

教師なし学習におけるクラスタリングを用いた ワークショップ会議録からの複数話題抽出の検討

瀧田美聡¹ 伊集院幸輝² 小早川真衣子³ 西村拓一² 中村嘉志¹

概要: 熟練作業者の高齢化を背景に、製造業や介護の現場における知識を集約し、人材育成やヒューマンエラーの予防に活用することが求められている。我々はこのような熟練者の持つ暗黙的な知識の集約と構造化のために、知識構造化ワークショップを実践してきた。この実践においては、事前準備、進行促進、振り返りの3つの作業をおこなうファシリテーターの役割が重要かつ不可欠である。本稿では、ファシリテーターの作業支援を目的として、3つ目の振り返り作業に着目する。振り返りでは、熟練作業者たちの自由闊達な議論におけるメタ知識としての話題を大まかに掴むことが求められる。そこで、時系列情報を含む発話間の類似度を用いた分析と、時系列情報を考慮しないクラスタリングを用いた分析の2つを組み合わせた抽出手法を提案し、ワークショップ会議録から単語ベースでの複数話題の抽出を試みる。提案手法と実際の会議録を用いた実験により、会議録中に含まれる話題と考え得る単語を複数抽出できることが確認できた。これにより、知識構造化ワークショップにおいてファシリテーターが議論の話題を大まかに把握できる可能性が示唆された。

キーワード: 話題抽出, k 平均法, ワークショップ, ファシリテーター, 介護

A study of topics extraction from a workshop speech record by using clustering method based on unsupervised learning

MISATO TAKITA^{†1} KOKI IJIN^{†2} MAIKO KOBAYAKAWA^{†3}
TAKUICHI NISHIMURA^{†2} YOSHIYUKI NAKAMURA^{†1}

Abstract: Against the background of the aging of skilled workers in manufacturing and nursing care field, it is required to aggregate their implicit knowledge in order to utilize for human resource development and prevention of human error. We have been practicing knowledge structuring workshops to collect and structured such implicit knowledge of skilled workers. In the workshop practice, the facilitator must perform three multiple tasks of preparation, supervision, and reflection on the spot. This paper focuses on the reflection task. In the reflection, it is essential that the facilitator roughly grabs the topics of conversation as meta-knowledge on the skilled workers' free talks. This needs much skill and patience. To support the facilitator's reflection task, this paper proposes a word-based topics extraction method from the conversation which is linked an utterance similarity analysis with time series and a clustering analysis without time series together. Using the real workshop transcriptions of conversation, experimental results show that the proposed method can extract the multiple single words that represent the topics of the discussions. The result suggests that the method can roughly grab the topics of conversation in the knowledge structuring workshop.

Keywords: Topics Extraction, k-means Clustering, Workshop, Facilitator, Nursing Care

1. はじめに

団塊世代の熟練者の定年退職による知識流出を背景に、介護現場や製造業において、熟練者の知識を集約し、人材育成[1]やヒューマンエラーの予防に活用できる技術が求められている。それらの知識流出対策の一環として、それぞれの現場では作業工程の各行為の手順の把握を目的としたマニュアルが生成されている。しかし、それらのマニュアルはその行為が果たすべき目的については記述されておらず、熟練者の判断等に関する知識を十分にカバーしきれていない。そのため、新人作業従事者は行為と順序が持つ重要性を把握できず、重要な行為を行わない、行為の順番を取り違えるといったヒューマンエラーが起きる可能性が

高い。高齢社会となった日本において、それぞれの現場における知識集約への支援が急務である。

作業工程における行為の目的を明確にする従来手法として、西村らは目的指向で行為を構造的に記述する知識表現モデル"CHARM"を提案した[2]。CHARMは、行為と目的に注目し、ある行為を、それを達成するために必要な行為の系列に分解することで、行為間の目的達成関係を明らかにしている(図1)。その結果として、各行為の目的の把握を容易にする。また、CHARMを適用する際は、作業行為に関する知識を教科書に基づいた形で記述し、作業従事者に共通知識として提供する。それにより、現場の従業員による主体的な知識の記述を促進することができる。これを知識発現として提案している[3]。

¹ 国士舘大学理工学部
Faculty of Science and Engineering, Kokushikan University
² 国立研究開発法人産業技術総合研究所
National Institute of Advanced Industrial Science and Technology

³ 千葉工業大学先進工学部
Faculty of Advanced Engineering, Chiba Institute of Technology

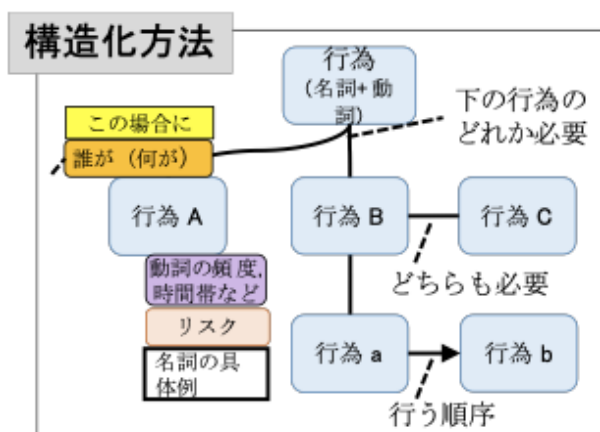


図 1 CHARM による目的指向構造化の概略図

しかし、知識構築専門家が共通知識を構築することが非効率、または困難なケースも存在する。介護現場や製造業における作業工程には独自の工程が多数あり、ある程度経験を積んだ現場の従業員でなければ共通知識の構築が難しい。例えば、介護現場における食事介助の場合、教科書には介護士と被介護者が一対一での食事介助をおこなう際の手順が記載されているが、実際の介護現場では、数人の介護士が多数の被介護者に対して食事の介助をおこなっている。このように、教科書に書き起こされた知識以上の知識が現場の作業員、すなわち熟練作業員の中に存在している。

そこで我々は、知識構築専門家による共通知識を必要とせず、作業員が自ら現場固有の知識を構造化することを可能とする新たな手法の提案と、その構造化を支援し円滑な知識の抽出をおこなう対話システムの構築をおこなっている[4]。前段階として、どのように人がワークショップで自身の持つ知識を発言するのか分析するため、人のファシリテーションによる支援のもとで、ワークショップ形式の知識構造化予備実験を実施している。この知識構造化ワークショップの結果、後述する作業フローと目的指向ツリーを組み合わせた提案手法によって、研究者による前準備の簡略化、ワークショップに参加した熟練者からの新たな目的の獲得がみられた。また、本予備実験より、熟練者の知識の表出には、ファシリテーターが重要な役割を果たしている可能性が高いことがわかった。

ここで、知識構造化ワークショップの運営に着目すると、ファシリテーターは、1) ワークショップ中の進行促進役を担うことのほか、2) 話し合いの場を設定するための事前準備、3) 集約した知識の構造化を含むワークショップ実施後の振り返りといった3つの作業をおこなう必要がある。このように、ワークショップを用いた現在の手法においては、ファシリテーターの役割が重要かつ不可欠であり、ファシリテーターにとって非常に負担の大きいものとなっている。

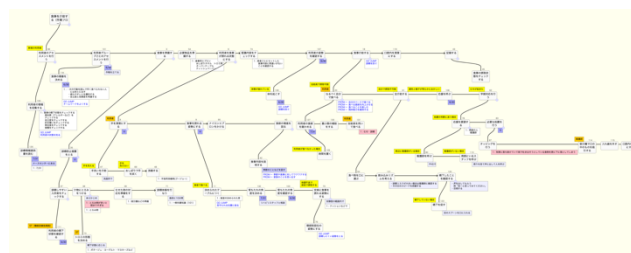


図 2 知識構造化ワークショップ実施後の振り返り作業で構築した作業フロー

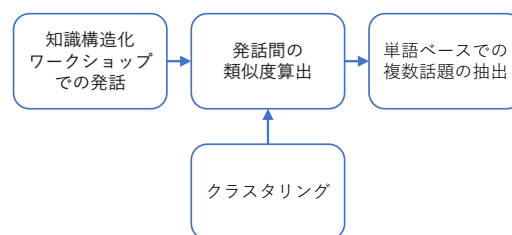


図 3 単語ベースでの複数話題抽出の流れ

特に、振り返り作業をおこなうときには、ファシリテーターは得られた知識をもとにして、最終的に図2のように集約した知識を構造化している。その際には、図1に示したようなルールに従って記述する。現在この作業は、ワークショップ実施中に記録したメモや記憶を頼りとして、全てファシリテーターによる手作業でおこなわれている。しかし、ワークショップの実施時間や、実施形態（複数のグループを並行して実施）によっては、ファシリテーターが全ての議論の流れを把握することが困難である。特に、現場主体での知識構造化を実現するためには、ファシリテーターの役割を果たせるよう、支援をおこなうことが必要不可欠である。

そこで本稿では、ワークショップ実施時に収録した動画データや、成果物を活用した知識構造化ワークショップ後の振り返り作業支援を目的とし、ワークショップ会議録からの単語ベースでの複数話題の抽出手法について検討する。具体的には、発言の書き起こしデータ（ワークショップ会議録）に着目し、時系列での発言間の類似度算出に加え、時系列を考慮しない出現単語のクラスタリングを用いることで、単語ベースでの複数話題の抽出を試みる（図3）。それらを実現するため、第2章では、既存のマニュアルから作成する作業フローと、熟練者の判断等に関する知識を記述する目的指向ツリーについて説明し、知識構造化手法の概要を記述する。第3章では、実際の知識構造化ワークショップの実施方法について述べる。第4章では、実際の知識構造化ワークショップで得たデータをもとに、ワークショップ会議録からの複数話題の抽出について検討する。第5章では、検証で得られた知見および今後の課題について述べる。第6章では、関連研究について記述し、最後に第7章でまとめる。



図 4 既存のマニュアルをもとに構築した作業フローの例

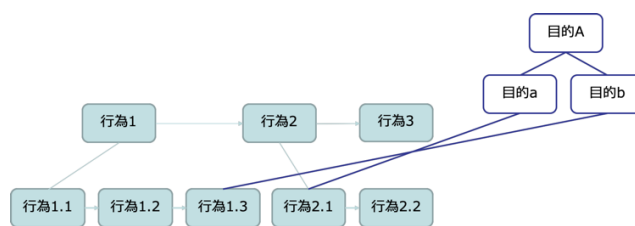


図 5 作業フローに対して目的指向ツリーを関連づけた例

2. 知識構造化手法の概要

2.1 知識構造化手法

多量の知識を容易に獲得できるように、既存マニュアルの作業工程フローチャートを活用し、現場主体によるボトムアップでの目的指向構造化知識の構築を実施する。本手法は、作業工程を時系列に並べたフローチャート（以下、作業フロー）と、作業工程にとって主とならないが、性能向上や作業の質の向上に貢献するような作業行為の目的を階層化した目的指向ツリー（以下、目的指向ツリー）の2つによって構成される。2つの形に分けることで、時系列情報の把握の容易さと、1つの行為と複数の目的との関係の表示へと対応できるようになっている。

2.2 作業フロー

作業フローは、特定の作業を達成するためにおこなうべき行為を行為ノードとして記述し、時系列順に連結・階層化した木構造の行為ノード群である。各行為ノードは、基本的に主語＋述語で記述されている。根ノードには、「特定の作業名」を記述し、下層に詳細の行為ノードを記述する。同層の行為ノードは、矢印、または直線で連結されており、矢印の場合はその順序で、直線の場合は順不同でその行為をおこなう必要がある。各ノードの階層は、上層の行為ノードを達成するためには、その行為ノードに連結された下位の行為ノードをおこなう必要があるといった、詳細な行為を記述している。これにより、作業工程の時系列情報を損なわずに情報が付与できる。また、行為ノードは、その行為の持つリスク、行為をおこなう条件、作業主体等の情報を付与できるような形となる（図1）。

作業フローの行為ノードについて、可能な限り既存のマニュアルや教科書の手順をそのまま活用するという記述ルールを設けている。具体的には、作業手順の最小単位として「ひとつの動詞」で表現されることをルールとして設定しており、複数の動詞で構成される作業手順については、動詞の数だけ行為ノードを生成している。また、作業手順の大項目・小項目をもとに階層関係を定義し、さらにその下に作業手順の行為ノードとなるようにしている（図4）。

2.3 目的指向ツリー

目的指向ツリーは、作業フローにおける階層関係と同様な記述ルールを持つが、行為の時系列情報を考慮しない木構造である。目的指向ツリーを構築する目的として、作業フローでは時系列情報を保持するために記述できなかった複数の目的を持つ行為に対して対応可能であることと、特定の作業を達成するために必要条件ではない目的を記述できることである（図5）。例として、介護現場における「食事介助をおこなう」という知識において、介護行為としての目的は、食事を被介護者に食べさせることにある。しかし、実際に介護士にヒアリングをおこなったところ、マニュアルに記述されていない、「被介護者の残存能力を維持する」等といった目的を持ちつつ、介護行為をおこなっていることが明らかとなった。これらの目的は、作業の達成には直接的に寄与しないものの、安全かつ良質な介護をおこなう上で非常に重要な要因と筆者らは捉え、またそれらが熟練介護士の持つ明文化されていない知識であると考えた。また、ある行為をおこなう際に、その小項目を達成するためには重要ではないが、のちの工程で手戻りが発生しやすくなる行為がある場合が存在し、熟練者は時系列情報の拘束がある作業フローで記述できない目的意識を持って行為に取り組んでいる可能性が高い。このような事例に対応するため、作業フローから独立した目的指向ツリーを構築することで、時系列情報を損なうことなく、熟練者の知を獲得できる形となる。

目的指向ツリーを構築する他のメリットとして、構築の容易さが挙げられる。現在の手法は、目的の階層化部分について作業フローの行為ごとに着目し、目的指向ツリーと連結することが可能であり、比較的容易に構築ができる。また、各行為に付与できる目的と目的指向ツリーの数に上限がないため、複数の熟練者から並列して知識の獲得が可能である。

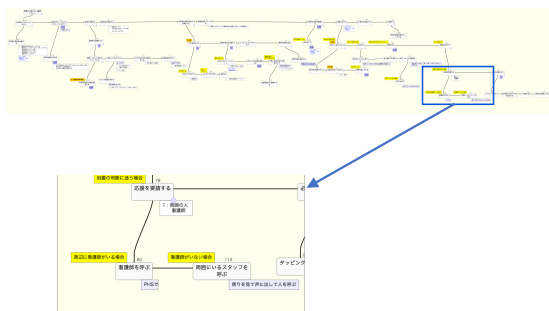


図 6 食事介助に関する作業フローの例

3. 知識構造化ワークショップの実施

現在の知識構造化ワークショップでは、熟練者を含む現場で働く数名の方に協力していただき、人のファシリテーションによる支援のもとで知識の集約をおこなっている。今回は、ある介護施設において、食事介助に関する約1時間の知識構造化ワークショップを実施した。当日は、5名の熟練介護士に協力していただき、本研究チームから2名がファシリテーターとして参加した。以下に知識構造化ワークショップの実施方法について記述する。

3.1 知識構造化ワークショップの事前準備

ワークショップの事前準備として、既存のマニュアルをもとに、作業フローの雛形をあらかじめ構築しておく。今回実験に協力していただいた介護施設は、食事介助に関する独自のマニュアルを保持していたため、作業フローの雛形はそのマニュアルをもとに構築した。構築した作業フローの雛形の例が図6である。

3.2 知識構造化ワークショップの実践

紙ベースでの知識構造化ワークショップを実施し、紙ベースの作業フローの修正と目的指向ツリーの構築を、参加者の熟練介護士5名に依頼した。事前準備で用意した作業フローをA1サイズに拡大印刷して使用している。また、本研究チームから参加した2名のファシリテーターが、ワークショップの趣旨と作業フローの説明や、熟練介護士への質問をおこなった。図7および図8は、今回のワークショップで実際に作成された作業フローと目的指向ツリーである。

3.3 ワークショップの振り返り

知識構造化ワークショップ終了後、ファシリテーターの振り返り作業として、得られた知識の整理、および作業フローと目的指向ツリーの更新を実施した。実際には、ワークショップの様子を記録した動画データや参加者の発言をテキストに起こした会議録を参照しながら振り返り作業をおこなった。そのため、約1時間の知識構造化ワークショップの振り返り作業に数日を要した。

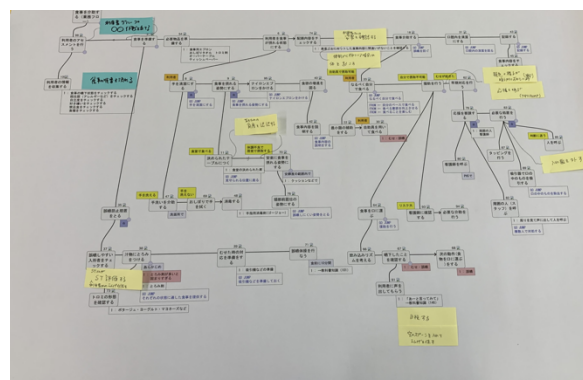


図 7 実際にワークショップで修正された作業フロー

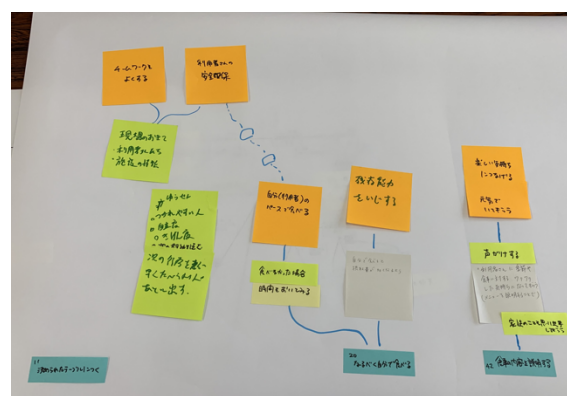


図 8 実際にワークショップで構築された目的指向ツリー

4. ワークショップ会議録からの複数話題抽出

本章では、第1章で述べた通り、ワークショップ会議録からの単語ベースでの複数話題の抽出について検討する。第3章で記述した、実際の知識構造化ワークショップで得た議事録の書き起こしデータをもとに検証をおこなう。具体的には、時系列での発話間の類似度算出に加え、時系列を考慮しないクラスタリングを用いることで、単語ベースでの複数話題の抽出を試みる。

現在は、知識構造化ワークショップ参加者の発言を手入力でテキスト化することで、ワークショップ会議録を作成している。入力として用いるワークショップ会議録は、知識構造化ワークショップの開始から終了までの参加者の発言を全て記録したものであるため、時系列情報を含んでいる。

本ワークショップ会議録で出現する発言は、比較的短い口語体によるものが多く、一般的な会議の議事録で出現する発言とは特徴が異なる。「見守りがやっぱり」、「立ち上がるとか、そういう危険な人」などがその一例である。そのため、本稿では、発言単位での分析ではなく、単語ベースでの話題検出を目指す。

実データを観察したところ、時系列順に並んだ発言において、同じ、または意味的に類似した単語が連続して出現するとき、その単語が話題のキーワードとなっており、そ

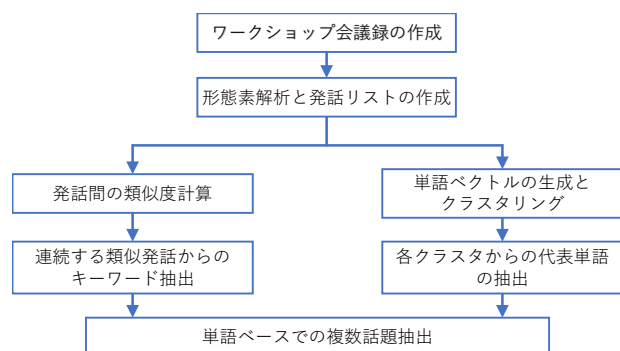


図 9 提案手法の概要

の話題についてキーワードが出現した単語の前後で議論されているケースが多かった。そのため、ここでは、時系列情報を考慮した類似単語を含む発話ペアの発見と、その類似単語がワークショップを通じてどのような意味合いを持って使われているか、ワークショップ会議録で出現した単語のクラスタリング分析を通じて明らかにする。具体的には、まず、時系列情報を含めた分析手法により、連続する類似発話ペアに共通して出現する単語の発見をおこなう。そこに、クラスタリングを用いることで、キーワードの分類と大まかな話題の推定が可能となる。図 9 に提案手法の概要を示す。

4.1 前処理

はじめに、ワークショップの動画記録をもとに発話のテキスト化をおこない、ワークショップ会議録を作成した。ワークショップ会議録には、参加者の発話のみが時系列順に記述されている。

その後、MeCab[5]による形態素解析の結果、ワークショップ会議録を 5225 単語に分割した。その中から名詞と動詞のみを取り出し、各発話の構成要素として発話リストを作成した。今回作成した発話リストには、1697 単語が含まれていた。なお、この時点では単語の重複を含んでいる。

4.2 時系列情報を考慮した分析

(1) 発話間の類似度計算

作成した発話リストをもとに、各発話に含まれる単語をそれぞれベクトル化する。本稿では、分散表現として Word2Vec[6]を用いた。Word2Vec は 2 層からなるニューラルネットワークを用いて単語の分散表現を学習する手法である。Word2Vec ではベクトルの演算をおこなうことで、単語間の意味的な関係を求めることが可能となっている。

次に、各発話に含まれる単語ベクトル同士の類似度をコサイン類似度によってそれぞれ算出する。ベクトル $\vec{x}$, $\vec{y}$ のコサイン類似度は、(1)式で計算される。2 つのベクトルのなす角度がコサイン値であり、値が大きいほど 2 つのベクトルの類似度が大きいと判断できる。

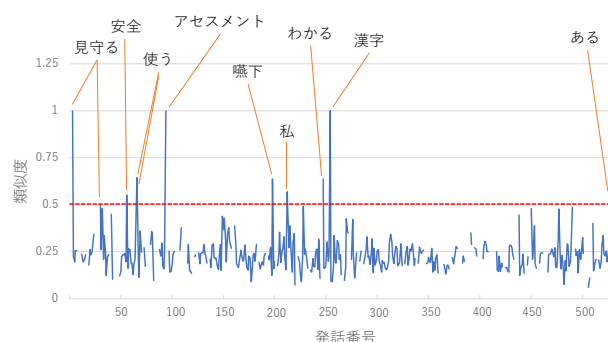


図 10 発話間の類似度（代表値：平均）

$$\cos(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}| |\vec{y}|} \quad (1)$$

連続する発話間の類似度を計算すると、複数の演算結果が出力される。演算結果が複数出力された場合には、その平均値を代表値とし、連続する発話の類似度の類似度とみなした。その結果を受け、今回は全体のペア数 533 ペア中の上位約 2 %にあたる、コサイン類似度が 0.5 以上の組み合わせを類似度が高いとみなした。図 10 に連続する発話間の類似度計算結果を示す。出現単語が全て未知語である発話との比較は空白で表示してある。

(2) 時系列情報を含んだ話題の抽出

時系列で見たときに連続している発話内で共通して出現する単語は、その近辺の議論における話題のキーワードとなっている可能性が高い。類似度を算出した結果、連続する類似発話とみなすことができた発話の組み合わせは 11 ペアあった。その 11 ペアのうち、ペア内で共通して出現していた単語を、以下に時系列順で示す。

1. 見守る
2. 見守る *時系列に抽出したため 1 と重複している
3. 安全
4. 使う
5. 使う *時系列に抽出したため 4 と重複している
6. わかる
7. アセスメント
8. 嚥下
9. 私
10. 漢字
11. ある

ここで、10 番に挙げられた「漢字」は、今回の知識構造化ワークショップのテーマであった介護現場における食事介助には、一見無関係のように思える。しかし、実際の議論の中では、作業フローに書き込んでいく場面で「介護の漢字は難しいものが多い」という話がなされていた。このことから、連続する発話間の類似度算出によりキーワード単語の抽出が可能であることがわかる。

4.3 時系列を考慮しない分析

前節では、時系列情報を考慮した連続した類似発話ペアを発見することで、キーワードとなり得る単語の抽出をおこなった。本節では、ワークショップ会議録で出現した単語のクラスタリング分析をおこなう。

(1) 単語のベクトル生成

はじめに、ワークショップ会議録に含まれる全単語のリストを作成する。今回作成した単語リスト(重複を含まず)の総単語数は 593 単語であった。この単語リストに含まれる 593 単語を全てベクトル化する。

4.2 節で発話間の類似度を算出したときと同様に、単語のベクトル生成には Word2Vec を用いた。今回、単語リストに含まれる 593 単語のうち、Word2Vec によってベクトルに書き換えることができたのは 530 単語であった。Word2Vec のモデル内に存在しない単語は、単語をベクトルに変換することができないため、この段階で 63 語が除外されている。除外されたものの一例が「誤嚥性肺炎」である。

(2) 単語のクラスタリング

ベクトル表現に書き表した 530 単語についてクラスタリングをおこなう。本稿では、クラスタリングに k-means 法を用いる。k-means 法は、非階層型クラスタリングの一種であり、クラスタリング手法として最も広く使われる手法の一つである。データの数 n 、クラスタの数を k としたとき、以下の流れでクラスタリングをおこなった。

- 各データ x_i ($i = 1, \dots, n$) に対してランダムにクラスタを割り振る。
- 割り振ったデータをもとに各クラスタの中心 V_i ($j = 1, \dots, k$) を計算する。
- 各 x_i と各 V_i との距離を求め、 x_i を最も近い中心のクラスタに割り当て直す。
- 上記の処理で全ての x_i のクラスタの割り当てが変化しなかった場合、あるいは変化量が事前に設定した一定の閾値を下回った場合に、収束したと判断して処理を終了する。そうでない場合は、新しく割り振られたクラスタから V_i を再計算して上記の処理を繰り返す。

今回の検証では、データの数 $n = 530$ 、クラスタ数 $k = 7$ と設定した。4.2 節での時系列情報を考慮した分析の結果、9 種類の単語が抽出できたので、はじめにクラスタ数を 9 に設定して検証をおこなった。しかし、今回は入力に用いるデータ数が 530 と少なく、クラスタ数が多いとうまく分類することができなかった。そのため、今回はクラスタ数を 7 に設定し、検証をおこなった。その結果を表 1 に示す。

表 1 単語のクラスタリング結果

クラスタ名	含まれていた単語の例
A	ある, 人, マニュアル, できる, 目的, 意欲, 安全, 場合, 全部, テーブル
B	ん, さん, しょう, やつ, きた, ふう, しょう, ほんま, たけ, いかん
C	食べる, 食事, とろみ, 嚥下, 既往, 摂取, 飲み込む, ポタージュ, 水分, アレルギー
D	自分, 自己流, 思う, 見守る, 言葉, 大事, 振り返る, ニュアンス, 一生懸命, 違和感
E	言う, 書く, 書き出す, 見る, 使う, 上がる, 起きる, 受け取る, 突っ込む, 呼ぶ
F	介助, 介護, ホスピタリティ, アセスメント, 病院, 教育, リハビリ, 老健, 福祉, ケア
G	そこ, 方, 辺, 入る, 部分, 位置, 向こう, 方面, 沿う, 寄り添う

(3) クラスタの重心ベクトル抽出および話題の抽出

4.2 節(2)のクラスタリングの結果をもとに、各クラスタがどのような単語群であるか決定するため、代表単語を算出する。本稿では、代表単語は、各クラスタの重心ベクトルに最も近い議事録に出現する単語とした。

Word2Vec のひとつの性質で、ベクトルから単語を求めることができるというものがある。この性質を利用して、各クラスタから求めた重心ベクトルと、意味的に類似している単語を類似度の高い順に出力した。その結果の一部を表 2 に、k-means 法によるクラスタリングのイメージを図 11 に示す。

表 2 重心ベクトルからの単語の出力結果

クラスタ名	出力した単語 (類似度の高い順)
A	もの, こと, の
B	ん, こう, もん
C	摂取, 食べる, 嚥下
D	お前, みたい, 思う
E	使う, 入れる, 作る
F	介護, 医療, 福祉
G	広がる, 見える, 辺り

クラスタの特徴として、クラスタ A では、「もの」、「こと」などといった、代名詞や一般名詞が多く含まれていた。クラスタ B には、助詞などの発話中において大きな意味を独立してもたない単語が含まれていた。クラスタ C には、食事や身体に関する単語が所属していた。クラスタ D には、名詞と動詞が幅広く含まれていた。クラスタ E, G には、一般動詞が多く含まれていた。クラスタ F には、介護に関する単語が所属していた。

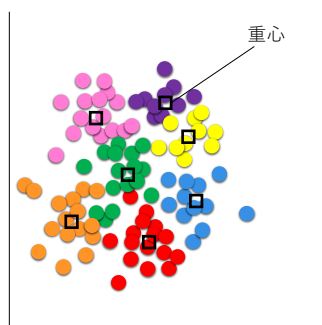


図 11 k-means 法によるクラスタリングのイメージ

4.4 単語ベースでの複数話題抽出

本節では、発話間の類似度計算結果から求めた 9 単語が、実際にどのクラスタに含まれるのかを示し、そのクラスタの代表単語との比較により、単語ベースでの話題抽出が可能であるかを検討する。前述した通り、発話間の類似度計算の結果から 9 単語、単語ベクトルのクラスタリング結果から各クラスタの代表単語としてそれぞれ 3 単語ずつを抽出できた。発話間の類似度計算の結果求めた 9 単語が所属していたクラスタについて表 3 に示す。

表 3 連続する発話間の類似度計算結果から抽出した単語が所属するクラスタ名

類似度計算結果から抽出した単語	所属するクラスタ名
安全, 漢字, ある	A(代名詞群)
嚙下	C(食事関連)
見守る, 私	D(一般動詞・名詞)
使う, わかる	E(一般動詞)
アセスメント	F(介護全般)

表 2 と表 3 を合わせて参照することで、キーワードの分類と大まかな話題の推定が可能となった。例えば、表 3 の「アセスメント」という単語は、表 2 より「介護, 医療, 福祉」のような単語に代表されるクラスタに所属していることが読み取れる。このことから、本手法によって、今回の知識構造化ワークショップの主題である食事介助中でも、食事に関する詳細な話題とは別に、介護, その中でも特に医療に関わるような話題が挙がっていたことが推定できた可能性がある。

5. 考察

知識構造化ワークショップ会議録からの、単語ベースでの複数話題の抽出について検証した結果、連続する発話間の類似度計算とクラスタリングを組み合わせた手法によって、知識構造化ワークショップにおける話題となり得る単語を複数抽出できた。しかし、今回抽出することができた単語の中には、知識構造化ワークショップにおける話題と

みなすことはできない単語も複数含まれていた。その原因について、時系列での発話間の類似度算出に関して(4.2 節), および時系列を考慮しないクラスタリングに関して (4.3 節), 以下でそれぞれ考察する。

まず、時系列での発話間の類似度算出によるキーワード抽出に関する考察を述べる。今回の検証では、各発話文に含まれる単語について個別に類似度を求め、その平均値を用いて各発話間の類似度とみなした。平均値を代表値として用いたことによって、外れ値の影響を受け、適切な代表値になり得ない部分が存在していた可能性がある。また、単語の分散表現には Word2Vec を用いた。Word2Vec のモデル作成時、コーパスとして Wikipedia のデータを使用している。この Word2Vec のモデル内に存在しない単語は扱うことができず、無視されてしまうことがある。知識構造化ワークショップでは専門用語が多数使用されることがあるため、コーパスとして用いた Wikipedia のデータに含まれておらず、扱われなかった可能性も否定できない。

次に、時系列を考慮しないクラスタリングに関する考察を述べる。単語のクラスタリングをおこなう際に、クラスタ数を変更することで所属する単語にも変化がみられるため、適切なクラスタ数を設定することがひとつの課題である。今回の教師なし学習におけるクラスタリングにおいて、総データ数は 530 であった。データ数が増加し、クラスタリングの規模が大きくなればより正確なクラスタリングをおこなうことができる可能性がある。また、前述したように Wikipedia から作られた Word2Vec を利用したが、この際に単語をベクトル化することができなかったものがあった。この点に関しては、モデルの作成時に専門用語に対応できるコーパスを用いることで、より詳細なクラスタリングを実現できる見込みがある。

また、教師なし学習におけるクラスタリングを用いることで、話題ごとのクラスタ生成がある程度達成でき、キーワードとなり得る単語の抽出に有用である可能性が示唆された。具体的には、食事介助に関するワークショップであるため、主題となる食事に関する単語が一つのクラスタに含まれていたことと、より広い観点から見た、多くの介助に共通して使われる「アセスメント」や「福祉」などといった単語は、別のクラスタに含まれており、異なる観点からワークショップを振り返る際に活用できる可能性がある。しかし、ワークショップの会議録内には一般的に用いられている動詞や名詞も多く含まれていたため、主題となる食事介助や介護全般に関する単語を詳細に分類できなかったため、改善が必要である。

時系列での発話間の類似度算出と時系列を考慮しないクラスタリングについて、それぞれ解決すべき課題は残されているものの、2 つの組み合わせによって知識構造化ワークショップにおけるキーワード分類と大まかな話題推定の可能性を示すことができたといえる。

6. 関連研究

クラスタリングを用いた時系列テキスト集合からの話題検出手法としては、福原ら[7]が、日付情報を含むテキスト集合の入力による、感情表現と用語のクラスタリングを用いる手法を提案している。具体的には、話題検出手法として以下の3点を提案している。1) ある感情について話題の推移を示すグラフを生成する手法、2) ある話題について感情の推移を示すグラフを生成する手法、3) キーワードの時間軸上の相関関係とテキスト内の共起情報を用いてキーワードをクラスタリングし話題ごとにキーワードを可視化する手法である。

クラスタリングを用いた話題推移の抽出に関する研究は、菊池ら[8]が、ニュース記事などの時系列テキスト集合を対象としておこなっている。話題の推移を表現したキーワード群を、階層型クラスタリングと複合語判定のための C-Value 法に基づいて抽出する方法を提案し、その有用性を確認している。

また、ファシリテーターの支援を目的としたひとつの手法として、分散表現を用いて自動で話題の変化を判定する手法が挙げられる。芳野ら[9]は、掲示板のようなプラットフォームをベースにした大規模意見集約支援システム COLLAGREE 上でおこなわれた議論のデータを対象とした手法を提案している。

上で述べた関連研究は、手法にクラスタリングを用いる点やファシリテーターの支援を目的としている点においては本稿と類似している。しかし、本稿では書き言葉ではなくワークショップ会議録という話し言葉がもとになったテキストを対象としているという点で異なっている。これによって文法のぶれや方言の影響を抑制することが課題となった。この課題に対し、影響を受けにくい名詞と動詞という単語に着目してそれらをベクトル化した上でクラスタリングをおこない、今度は逆に、各クラスターの重心ベクトルを求めてシソーラス中における最近傍単語を求める手法を提案し、話題を単語ベースで抽出している。

7. まとめ

本稿では、熟練者の持つ知識の効率的な集約と構造化を目的として実施している、現在の知識構造化ワークショップについて説明し、ワークショップ会議録からの単語ベースでの複数話題抽出手法、およびその検証と結果について述べた。検証の結果、時系列情報を含む発話間の類似度を用いた分析と、時系列情報を考慮しないクラスタリングを用いた分析を組み合わせることで、ワークショップ会議録から話題と考え得る単語を複数抽出することができた。それにより、単語ベースではあるものの、ファシリテーターが知識構造化ワークショップにおける議論の話題を大まかに把握するための支援の可能性を示唆することができた。

しかしながら、現手法には考察で述べたように残された課題も多く、知識構造化ワークショップにおける正確な話題の把握に至ったとはいえない。今後の課題として、まずは発話間の類似度計算による単語抽出の精度とクラスタリングの精度を高めることが挙げられる。

以上、本稿は未だ課題が残されているが、知識構造化ワークショップにおける複数の話題把握に関して、ひとつの方法論を示すことができた。これにより、本稿の目的であるファシリテーターの知識構造化ワークショップ後の振り返り作業の支援を実現できる可能性がある。今後は、上記の課題解決を進めるとともに、知識構造化ワークショップにおけるファシリテーター支援によって、円滑な知識の集約と構造化を目指す。

謝辞

本研究をおこなうにあたって知識構造化ワークショップ等にご協力いただいた社会医療法人財団董仙会に感謝申し上げます。

参考文献

- [1] 経済産業省: 2018 年版ものづくり白書 (2018).
- [2] Nishimura, S., Kitamura, Y., Sasajima, M., et al.: CHARM as Activity Model to Share Knowledge and Transmit Procedural Knowledge and its Application to Nursing Guidelines Integration, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol.17, No.2, pp.208-220 (2013).
- [3] 西村悟史, 毛利陽子, 山中泉ほか: 社会福祉法人内の介護マニュアル統一を通じた知識発見の改良, 情報処理学会デジタルプラクティス, Vol.10, No.1, pp.244-266 (2019).
- [4] 伊集院幸輝, 小早川真衣子, 西村拓一: 現場主体による目的指向構造化知識の手法と対話システムによる支援の検討, 知識・技術・技能の伝承支援研究会資料 (2019).
- [5] “MeCab: Yet Another Part-of-Speech and Morphological Analyzer”. <https://taku910.github.io/mecab/>, (参照 2019-12-13).
- [6] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean: Distributed Representations of Words and Phrases and their Compositionality, *Proc. of NIPS*, pp.3111- 3119 (2013).
- [7] 福原知宏, 中川裕志, 西田豊明: 感情表現と用語のクラスタリングを用いた時系列テキスト集合からの話題検出, 第 20 回人工知能学会全国大会 (2006).
- [8] 菊池匡晃, 岡本昌之, 山崎智弘: 階層型クラスタリングを用いた時系列集合からの話題推移抽出, データ工学ワークショップ DEWS (2008).
- [9] 芳野魁, 伊藤孝行: ファシリテーター支援を目的としての分散表現を用いた話題変化判定, 第 32 回人工知能学会全国大会 (2018).