる危険な疾患の組み合わせに含まれていない単語にも、適切に依存型の重みを求めることで、より多くの医学単語を重み付けの対象にできるからである。まず、危険な疾患の

表 5.10: タイプ 1 の医学単語の例

単語名	ICD コード	危険な疾患の組み合わせ要素
Liver failure	K72.9	K72.9, N19
Renal failure	N19	K72.9, N19

組み合わせに含まれている医学単語をタイプ 1 とする。表 5.10 は、タイプ 1 の医学単語の例である。表 5.10 にはタイプ 1 の単語名とその ICD-10 コード、その単語を含む危険な疾患の組み合わせを構成する ICD-10 単語が示されている。

組み合わせに該当しないが、単語を意味のベクトルで表現する Word2Vec[25] を活用した類似度計算で、その組み合わせとの類似度の高い医学単語をタイプ 2 とし、類似度の低い医学単語をタイプ 3 とする。本研究では、タイプ 1 と 2 に属する医学単語に対して依存型の医学重要度の重みを求める。

タイプ 1 の単語は、危険な疾患の組み合わせのチャールソン併存疾患指数のスコアにより以下の式のように依存型の重みが求められる。

$$CCI_j = \sum_{i=1}^{m} s_{ji}$$

$$CCI_j^{'} = \frac{CCI_j - min\,(CCI)}{max\,(CCI) - min\,(CCI)}$$

$$w_{ji}^{1} = CCI_j^{'}$$

ここで、CCI_j は文書 j に危険な組み合わせがある場合の文書 j の患者の危険度である。これは危険な組み合わせ内に含まれる ICD-10 コードのチャールソン併存疾患指数のスコア s_{ji} の合計で導かれる。危険な組み合わせが含まれていない文書 j の CCI_j の値はゼロである。m は文書 j の危険な組み合わせ内の ICD-10 コードの数である。CCI_j は Min-Max 正規化により正規化され、$CCI_j^{'}$ が正規化後の値である。CCI はチャールソン併存疾患指数のスコアのセットである。危険な組み合わせ内にある単語の依存型の医学重要度の重み w_{ji}^{1} は、正規化後の $CCI_j^{'}$ となる。危険な組み合わせ内に同じ ICD-10 コードを持つ単語が複数含まれている場合、チャールソン併存疾患指数のスコアを求める際に 1 つの ICD-10 コードとしてみなす。

タイプ 2 の重みは、患者状態との類似度により以下の式のように依存型の重みが求められる。

$$Sim\left(t_{ji},\, t_j^1\right) = \frac{\sum_{k=1}^{n} cos\left(t_{ji},\, t_{jk}^1\right)}{n}$$

$$\phi_{ji} = \begin{cases} 1\, if\ Sim\left(t_{ji},\, t_j^1\right) \geq \beta \\ 0\ otherwise \end{cases}$$

$$w_{ji}^2 = CCI_j^{'} \times Sim\left(t_{ji},\, t_j^1\right) \times \phi_{ji}$$

ここで、$Sim\left(t_{ji},\, t_j^1\right)$ は文書 j のタイプ 1 ではない単語 t_{ji} と文書 j のタイプ 1 の単語 t_j^1 との間の類似度である。その類似度は、単語 t_{ji} とタイプ 1 のそれぞれの単語 t_{jk}^1 のコサイン類似度の平均で導かれる。n は文書 j のタイプ 1 の単語の数である。類似度の計算の際、それぞれの単語は Word2Vec によりベクトルで表現されている。このコサイン類似度の平均が、単語 t_{ji} の文書 j の危険な組み合わせに対する依存度となる。ϕ_{ji} は、単語 t_{ji} が文書 j 内の危険な疾患の組み合わせ t_j^1 と類似しているかどうかを判断する指示関数である。コサイン類似度の平均が $\beta\,(= 0.5)$ 以上の場合に ϕ_{ji} の値が 1 となり、そうでない場合は 0 となる。よって、タイプ 2 の単語とみなされるのは、コサイン類似度の平均が 0.5 以上の場合であり、そうでない場合は、タイプ 3 の単語となる。タイプ 2 の単語の重み w_{ji}^2 は、正規化後の $CCI_j^{'}$ と危険な組み合わせへの依存度で求められる。ただし、タイプ 2 の単語の依存型の重みは、文書 j 内に危険な組み合わせが含まれている場合に計算がされる。危険な組み合わせ内に同じ ICD-10 コードを持つ単語が複数含まれている場合、危険な組み合わせ内のそれぞれの単語との類似度で最大となる値を選択する。表 5.11 はタイプ 2 と 3 の単語の例である。表 5.11 には単語名と、単語タイプ、危険な疾患の組み合わせの要素の ICD-10 コード、危険な組み合わせへのコサイン類似度の平均、危険な組み合わせ内のそれぞれの要素とのコサイン類似度が示されている。

ステップ 4 の重みの組み合わせでは、以下のように TFIDF による重み w_0 と独立型及

表 5.11: タイプ 2 と 3 の医学単語の例

単語名	単語タイプ	類似度の平均	類似度	危険な疾患の組み合わせ要素
Respiratory failure	Type 2	0.6228	0.6502, 0.5954	K72.9, N19
Hypotension	Type 2	0.7087	0.69, 0.7273	K72.9, N19
Hypertension	Type 3	0.289	0.2771, 0.3008	K72.9, N19
Shortness of breath	Type 3	0.1434	0.1281, 0.1588	K72.9, N19

び依存型の医学重要度による重み w_1、w_2 が組み合わされる。

$$w = \alpha_0 \times w_0 + \alpha_1 \times w_1 + \alpha_2 \times w_2$$

ここで、a_0 と α_1 は TFIDF による重みと独立型の医学重要度の重みに対する係数である。これらの値はそれぞれ 1、0.75、0.5、0.25 の 4 つのパターンを用いる。依存型の医学重要度の重みの係数は α_2 は、その有効性を評価するために値を 1 と固定する。

5.6.3 実験による評価

改良手法の有効性を検証するために死亡予測の実験を行う。データセットはこれまでと同様の MIMIC II を用いる。本実験では SVM (linear)、Naive Bayes、K-Nearest Neighbor、Random forest、Decision tree の 5 つの分類器を用いる。これらは L2 正則化による特徴選択の後に実行される。ここで、L2 正則化と分類器のパラメータはデフォルトである。実験ではベースラインの TFIDF、TFIDF と独立型の医学重要度の組み合わせた手法、TFIDF と独立型及び依存型の医学重要度の組み合わせた手法の F1 スコアを比較する。それぞれの手法は TFIDF、TFIDF-MED-1、TFIDF-MED-2 と実験結果の中で表記される。F1 スコアを求める際は、全ての電子カルテを用いて 5-fold stratified cross validation を行う。

適切なパラメータ Δ と a の値はモデルの学習を通して得られた。適切なパラメータ Δ は概して 0.01 であった。それぞれの分類器ごとの適切なパラメータ a（a_0 と a_1）は、依存型の重みを考慮せずに TFIDF と独立型の医学重要度を組み合わせた手法を用いて求められた。表 5.12 は分類器ごとの適切なパラメータ a である。表 5.12 には 5 つの分類器（SVM

(linear)、Naive Bayes、K-Nearest Neighbor、Random forest、Decision tree）ごとに適したパラメータ a_0 と a_1 の値が示されている。

表 5.12: 分類器ごとの適切なパラメータ α

Method	α_0	α_1
SVM	1	0.5
NB	0.25	1
KNN	0.25	0.5
DT	0.5	0.5
RF	0.25	0.25

表 5.13 は死亡予測の実験結果である。表 5.13 には 5 つの分類器を用いた場合における、ベースラインの TFIDF と適切なパラメータを用いた場合の TFIDF-MED-1、TFIDF-MED-2 それぞれの F1 スコアが示されている。結果から、TFIDF-MED-2 が TFIDF と TFIDF-MED-1 と比べて、全ての分類器において高いスコアを出した。特に、Random forest を分類器とした場合に TFIDF-MED-2 と TFIDF-MED-1 の差は 0.26 であった。ベースラインの TFIDF は、SVM と NB、KNN の 3 つの分類器を用いた場合で F1 スコアの値が 0.1 以下であった。

5.6.4 考察と結論

本研究では、患者状態の重症度に基づいて独立型だけでなく依存型の医学重要度を捉える重み付け手法を提案した。依存型の医学重要度は患者状態の重症度を危険な疾患の組み合わせとして、チャールソン併存疾患指数の活用により、組み合わせの危険度とさらに Word2Vec の活用による依存度の観点から重みを導き出した。

表 5.13: 改良手法 1 の実験結果

	SVM	NB	KNN	DT	RF
TFIDF	0.0643	0	0.071	0.7826	0.8348
TFIDF-MED-1	0.8088	0.754	0.4103	0.7811	0.8373
TFIDF-MED-2	**0.8126**	**0.7555**	**0.4209**	**0.7912**	**0.8633**

改良手法の有効性を検証するために行われた死亡予測の実験から、独立型だけでなく依存型の医学重要度を考慮した手法による F1 スコアが、5 つの分類器全てにおいて最も高かったことから、疾患の危険な組み合わせという患者状態に基づいて患者ごとで医学単語の重みを可変させた方法の有効性が示されたと考えられる。このような結果から、文書の文脈を考慮した改良手法は、患者の重症度をより正確に捉えられたと考えられる。また、依存型の医学重要度の重み係数は 1 と固定し、TFIDF と独立型の医学重要度の重み係数の適切な値は、TFIDF と独立型の医学重要度を組み合わせた手法を用いて求められた。このように、改良手法の有効性は重み係数に依存しないため、本研究で提案する依存型の医学重要度の重要性が示されたと考えられる。

本研究ではチャールソン併存疾患指数に基づいて危険な疾患の組み合わせを患者状態としたが、今後の課題として別の視点から患者状態を捉えることを検討する。

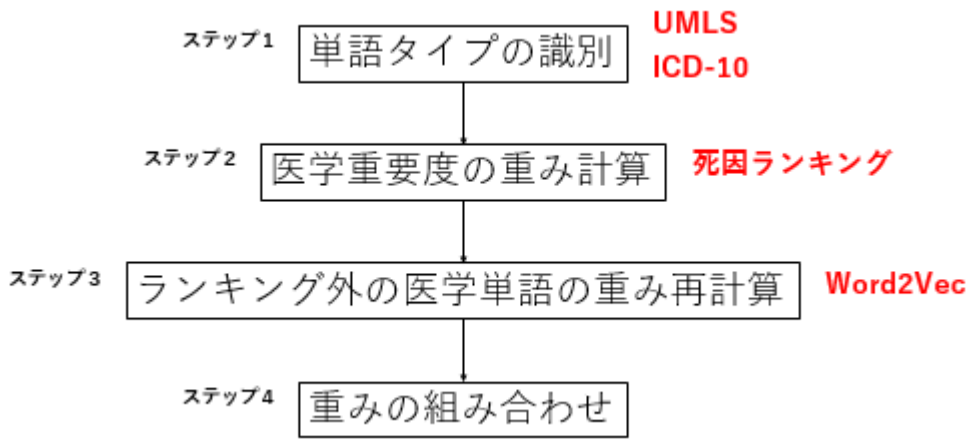


図 5.7: 改良手法 2 のフレームワーク

5.7 改良手法 2（ランク単語の類似度を用いた医学単語の重み付け手法の改良）

5.7.1 目的

これまでの提案手法で活用していた死因ランキングは 15 位までと限られており、ランキング外の医学単語に対しては暫定的にランキング内の単語よりも低い重みを患者の重症度の観点として与えていた。したがって、本研究では死因ランキング内にあるランク単語の類似度を活用して、ランキング外の医学単語に対して患者の重症度による重みを再計算することで提案手法の改良を行う。

5.7.2 改良方法

改良手法 2 のフレームワークは図 5.7 となる。

ステップ 1 の単語タイプの識別は、本論文で示している共通の重み付けフレームワークにより行われる。

ステップ 2 の独立型の医学重要度の重み計算は研究 2 で示したように重みを求めるが、ここでは、4 パターンでパラメータ Δ を変化させず、パラメータ Δ が 0.04 の場合と同程

度の重みとなる研究 2 の元となる論文 [60] で提示した以下の式を用いて改良手法の開発を行う。

$$v_i = \frac{(v_{max} - v_{min}) \times (\xi - i + 1)}{\xi} \tag{5.1}$$

v_i は順位 i 番目のランクに対応する医学重要度である。v_{max} と v_{min} は医学重要度の最大値と最小値にあたり、それぞれ 0.9、0.2 と設定している。ξ はランクにより区分けされるグループの総数を意味する。死因ランキングによる 15 のグループと ICD-10 ランク外単語のグループ、非 ICD-10 単語のグループの計 17 のグループを扱うため、ここで用いる ξ の値は 17 となる。表 5.14 は死因ランキングの死因名とそれに対応する ICD-10 コード、医学重要度の値を表している。この表には記載されていないが、ICD-10 ランク外単語の 16 番目のグループと非 ICD-10 単語の 17 番目のグループも同じように計算し、医学重要度はそれぞれ 0.08、0.04 となる。非医学単語に対しては医学重要度の重みはゼロとする。表で示すように i の値が高ければ高いほど医学重要度の値は小さくなる。

表 5.14: 死因ランキングの死因名と対応する ICD-10 コード及び医学重要度の値

The rank	The name of cause of death	The ICD-10 code(s)	The value
1	Disease of heart	I00-I09, I11, I13, I20-I51	0.7
2	Malignant neoplasms	C00-C97	0.66
3	Chronic lower respiratory diseases	J40-J47	0.62
4	Cerebrovascular diseases	I60-I69	0.58
5	Accidents (unintentional injuries)	V01-X59, Y85-Y86	0.54
6	Alzheimer's disease	G30	0.49
7	Diabetes mellitus	E10-E14	0.45
8	Nephritis, nephritic syndrome and nephrosis	N00-N07, N17-N19, N25-N27	0.41
9	Influenza and pneumonia	J09-J18	0.37
10	Intentional self-harm (suicide)	U03, X60-X84, Y87.0	0.33
11	Septicemia	A40-A41	0.29
12	Chronic liver disease and cirrhosis	K70, K73-K74	0.25
13	Essential hypertension and hypertensive renal disease	I10, I12, I15	0.21
14	Parkinson's disease	G20-G21	0.16
15	Pneumonitis due to solids and liquids	J69	0.12

ステップ 3 では、これまで暫定的に求めたランキング外のグループ 16 及び 17 に該当する医学単語の医学重要度の重みに対して、死因ランキング内のランク単語の類似度により重みの再計算を行う。本研究では、Word2Vec[25] により単語をベクトルに変換し、類似度計算においてよく用いられるコサイン類似度により死因ランキング内のランク単語との類似度を計算する。

まず、グループ 16 と 17 の医学単語全てに対してランキング内のランク単語との類似度を計算する。$S(t)$ は医学単語 t のそれぞれのランク内のランク単語に対する類似度で、$S_i(t)$ は医学単語 t のランク i 内のランク単語に対する類似度である。ランク i 内に複数のランク単語がある場合、$S_i(t)$ には複数の類似度が含まれる。次に、それぞれのランク i 内のランク単語に対する類似度の平均を求める。$avg(S_i(t))$ は医学単語 t のランク i 内のランク単語に対する類似度の平均である。ここで、$avg(S_i(t))$ はグループ 16 及び 17 に該当する他の医学単語のランク i 内のランク単語に対する類似度に基づいて Min-Max 正規化を行い、0 から 1 の範囲の値をとる。

重み再計算の際は、それぞれのランク内のランク単語に対する類似度の最大値を参照する。類似度の閾値を 0.8 とし、その閾値以上の類似度がある場合、グループ 16 と 17 に該当する医学単語の重みを再計算する。ただし、複数のランクで閾値以上の高い類似度を持つ場合は重み再計算から除外する。ある特定のランクで 0.8 以上の類似度を持つ医学単語 t において、$R(S(t))$ はそのランク番号となる。類似する特定のランク内のランク単語との類似度の最大値は、i が $R(S(t))$ の場合の $max(S_i(t))$ であり、そのランクの重みは研究 2 で示している v_i となる。

このように、類似する特定のランク番号 $i = R(S(t))$ によるランクの重み v_i と、そのランクとの類似度の最大値 $max(S_i(t))$、及び平均値 $avg(S_i(t))$ の 3 つの要素を用いて、医学重要度の重みを再計算する。その際、以下の足し算または掛け算の組み合わせを用いる。

$$\dot{w}(t) = w(t) + (v_i \times max(S_i(t)) \times avg(S_i(t))) \tag{5.2}$$

$$\dot{w}(t) = w(t) \times \{1 + (v_i \times max(S_i(t)) \times avg(S_i(t)))\} \tag{5.3}$$

ここで、i は類似する特定のランク番号 $R(S(t))$ である。$w(t)$ は医学単語 t の元の医学重要度の重みであり、$\dot{w}(t)$ は再計算後の重みである。

患者状態の重症度を考慮した医学重要度の重み w_1 は以下の式により TFIDF の重み w_0 と組み合わされる。

$$w = \alpha_1 \times w_0 + \alpha_2 \times w_1 \tag{5.4}$$

ここで、α_1 と α_2 はそれぞれ TFIDF の重みと医学重要度の重みの係数であり、本研究では表 5.15 の 10 通りの組み合わせを用いる。係数の組み合わせでは、特に医学重要度の重みを際立たせるように設定している。

表 5.15: 10 通りの係数の組み合わせ

	α_1	α_2 (= 2.0 - α_1)
Case 1	0.1	1.9
Case 2	0.2	1.8
Case 3	0.3	1.7
Case 4	0.4	1.6
Case 5	0.5	1.5
Case 6	0.6	1.4
Case 7	0.7	1.3
Case 8	0.8	1.2
Case 9	0.9	1.1
Case 10	1.0	1.0

5.7.3 実験による評価

医学重要度の重み再計算の有効性を検証するために死亡予測の実験を行う。用いるデータセットはこれまでと同じ MIMIC II である。分類器を Random forest とし、特徴選択を L2 正則化とする。

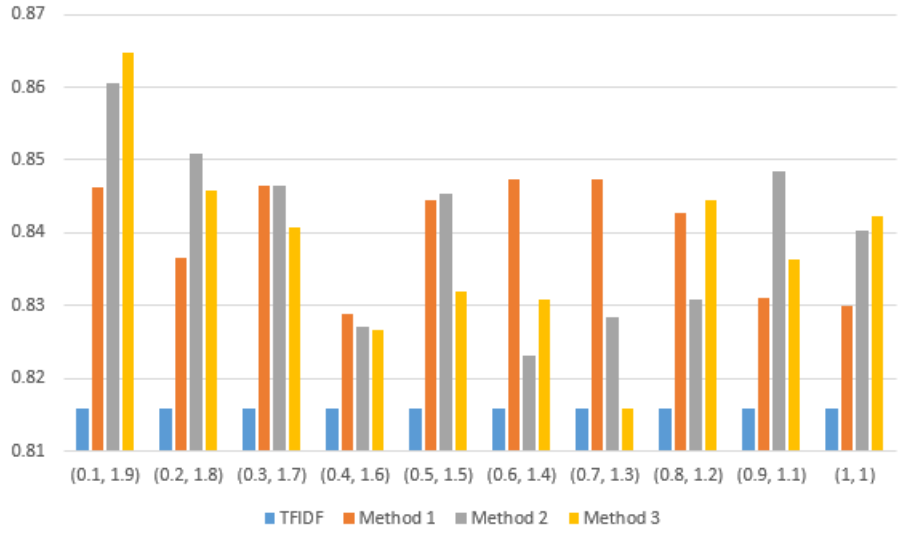


図 5.8: 10 通りの係数の組み合わせごとの 4 つの手法の F1 スコア

ベースラインを TFIDF とし、医学重要度を考慮した手法において、重み再計算なしの手法を Method 1、重み再計算ありで再計算時の組み合わせを足し算とした手法を Method 2、組み合わせを掛け算とした手法を Method 3 と記述する。これらの手法の比較は 13,026 文書を用いた 5-fold クロスバリデーションによる F1 スコアを用いて行う。

実験結果は図 5.8 である。x 軸は 10 通りの係数の組み合わせで、y 軸は F1 スコアの値で目盛りの最小値と最大値はそれぞれ 0.81、0.87 であり、組み合わせごとで 4 つの手法の結果を示している。結果から、死亡予測のタスクにおいて最も高い F1 スコアを出したのは Method 3 の医学重要度を考慮し、かつ、重み再計算の際に組み合わせを掛け算とした手法で、係数 α_1 と α_2 が 0.1、1.9 の場合であった。

表 5.16: 4 つの手法の F1 スコアの最大値

	TFIDF	Method 1	Method 2	Method 3
F1 スコア	0.8159	0.8474	0.8606	**0.8647**

F1 スコアの最大値を手法ごとで比較した表 5.16 によると、Method 3 が他の TFIDF、Method 1、Method 2 と比べて百分率にするとそれぞれ 4.88%、1.73%、0.41%高い結果となった。

5.7.4 考察と結論

本研究は死因ランキングを活用して患者の重症度を考慮したこれまでの提案手法である医学単語の重み付け手法において、死因ランキングのランク単語の類似度を活用して、ランキング外の医学単語の重症度を考慮し、医学重要度の重みを再計算する方法を提案した。重みを再計算する際は、単語を Word2Vec によりベクトルに変換してコサイン類似度によりランク単語との類似度を求めた。そして、類似する特定のランクの重みと、そのランクとの類似度の最大値と平均値という 3 つの要素により重みの再計算を行った。重み再計算を加えた提案手法の有効性を検証するために、死亡予測の実験を行った結果、重み再計算をした手法の F1 スコアが最も高かったことから、死因ランキング外の医学単語の医学重要度の重みをランキング内のランク単語との類似度を活用して再計算することの有効性が示されたと考えられる。また、最も高い F1 スコアの際の係数は α_1 が 0.1、α_2 が 1.9 の場合で、これは TFIDF の影響度を低くし、医学重要度の影響力を際立たせた場合であったことから、再計算後の医学重要度の有効性が示されたと考えられる。また、係数がどちらも 1.0 の場合であっても、Method 3 が最も高いスコアを出していたことから、重み係数に関わらず医学重要度の重みを再計算する方法の有効性が示されたと考えられる。Method 2 と比較すると、Method 3 は重み再計算の際に組み合わせを足し算ではなく掛け算としていたことから、再計算により重みを大幅に上昇させるのは有効ではないことがわかった。

本研究では、15 のランクを含む死因ランキングという限られた医学知識を活用する際に、Word2Vec を用いて、ランキング外の医学単語に対しても、ランク単語との類似度により患者の重症度という観点から重みを再計算することを可能にしており、このような限られた知識を拡張してコンピュータにとって語の意味理解を促進する方法は、意味を考慮して重みを与える様々な手法において有効であると考えられる。

第6章　結論

6.1　まとめ

本論文では、膨大なデータが存在し、日々増加しているテキストデータの分析の基盤となる、単語に対して重要度の観点から重みを数値で与えて文書を計算可能な形式に変換する単語の重み付け手法について、医学文書における意味を考慮した単語の重み付け手法を開発した。

第1章では、テキストデータの分析の重要性と伝統的な手法である TFIDF が単語の頻度情報により重みが付与されていて単語の意味が考慮されていないことを述べた。次に、医学分野における医学論文や電子カルテといった医学文書の分析とその課題について述べた。課題においては、理論的課題と実務的課題の2つを設定した。理論的課題は、文書分析で扱うタスクによって意味の捉え方が変化する中で、適切な医学知識を組み込みながら、様々な意味の視点を計算機で表現することとした。その背景としては、医学文書を対象とした単語の重み付け手法に関する研究では、TFIDF を用いた重み付けだけでなく、TFIDF を拡張して単語の意味を考慮した重み付け手法がこれまで多く開発されている。しかしながら、それらは情報検索や分類、センチメント分析といったある特定のタスクに対する手法となっており、タスクに依存しない意味を考慮した単語の重み付けのフレームワークはこれまでみられなかった。したがって、本論文では上記の理論的課題を設定した。ビジネスにおける実務的な課題は、意味を考慮した単語の重み付けという基礎的な研究を効果的に実践に結びつけるフレームワークを構築することとした。その背景としては、意味を考慮した単語の重み付け手法による変換後の文書を二次利用し分析に適用するには、どのような文書を用いているのかやどのような分析に適用するのか、どのよう

な意味を捉えるのかといった問題に対応する必要があると考えられる。しかしながら、そのためのフレームワークがこれまで存在していなかった。したがって、本論文では上記の実務的課題を設定した。そして、これらの課題を解決するための本論文の着眼点として、単語の重み付けという計算プロセスを人が行ったらどうなるかをまず考え、人の認知プロセスに着目すると、人が想起する状況によって単語の重要度は変化し得ることがみえてくることを述べた。これは、状況を特徴づけるフレームが単語の意味を割り当てるとするフレーム意味論の考えで説明でき、この考え方を単語の重み付け手法に取り入れることで、単語の意味は様々な視点で捉えられること、また具体的な応用に適する百科事典的意味という観点から計算機が語の意味を理解することが可能になることを述べた。そこで、本論文の目的として、フレーム意味論の考え方を用いて、設定する文書分析のある目的に適した意味を考慮した医学単語の重み付け手法を開発することとした。本論文では、文書分析として特に情報検索と予測という医学分野における重要な課題でかつ研究が盛んに行われている2つを設定し、医学文書の情報検索と電子カルテを用いた患者状態の予測という2つの具体的な目的に適した意味を考慮した単語の重み付け手法の開発を行うこととした。本論文の目的を達成するために、2つのフレームワークを提案することを述べた。1つは、フレーム意味論の考え方を単語の重み付けに取り入れる手順を述べ、その提案手順に基づいた重み付けフレームワークを示した。もう1つは、医学知識表現に基づいた、医学文書分析に共通する意味を考慮した単語の重み付けフレームワークを述べた。章の最後に本論文の5つの貢献と、本論文の以降の構成を述べた。

第2章では、意味を考慮した単語の重み付け手法の関連研究として、オントロジーと単語のクラス情報について述べ、それらを活用した先行研究の概要を述べた。そして文書分析ごとの課題として、特に医学論文の情報検索と電子カルテの死亡予測における関連研究とその課題を述べ、提案する手法の概要を述べた。

第3章では、本論文で提案するフレームワークを述べた。始めにフレーム意味論の考え方を単語の重み付け手法へ適用するための手順を述べ、その提案手順に基づく意味を考慮した重み付けのフレームワークを述べた。このフレームワークにより、フレーム意味論の

考え方を取り入れながら、関連研究と同様にオントロジーと単語のクラス情報を活用して、意味を考慮した単語の重み付け手法を開発した。次に、医学知識表現に基づいた、医学文書分析に共通する意味を考慮した単語の重み付けフレームワークを述べた。提案する医学知識表現は、医学文書内の単語を階層で区分けし、その区分けしたカテゴリーに対して重みを与え、そのカテゴリー内に属する単語は連続値で重みを付与できる場合を除き、全てにそのカテゴリーの重みを同じように付与することであった。意味の視点を計算機で表現するために、それらのカテゴリー内の単語集合により様々な意味の視点を計算機に理解させた。最後に、提案手法の本論文内の位置づけを述べた。

第 4 章と第 5 章では、第 3 章で提案したフレームワークを用いて、設定した情報検索と予測というそれぞれの目的に適した重み付け手法を提案し、その有効性を示した。医学論文の情報検索では、機械学習の手法（LDA）によるトピック情報を用いて、医学単語の一般性及び特定性を捉えた重み付け手法を開発した。電子カルテを用いた死亡予測では、死因ランキングという医学知識を活用することで、患者の重症度という観点から単語の重要度を捉えた重み付け手法を開発した。電子カルテを用いた死亡予測では、2 つの改良手法を述べ、チャールソン併存疾患指数や機械学習手法の Word2Vec を活用することで改良の有効性も示された。

第 6 章では、これまで本論文のまとめを述べてきた。本論文の目的は、フレーム意味論の考え方を用いて、目的に適した意味を考慮した医学単語の重み付け手法を開発することであった。フレーム意味論の考え方を取り入れることで、主な前提として、単語の重要度の決定の仕方は一意に決まらず、重み付けをどのような文書分析に適用するか、どのような文書を用いるか、といった想起する状況（視点）に依存するため多様であることを保持した。また、主な意味として、計算機に具体的な応用に適した百科事典的意味を理解させることができた。第 4 章と第 5 章で提案した情報検索と予測の目的に適した開発手法において、医学文書を用いた情報検索では、医学単語の特定性として共起の偏りの視点による百科事典的意味を学習により捉え、その意味を考慮した単語の重み付け手法を開発し、情報検索の実験でその有効性を示した。電子カルテを用いた患者状態の予測では、患者の

重症度の視点による百科事典的意味を既存の百科事典的知識である死因ランキングをもとに捉え、その意味を考慮した単語の重み付け手法を開発し、死亡予測の実験でその有効性を示した。さらに、より適切に患者の重症度を捉えるための 2 つの改良手法を開発し、それぞれの改良の有効性を示した。したがって、本論文の目的である、フレーム意味論の考え方を用いて、目的に適した意味を考慮した医学単語の重み付け手法の開発が達成されたと考えられる。以降では、本論文の提案手法の特徴と課題、効果を述べる。

6.2 本論文の提案手法の特徴・課題・効果

6.2.1 本論文の提案手法の特徴

6.2.1.1 人の認知プロセスを参考にした重み付け

本論文では、人間の振る舞いを参考に、単語の重み付けという計算手法を照射し直した。それによって、単語の重要度の決定の仕方は、人が想起する状況によって変化する、という前提を保持していることが特徴である。この状況を説明できるフレーム意味論の考え方（状況を特徴付けるあるフレームを考えることで、そのフレームがある単語の意味を割り当てる）を取り入れることで、単語の意味は様々な視点で捉えられることを前提として保持し、かつ、具体的な応用に適した百科事典的意味の観点から計算機が単語の意味を理解することを可能にした。それによって、実践を重視した目的志向の意味を考慮した重み付けの開発を可能にした点が本論文の特徴であると考えられる。

6.2.1.2 意味階層からみる二段階の重み付け

単語の意味を階層としてみると、本研究の特徴は、単語がどの領域に属しているかという一階層の意味をオントロジーにより捉えるだけではなく、さらにその領域単語間での百科事典的意味の観点による意味的な違いという単語の二階層の意味も複数の医学知識や機械学習により捉えた重み付けであると考えられる。

6.2.1.3　オントロジーの活用による百科事典的知識の参照

オントロジーを活用することで、医学文書内の単語の概念の観点から、医学単語かそうでないかを識別できた。また、単語の概念情報を活用することにより、単語を ICD-10 の医学知識や死因ランキングという百科事典的知識へマッピングすることが可能になった。このように、オントロジーの活用により計算機が語の意味を理解するために百科事典的知識を参照する方法を提案したことが本論文の特徴であると考えられる。

6.2.1.4　サービス視点

単語の重み付けという計算手法にサービスの視点を取り入れている点も特徴である。サービスの価値はそれを受け取る人によって決まるというサービスの考え方と同様に、本研究もある目的を設定し、その目的に適した意味の捉え方を計算機にさせていることから、上記のサービス概念が提案手法の基礎をなしている。

6.2.1.5　応用可能性

単語の重み付けは、単語に対して重要度の観点から重みを与え、文書を計算可能な形式に変換する文書分析の基盤となる手法である。本論文で開発してきた重み付け手法は、伝統的な手法である TFIDF をベースにしているため、文書をベクトル空間上で簡易に医学文書を計算可能な形式で表現し、蓄積できる。さらに、フレーム意味論の考え方を取り入れていることで、多様な視点の意味を保持しながら、データの二次利用として様々な文書分析への適用が可能になると考えられる。本研究では、オントロジーを活用して医学単語の識別と他の医学知識へのマッピングをすることで、医学文書内の単語を階層的に区分けし、その区分けしたカテゴリーを重み付けに活用しているため、提案手法はオントロジーが整備されている領域の文書分析への応用が特に期待できる。

6.2.1.6　新規性

単語間の概念関係を階層構造でシステマティックに表現するオントロジー [9] により、単語の意味的側面を考慮しながら TF-IDF を改良し、医学文書に適用する研究は多くある（先行研究の例 [56, 21]）。オントロジーだけでなく、クラス情報を活用したハイブリッドな重み付け手法も医学論文データを用いて開発されている [53, 58]。先行研究と同様に、本論文もオントロジーとクラス情報を活用した重み付け手法を行ってきた。特に、UMLS のオントロジーと単語のタイプやトピック、死因ランクといったクラス情報を活用した重み付けを提案してきた。先行研究と異なる点は、フレーム意味論の考え方を取り入れて、実践を重視した目的志向の重み付けフレームワークを用いていること、単語の意味を百科事典的意味の観点から二階層で捉えていること、具体的には、UMLS オントロジーとそこから得られる単語タイプだけでなく、他の医学知識による死因ランクや学習により取得する単語のトピックというクラス情報を活用して、情報検索に適した医学単語の特定性の重みと、死亡予測に適した患者状態の重症度の重みを捉えた手法を開発したことであると考えられる。さらに、本論文では、単語の重み付けにおいて計算機に人が想起する単語の様々な意味の視点を理解させる新たな医学知識表現を提案したことが、先行研究と異なる点であると考えられる。医学知識表現は、医学知識（特にランキング）や機械学習を用いながらオントロジーや単語のクラス情報を活用して、医学文書内の単語を階層的に区分けし、それぞれのカテゴリー内の単語集合を意味の視点として計算機に表現させることを可能とした。そして、この医学知識表現を活用して、カテゴリーに対して重みを付与することで（連続値を付与できる場合を除いて）、カテゴリー内の単語に同じ重みを付与させるという文書分析に共通する重み付けフレームワークを提案した。

6.2.2　本論文の提案手法の課題

本論文で提案した 2 つのフレームワークについて、フレーム意味論の考え方を取り入れた重み付けフレームワークの課題では、フレーム意味論の考え方を取り入れる手順はど

の分野の文書分析にも適用できると考えられるが、ある分野の単語の百科事典的意味を捉える方法は、その分野の文書の特徴や、分野で利用可能な知識などに依存すると考えられる。提案する医学知識表現に基づく文書分析に共通する重み付けフレームワークでは、医学単語の識別や他の医学知識へのマッピングの際に、特にオントロジーに基づいているため、オントロジーが整備されていない分野への適用は困難であると考えられる。医学知識表現においては、表現を拡張させていく際に、is-a 関係だけでなく、概念間の他の関係性を考慮して記述していく必要があると考えられる。

本論文の対象は医学文書であるため、得られた結果（データ）を医療へ応用することは、患者の回復や安全に大きな影響を及ぼすことになる。また、結果の解釈には高い専門性が必要であると考えられる。さらに解釈の仕方は受け取る人の立場や思いなどによって異なると考えられる。したがって、得られた結果を医療へ応用するために、患者や専門家（医者）、技術者などの思いや考えを考慮する必要があると考えらえる。

6.2.3 本論文の提案手法の効果

6.2.3.1 学術への効果

これまで医学文書を対象とした単語の重み付け手法に関する研究では、TFIDF という単語の頻度情報のみを考慮した重み付けだけでなく、TFIDF を拡張して単語の意味を考慮した重み付け手法が多く開発されてきている。しかし、それらは情報検索や分類、センチメント分析といったある特定の分析に対する手法であり、分析に依存しない意味を考慮した単語の重み付けフレームワークはこれまでみられなかった。したがって、理論的な課題として、対象とする文書分析によって意味の捉え方が変化する中で、適切な医学知識を組み込みながら、様々な意味の視点を計算機で表現することと設定したこの課題を解決するために、本論文では、死因ランキングといった医学知識や機械学習を用いながらオントロジーや単語のクラス情報を活用することで、医学文書内の単語を階層的に区分けし、それぞれのカテゴリー内の単語集合を意味の視点として計算機に表現させ、連続値を付与

できる場合を除いてカテゴリーに対して重みを付与することで、カテゴリー内の単語に同じ重みを与えるという文書分析に共通する重み付けフレームワークを提案した。このように、単語の重み付けにおいて計算機に人が想起する単語の様々な意味の視点を理解させる新たな医学知識表現の提案により、様々な意味の視点を計算機で表現できるようになったことが本論文の学術的な効果であると考えられる。

6.2.3.2　社会への効果

これまで意味を考慮した単語の重み付けという基礎的な研究を効果的に実践に結びつけるためのフレームワークが存在していなかった。したがって、ビジネスにおける実務的な課題として、単語の意味を考慮した重み付けによる計算可能な形式に変換後の文書を二次利用し分析に適用するにあたって、どのような文書を用いているのか、どのような分析に適用するのか、どのような意味を捉えるのかといった問題に対応する実践に即したフレームワークを構築することと設定した。この課題のために本論文では、単語の重み付けという計算手法に対してフレーム意味論の考え方を適用することによって、単語の意味は様々な視点で捉えられることを前提として保持し、具体的な応用に適した百科事典的意味という観点から計算機が単語の意味を理解することを可能にした。それにより、単語の重み付けという文書分析の基礎的手法を効果的に実践に結びつけるため、実践を重視した目的志向の意味を考慮した単語の重み付け手法の開発が可能となったことが本論文の社会的な効果であると考えられる。

本論文では医学文書を対象に意味を考慮した重み付けをしたが、提案する重み付けのフレームワーク（百科事典的意味の観点からどのような視点で単語の重要度を捉え、どのような文書分析の目的に適用するかを考える→そのために計算機にとって必要な知識や学習は何かを考える→設定した応用の場面に適した手法のための適切な知識や学習を組み込む重み付けプロセスを設定する）は、他の様々な分野の文書分析にも適用が可能であると考えられる。

特に、本論文では医学論文の情報検索と電子カルテを用いた患者状態の予測（死亡予

測）に適した重み付け手法を開発してきたため、これらがどのように社会に応用できるかについて医者・患者・病院という 3 つの視点から以下に述べる。

1. 医者にとっては、患者状態の予測に適した重み付け手法による計算可能な形式に変換後の文書を用いて、データの二次利用の観点から、特定の疾患の予測や、患者状態の重症度に基づいた類似症例検索といった、患者のリスクに関する様々な文書分析による診断支援が可能になると考えられる。医学論文の情報検索に適した提案する重み付け手法の結果からは、医学論文から診療に役立つ情報の効果的な取得が期待できる。

2. 患者にとっては、患者状態の予測に適した重み付け手法の結果から、セカンドオピニオンのような意思決定支援の機会が得られると考えられる。また、医学論文の情報検索に適した提案する重み付け手法の結果から、インフォームドコンセントのため、診療に係る情報提供による意思決定支援の機会も得られると考えられる。

3. 病院にとっては、情報検索に適した重み付け手法の結果から、医者が診療のために探したい情報を医学論文から効果的に取得できることにより、診療時間の短縮による病院経営の効率化が期待できる。また、患者状態の予測に適した重み付け手法の結果から、医者が患者のリスクの観点から患者状態を適切に把握することでも病院経営の効率化が期待できる。

謝辞

始めに、北陸先端科学技術大学院大学知識科学研究科の先生方や関係者の方々に、知識科学という学問に関する講義の受講や TA の業務をさせて頂けたこと、研究のアドバイスやディスカッションをして頂いたことに御礼を申し上げます。また、研究活動や日常生活に係る諸手続きやアドバイスをして頂いた学内の方々に御礼申し上げます。博士後期課程在学時の経済援助として、Doctoral Research Fellow (DRF) の奨励金と Research Promotion Award (RPA) の雇用型研究奨励金を北陸先端科学技術大学院大学から頂きました。御礼申し上げます。また、日本学生支援機構奨学金の援助に御礼を申し上げます。このような経済的な援助によって、博士後期課程で研究活動に集中する環境が得られました。

研究活動では、まず、所属研究室の主指導教員である Ho Tu Bao 教授には多くの時間を割いてご指導をして頂きました。御礼を申し上げます。主指導教員から学んだ多くのことは今後に活かさせて頂きます。また、所属学会の一般社団法人日本医療情報学会や一般社団法人人工知能学会、知識共創フォーラム、国際学会の ACIS 等での学会発表や論文の査読において、アドバイスやディスカッションをして頂いた学内外の先生方に御礼申し上げます。そして、予備審査で博士論文修正のために多くのご指摘をして頂いた審査員の先生方に御礼申し上げます。

最後に、私が望んでいた研究活動がこれまで行えてきたのは、何よりも両親と家族の存在があったからです。両親と家族に感謝を致します。

研究業績

学術誌掲載論文

松尾亮輔, Ho Tu Bao. (2017). 診断用の医学論文検索における 4 つの医学単語の重み付け手法の分析, 医療情報学 37 巻 3 号, 105-115. (査読あり)

国際学会発表論文（ポスター発表を含む）

Matsuo R., Ho T.B. (2014). An Ontology Based Term Weighting Method for Exploiting EMR, Asia Conference on Information Systems ACIS 2014, Proceedings of Asian Conference on Information Systems ACIS 2014, pp.306-313, 1-3 December, Nha Trang. (口頭発表) (査読あり)

Matsuo R., Ho T.B. (2015). A Severity based Information Retrieval for Electronic Medical Records, The 5th Annual Translational Bioinformatics Conference, November 7-9, Tokyo. (ポスター発表) (査読あり)

国内学会発表論文（ポスター発表を含む）

松尾亮輔, Ho Tu Bao. (2016). 世界知識による医学単語の意味を考慮した単語重み付け手法, 第 6 回知識共創フォーラム, 2016 年 3 月 12-13 日, 石川県金沢市. (ポスター発表) (査読あり)

松尾亮輔, Ho Tu Bao. (2016). 医学文書の情報検索における 4 つの医学単語の重み付け手法の分析, 第 20 回日本医療情報学会春季学術大会, プログラム抄録集, O-3, 2 ページ, 2016 年 6 月 2- 4 日, 島根県松江市. (口頭発表) (査読あり)

松尾亮輔, Ho Tu Bao. (2016). 重症度を考慮した医学単語重み付け手法による死亡予測, 第 30 回人工知能学会全国大会, 第 30 回人工知能学会全国大会論文集, 4D1-4in2, 4 ページ, 2016 年 6 月 6- 9 日, 福岡県北九州市. (口頭発表とポスター発表) (査読あり)

松尾亮輔, Ho Tu Bao. (2016). 患者状態の重症度を考慮した独立型と依存型の重み結合による医学単語の重み付け手法, 第 36 回医療情報学連合大会 (第 17 回日本医療情報学会学術大会), 4 ページ, 2016 年 11 月 21-24 日, 神奈川県横浜市. (ポスター発表) (査読あり)

松尾亮輔, Ho Tu Bao. (2016). ランク単語の類似度を用いた医学単語重み付け手法の改良, 第 2 回日本医療情報学会「医用知能情報学研究会」人工知能学会「医用人工知能研究会」合同研究会, SIG-AIMED-002-05, 5 ページ, 2016 年 11 月 9 日, 神奈川県横浜市. (口頭発表) (査読なし)

松尾亮輔, Ho Tu Bao. (2016). オントロジーの活用によるフレーム意味論の単語重み付けへの適用事例, 第 40 回 SWO 研究会-セマンティックウェブとオントロジー研究会, SIG-SWO-040-05, 4 ページ, 2016 年 11 月 10 日, 神奈川県横浜市. (口頭発表) (査読なし)

松尾亮輔, Ho Tu Bao. (2017). 在院日数予測のための意味を考慮した単語重み付け手法, 第 3 回日本医療情報学会「医用知能情報学研究会」人工知能学会「医用人工知能研究会」合同研究会, SIG-AIMED-003-05, 6 ページ, 2017 年 3 月 9-10 日, 神奈川県三浦市. (口頭発表) (査読なし)

松尾亮輔, **Ho Tu Bao. (2017).** 医学文書分析における意味を考慮した単語重み付けのための医学知識表現, 第 **7** 回知識共創フォーラム, **6** ページ, **2017** 年 **3** 月 **21-22** 日, 大阪府大阪市. (口頭発表) (査読あり)

松尾亮輔, **Ho Tu Bao. (2017).** 意味を考慮した単語重み付けによる重症度に基づく 類似症例検索, 第 **21** 回日本医療情報学会春季学術大会, **2** ページ, **2017** 年 **6** 月 **1- 3** 日, 福井県福井市. (ポスター発表) (査読あり)

参考文献

[1] Alan R Aronson. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In *Proceedings of the AMIA Symposium*, pages 17–21. American Medical Informatics Association, 2001.

[2] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3(Jan):993–1022, 2003.

[3] David Blumenthal and Marilyn Tavenner. The “meaningful use” regulation for electronic health records. *New England Journal of Medicine*, 363(6):501–504, 2010.

[4] Olivier Bodenreider. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Research*, 32(suppl 1):D267–D270, 2004.

[5] Mary E Charlson, Peter Pompei, Kathy L Ales, and C Ronald MacKenzie. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases*, 40(5):373–383, 1987.

[6] Aaron M Cohen and William R Hersh. A survey of current work in biomedical text mining. *Briefings in Bioinformatics*, 6(1):57–71, 2005.

[7] Ronen Feldman and James Sanger. *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge University Press, 2007.

[8] Charles Fillmore. Frame semantics. *Linguistics in the Morning Calm*, pages 111–137, 1982.

[9] Thomas R Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993.

[10] Mark Hoogendoorn, Peter Szolovits, Leon MG Moons, and Mattijs E Numans. Utilizing uncoded consultation notes from electronic medical records for predictive modeling of colorectal cancer. *Artificial Intelligence in Medicine*, 69:53–61, 2016.

[11] Rein Houthooft, Joeri Ruyssinck, Joachim van der Herten, Sean Stijven, Ivo Couckuyt, Bram Gadeyne, Femke Ongenae, Kirsten Colpaert, Johan Decruyenaere, Tom Dhaene, et al. Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores. *Artificial Intelligence in Medicine*, 63 (3):191–207, 2015.

[12] The International Health Terminology Standards Development Organisation (IHTSDO). SNOMED-CT. http://www.ihtsdo.org/snomed-ct/（アクセス日 2017 年 1 月 5 日）.

[13] Fernando Jiménez, Gracia Sánchez, and José M Juárez. Multi-objective evolutionary algorithms for fuzzy classification in survival prediction. *Artificial Intelligence in Medicine*, 60(3):197–219, 2014.

[14] Liping Jing, Lixin Zhou, Michael K Ng, and J Zhexue Huang. Ontology-based distance measure for text clustering. In *Proceedings of SIAM DM workshop on text mining*, 2006.

[15] Michel Joubert, Marius Fieschi, Jean-Jacques Robert, Françoise Volot, and Dominique Fieschi. Umls-based conceptual queries to biomedical information databases. *Journal of the American Medical Informatics Association*, 5(1):52–61, 1998.

[16] Ramakanth Kavuluru, Anthony Rios, and Yuan Lu. An empirical evaluation of

supervised learning approaches in assigning diagnosis codes to electronic medical records. *Artificial Intelligence in Medicine*, 65(2):155–166, 2015.

[17] Youngjoong Ko. A new term-weighting scheme for text classification using the odds of positive and negative class probabilities. *Journal of the Association for Information Science and Technology*, 66(12):2553–2565, 2015.

[18] Michael Krauthammer and Goran Nenadic. Term identification in the biomedical literature. *Journal of Biomedical Informatics*, 37(6):512–526, 2004.

[19] Man Lan, Chew Lim Tan, and Hwee-Boon Low. Proposing a new term weighting scheme for text categorization. In *AAAI*, volume 6, pages 763–768, 2006.

[20] Ronald W Langacker. *Foundations of cognitive grammar: Theoretical prerequisites*, volume 1. Stanford University Press, 1987.

[21] Qiming Luo, Enhong Chen, and Hui Xiong. A semantic term weighting scheme for text categorization. *Expert Systems with Applications*, 38(10):12708–12716, 2011.

[22] Christopher D Manning, Prabhakar Raghavan, Hinrich Schütze, et al. *Introduction to information retrieval*, volume 1. Cambridge University Press, 2008.

[23] Justin Martineau and Tim Finin. Delta tfidf: An improved feature space for sentiment analysis. In *Proceedings of the Third AAAI International Conference on Weblogs and Social Media*, 2009.

[24] Genevieve B Melton and George Hripcsak. Automated detection of adverse events using natural language processing of discharge summaries. *Journal of the American Medical Informatics Association*, 12(4):448–457, 2005.

[25] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Dis-

tributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, pages 3111–3119, 2013.

[26] Harvey J Murff, Fern FitzHenry, Michael E Matheny, Nancy Gentry, Kristen L Kotter, Kimberly Crimin, Robert S Dittus, Amy K Rosen, Peter L Elkin, Steven H Brown, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA*, 306(8):848–855, 2011.

[27] Sherry L Murphy, Jiaquan Xu, and Kenneth D Kochanek. Deaths: final data for 2010. *NVSR*, 61(4):1–118, 2013.

[28] Giulio Napolitano, Adele Marshall, Peter Hamilton, and Anna T Gavin. Machine learning classification of surgical pathology reports and chunk recognition for information extraction noise reduction. *Artificial Intelligence in Medicine*, 70:77–83, 2016.

[29] Natalya F Noy, Nigam H Shah, Patricia L Whetzel, Benjamin Dai, Michael Dorf, Nicholas Griffith, Clement Jonquet, Daniel L Rubin, Margaret-Anne Storey, Christopher G Chute, et al. Bioportal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research*, 37:W170–W173, 2009.

[30] The National Library of Medicine (NLM). MeSH. https://www.nlm.nih.gov/mesh/（アクセス日 2017 年 1 月 5 日）.

[31] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(Oct):2825–2830, 2011.

[32] Janet FE Penz, Adam B Wilcox, and John F Hurdle. Automated identification of adverse events related to central venous catheters. *Journal of Biomedical Informatics*, 40(2):174–182, 2007.

[33] Juan Ramos. Using tf-idf to determine word relevance in document queries. Technical report, Department of Computer Science, Rutgers University, 2003.

[34] Graeme Richards, Victor J. Rayward-Smith, PH Sönksen, S Carey, and C Weng. Data mining for indicators of early mortality in a database of clinical records. *Artificial Intelligence in Medicine*, 22(3):215–231, 2001.

[35] Rachel L Richesson, Jimeng Sun, Jyotishman Pathak, Abel N Kho, and Joshua C Denny. A survey of clinical phenotyping in selected national networks: demonstrating the need for high-throughput, portable, and computational methods. *Artificial Intelligence in Medicine*, 71:57–61, 2016.

[36] Vicent J Ribas Ripoll, Alfredo Vellido, Enrique Romero, and Juan Carlos Ruiz-Rodríguez. Sepsis mortality prediction with the Quotient Basis Kernel. *Artificial Intelligence in Medicine*, 61(1):45–52, 2014.

[37] Raul Rodriguez-Esteban. Biomedical text mining and its applications. *PLoS Computational Biology*, 5(12):e1000597, 2009.

[38] Kari L Ruud, Matthew G Johnson, Juliette T Liesinger, Carrie A Grafft, and James M Naessens. Automated detection of follow-up appointments using text mining of discharge records. *International Journal for Quality in Health Care*, 22(3): 229–235, 2010.

[39] Mohammed Saeed, Mauricio Villarroel, Andrew T Reisner, Gari Clifford, Li-Wei Lehman, George Moody, Thomas Heldt, Tin H Kyaw, Benjamin Moody, and Roger G Mark. Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): A

public-access intensive care unit database. *Critical Care Medicine*, 39(5):952–960, 2011.

[40] Mohammed M Sakre, Mohammed M Kouta, and Ali MN Allam. Weighting query terms using wordnet ontology. *International Journal of Computer Science and Network Security*, 9:349–358, 2009.

[41] Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, 24(5):513–523, 1988.

[42] Matthew S Simpson, E Voorhees, and William Hersh. Overview of the trec 2014 clinical decision support track. In *Proceedings of the 2014 Text Retrieval Conference*, 2014.

[43] Irena Spasic, Sophia Ananiadou, John McNaught, and Anand Kumar. Text mining and ontologies in biomedicine: Making sense of raw text. *Briefings in Bioinformatics*, 6(3):239–251, 2005.

[44] Vijaya Sundararajan, Toni Henderson, Catherine Perry, Amanda Muggivan, Hude Quan, and William A Ghali. New icd-10 version of the charlson comorbidity index predicted in-hospital mortality. *Journal of Clinical Epidemiology*, 57(12):1288–1294, 2004.

[45] V Sureka and SC Punitha. Approaches to ontology based algorithms for clustering text documents. *International Journal of Computer Technology and Applications*, 3 (5):1813–1817, 2012.

[46] Hmway Hmway Tar and Thi Thi Soe Nyunt. Ontology-based concept weighting for text documents. *World Academy of Science, Engineering and Technology*, 81: 249–253, 2011.

[47] Princeton University. WordNet. https://wordnet.princeton.edu/（アクセス日　2017年1月5日）.

[48] Giannis Varelas, Epimenidis Voutsakis, Paraskevi Raftopoulou, Euripides GM Petrakis, and Evangelos E Milios. Semantic similarity methods in wordnet and their application to information retrieval on the web. In *Proceedings of the 7th Annual ACM International Workshop on Web Information and Data Management*, pages 10–16. ACM, 2005.

[49] Richard Wicentowski and Matthew R Sydes. Using implicit information to identify smoking status in smoke-blind medical discharge summaries. *Journal of the American Medical Informatics Association*, 15(1):29–31, 2008.

[50] World Health Organization. *International statistical classification of diseases and related health problems*, volume 1. World Health Organization, 2004.

[51] Xu Yabin and Chen Jian. Research on fuzzy clustering of EMR based on XML. In *2010 International Conference on Biomedical Engineering and Computer Science*, pages 1–4. IEEE, 2010.

[52] Christopher C Yang and Pierangelo Veltri. Intelligent healthcare informatics in big data era. *Artificial Intelligence in Medicine*, 65(2):75–77, 2015.

[53] Hong Yu and Yong-gang Cao. Using the weighted keyword models to improve information retrieval for answering biomedical questions. *Summit on Translational Bioinformatics*, pages 143–147, 2009.

[54] John Zakos and Brijesh Verma. Concept-based term weighting for web information retrieval. *International Journal of Computational Intelligence and Applications*, 6 (02):193–207, 2006.

[55] Qing T Zeng, Sergey Goryachev, Scott Weiss, Margarita Sordo, Shawn N Murphy, and Ross Lazarus. Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system. *BMC Medical Informatics and Decision Making*, 6(1):30, 2006.

[56] Xiaodan Zhang, Liping Jing, Xiaohua Hu, Michael Ng, and Xiaohua Zhou. A comparative study of ontology based term similarity measures on PubMed document clustering. In *International Conference on Database Systems for Advanced Applications*, pages 115–126. Springer, 2007.

[57] Xiaodan Zhang, Liping Jing, Xiaohua Hu, Michael Ng, Jiali Xia Jiangxi, and Xiaohua Zhou. Medical document clustering using ontology-based term similarity measures. *International Journal of Data Warehousing and Mining*, 4(1):62–73, 2008.

[58] Weizhong Zhu, Xuheng Xu, Xiaohua Hu, Il-Yeol Song, and Robert B Allen. Using UMLS-based re-weighting terms as a query expansion strategy. In *IEEE International Conference on Granular Computing*, pages 217–222, 2006.

[59] 串間宗夫, 荒木賢二, 鈴木斎王, 荒木早苗, and 仁鎌照絵. 電子カルテネットワークシステム (IZANAMI) のテキストデータ分析. In **電子情報通信学会技術研究報告**. *IN,* **情報ネットワーク**, volume 109, pages 383–388, 2010.

[60] 松尾 亮輔 and Ho Tu Bao. 重症度を考慮した医学単語重み付け手法による死亡予測. In **第** *30* **回人工知能学会全国大会**, pages 4D1–4in2, 2016.

[61] 今井健, 荒牧英治, 梶野正幸, 美代賢吾, and 大江和彦. 構文情報と医学用語属性を用いた画像診断所見オントロジーの構築の試み. **医療情報学**, 25(6):395–403, 2005.

[62] 国府裕子, 周俊, 古崎晃司, 今井健, 大江和彦, and 溝口理一郎. 臨床医療オントロジーの構築に関する基礎的な考察. In **第** *22* **回人工知能学会全国大会**, pages 2E3–01, 2008.

[63] 大江和彦 and 今井健. 臨床医学知識処理を目指した医療オントロジー開発. **人工知能学会誌**, 25(4):493–500, 2010.

[64] 小原京子. フレーム意味論と日本語フレームネット (特集文の意味と語の意味). **日本語学**, 25(6):40–52, 2006.

[65] 小野大樹, 高林克日己, 鈴木隆弘, 横井英人, 井宮淳, and 里村洋一. テキストマイニングによる退院サマリー自動分類の試み. **医療情報学**, 24(1):35–44, 2004.

[66] 岡本和也, 竹村匡正, 黒田知宏, 長瀬啓介, and 吉原博幸. 文脈に基づく類似診療文書検索システム. **生体医工学**, 44(1):199–206, 2006.

[67] 溝口理一郎. **オントロジー工学**. オーム社, 2005.

[68] 籾山洋介. **言語科学の百科事典**, chapter 1-8. 認知言語学, pages 157–177. 2006.

[69] 籾山洋介. メタファーの認知的基盤と経験的基盤. In **言語文化研究叢書**, volume 7, pages 97–111. 2008.

[70] 藤井聖子 and 小原京子. 構文研究の理論と実践 (6) フレーム意味論とフレームネット. **英語青年**, 149(6):373–376, 2003.

[71] 谷恵里香 and 砂山渡. 電子カルテにおける新人とベテランの特徴比較支援システム. In **第 *3* 回人工知能学会インタラクティブ情報アクセスと可視化マイニング研究会資料**, pages 37–43, 2013.