

一般口演 | 知識工学

## 一般口演9

### 自然言語処理・テキストマイニング

2021年11月20日(土) 09:10 ~ 11:10 E会場 (2号館2階222+223)

#### [3-E-1-04] Fuzzy-C-Means法を用いた痛みに関する看護用語の数値化

\*韓 豊<sup>1</sup>、張 孜恒<sup>1</sup>、大橋 和貴<sup>1</sup>、張 洪健<sup>1</sup>、小笠原 克彦<sup>1</sup>（1. 北海道大学大学院保健科学院）

\*Feng Han<sup>1</sup>, ZiHeng Zhang<sup>1</sup>, kazuki oohasi<sup>1</sup>, HongJian Zhang<sup>1</sup>, katsuhiko ogasawara<sup>1</sup>（1. 北海道大学大学院保健科学院）

キーワード：Fuzzy C Means, Natural Language Processing(NLP), Fuzzification, Nursing terminology, Pain Assessment

背景：“痛み”は患者の主観的な感覚であり、その表現も多岐にわたるため、コンピューターが分析するのが難しいだけでなく、人間が直感的に理解することも難しい。近年、医療支援システムにAIが導入され、精度が大幅に向上した。医療支援システムの精度をさらに向上させるため、コンピューターに痛みに関する表現を分析させ、数値化させる必要がある。目的：本研究では、医療支援システムが看護用語の痛みに関する表現から痛みの概念を共有できるようにするため、Fuzzy-C-MeansおよびGaussianメンバーシップ関数を使用し、痛みの表現を数値化する。方法：PubMedから収集された335,367件の看護関連の「abstract」とした。335367件のAbstractのうちキーワードを含むものが199764件であった。次に、word2vec法を使用して形態素解析を行って単語ベクトルに変換した。ベクトル化された「abstract」をFuzzy-C-Means法を使用してキーワードによって痛みを表す単語を分類し、抽出した。最後に、Gaussianメンバーシップ関数を使用して、単語を数値化した。結果：数値化された形式は範囲値であり、代表値に近いほど、その分類の所属値（可能性）が高くなる。以下では、「pain」の例を挙げて、各キーワードに対応する代表値の結果を示す：Phantom pain(pain:2.2729), Pain control(pain:1.4710), Increased pain(pain:2.3849), Reduced pain(pain:2.4692), Acute pain(pain:2.5735), Chronic pain(pain:2.3522), Abdominal pain(pain:2.1726)。

# Fuzzy-C-Means 法を用いた痛みに関する看護用語を数値化

韓 豊<sup>\*1</sup>、張 孜恒<sup>\*1</sup>、大橋 和貴<sup>\*1</sup>、張 洪健<sup>\*1</sup>、小笠原 克彦<sup>\*1</sup>、

<sup>\*1</sup> 北海道大学大学院 保健科学院、

## Quantify nursing terms related to pain by using the Fuzzy-C-Means method

Feng Han<sup>\*1</sup>, ZiHeng Zhang<sup>\*1</sup>, kazuki oohasi<sup>\*1</sup>, HongJian Zhang<sup>\*1</sup>, Katsuhiko Ogasawara<sup>\*1</sup>,

<sup>\*1</sup> Graduate School of Health Sciences, Hokkaido University,

There are still subjective words that are difficult for humans to understand intuitively in the standardized terminology used in nursing. At present, the accuracy of using self-evaluation scales is low, and it is not suitable for computers. We propose to use the Fuzzy-C-Means method and Gaussian membership function to allow the computer to analyze and quantify the word of pain in the nursing text to improve the quantification accuracy. Further improve the accuracy of the medical support system. We use the Fuzzy-C-Means method to classify the 199764 data collected by PubMed by keywords. Then extract the pain words in the category of keywords, and finally use the Gaussian membership function to quantify the extracted pain intensity words. The quantified form is a range value. The closer to the midpoint value, the greater the degree of membership. Out of range is 0. The following are the results of the pain range corresponding to some keywords. Phantom pain (pain: 2.2729), Bone pain (pain: 2.417), Cancer pain (pain: 2.5692). It has been proved that the Fuzzy-C-Means method and Gaussian membership function can quantify subjective words with high precision. Further research is needed to evaluate the impact of using this model on the medical support system.

**Keywords:** Fuzzy-C-Means, natural language processing (NLP), fuzzification, Nursing terminology, Pain Assessment

### 1. 結論

統一された看護用語は、看護師が患者のニーズと感情をより正確に表現するのに役立つ。さらに、標準化された用語を使用すると、看護師間や他の医療専門家との情報共有に貢献する。また、標準化された用語は、電子健康記録に特定の構造を提供し、看護記録の精度を向上させるのに役立つ。しかし、標準化された看護記録で痛みの強さに表す単語などの主観的な言葉がまだあり、そのような単語は人間が直感的に理解することも難しく、コンピューターが分析することはさらに困難である。

また、医療支援システムに AI が導入され、精度が大幅に向上した。医療支援システムの精度をさらに向上させるためには、コンピューターに痛みに関する表現を分析させ、数値化させる必要があります。現在、痛みの強さを数値化するために、自己評価尺度 (Visual Analogue Scale や Numerical Rating Scale など) が使用されている。これらの数値化の方法は、コンピューターの精度が低く、コンピューターの使用には適していない。

そこで本研究では、看護記録で痛みの概念をよりよく共有し、医療支援システムが痛みの強さをさらに分析できるようにするため、Fuzzy-C-Means および Gaussian メンバーシップ関数を使用し、痛みの強さを数値化することを提案する。

### 2. 方法

この研究では、Pubmed によって収集された 335,367 件の看護関連の abstract を使用した。最初に、335,367 件の抄録が、International Classification for Nursing Practice (ICNP) によって収集された 20 の痛みに関連する用語に従って分類された。また、処理したデータに対し、特徴抽出方法である Word2Vec を使用し、自然言語をコンピューターで処理できる単語ベクトルに変換した。

次に、クラスタリングのグループ数を決める。最適なグループ数を決めるために、分別に 2, 3, 4 グループを対象として (グループの単語の差、グループの中心値の平均距離と各単語と中心点の平均距離) 3 つの指標を使用して検討した。さらに、グループの違いを直感的に観察できるようにするために、主成分分析 (PCA) 法を使用してデータの次元数を 3 次元に縮小した。また、Fuzzy-C-Means 法で痛みの強さに表す単語を抽出する。そして、抽出された単語のベクトルを対象として中心値を計算した。

最後に、Gaussian メンバーシップ関数を使用して、クラスタリングで計算の中心値と単語と中心値の距離に基づいて数値化の結果を計算して数値範囲に変換した。

### 3. 結果

#### 3.1 クラスタリング

結果が多いため、Abdominal pain の一部結果を例として示す。Abdominal pain に対するグループ数の指標を表 1 に示す。グループ数が 2 の時は、他のグループ数より各単語と中心点の平均距離が長かった。抽出結果の一部を表 2 に示す。

表 1 Abdominal pain の対するグループ数の指標

| グループ数 | グループの単語の差 | 中心点の最小距離 | 各単語と中心点の平均距離 |
|-------|-----------|----------|--------------|
| 2     | 0.124     | 9.2E-05  | 1.085115922  |
| 3     | 0.080     | 5.4E-05  | 1.062938298  |
| 4     | 0.376     | 4.14E-05 | 1.02158934   |

表 2 Abdominal pain の一部抽出結果

| 痛みに関連する用語      | 抽出結果 (前 10 個)                                                              |
|----------------|----------------------------------------------------------------------------|
| Abdominal pain | ache, pain, agony, sore, discomfort, stab, bad, unbearable, acute, chronic |

### 3.2 数値化

Gaussian メンバーシップ関数で各抽出された単語を数値範囲に変換した。単語全体の数が多いため、Pain の例を挙げて結果と出力する特徴を表 3 に示す。図 1 は、例 (Abdominal pain の pain) のメンバーシップ図である。中心点からの距離が遠くなると、メンバーシップの度合いは低下した。メンバーシップ値の範囲は [0, 1] であり、中心点が最大値、1.800 以下と 1.950 以上が最小値であった。

表 3 各用語の Pain の中心点の結果

| 痛みに関連する用語      | Pain に対する中心点 | 痛みに関連する用語                    | Pain に対する中心点 |
|----------------|--------------|------------------------------|--------------|
| Abdominal pain | 1.8710       | Increased pain               | 2.3849       |
| Acute pain     | 2.5735       | Ischemic pain                | 3.7298       |
| Arthritis pain | 2.1189       | Muscle pain                  | 2.4125       |
| Assessing pain | 0.8324       | Musculoskeletal pain         | 1.8889       |
| Bone pain      | 2.4170       | Neurogenic pain              | 1.6440       |
| Breast pain    | 2.7685       | Pain control                 | 1.4710       |
| Cancer pain    | 2.5692       | Pain management/m medication | 1.5148       |
| Chronic pain   | 2.3522       | Phantom pain                 | 2.2729       |
| Cutaneous pain | 2.3989       | Reduced pain                 | 2.4692       |
| Fracture pain  | 2.1899       | Wound pain                   | 2.3522       |

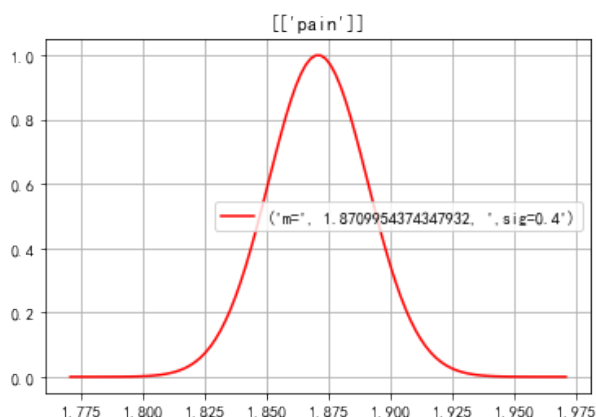


図 1 Abdominal pain の pain のメンバーシップ図

### 4. 考察

看護用語に関する研究は、ほとんど既存の用語または分類システム (Unified Medical Language System (UMLS) Systematic Nomenclature of Medicine-Clinical Terms (SNOMED-CT)) を使用し、各地域や分野での応用可能性を検証した<sup>[1]</sup>。Kang らは、専門家を招集して看護用語を評価することにより、患者の状態を予測した<sup>[2]</sup>。

さらに、非構造化テキストからの医学的知識の抽出、医学的予測および診断に関する多くの研究もある。Hashem らは、サポートベクターマシンモデル (SVM) を使用し、University of California at Irvine (UCI) リポジトリで利用可能な 2 つのデータセットを分類し、最高ランクの機能セットに基づいて肝疾患を評価した<sup>[3]</sup>。Albion らは、サポートベクターマシンモデル (SVM)、Fuzzy-C-Means (FCM)、およびランダムフォレスト (RF) の組み合わせ方法を使用し、ICU 患者のリスクレベルを予測した<sup>[4]</sup>。

Mostafa らは、Fuzzy-C-Means (FCM) とファジーメンバーシップ関数を使用して肝疾患を予測した<sup>[6]</sup>。

以上の研究ではフリーテキストや看護用語の研究を行ってきたが、看護言語を標準としてフリーテキストから痛みに関連する単語を抽出する研究はない。

本研究では、FCM やガウスメンバーシップを用い、痛みに関連する単語を対象として抽出し、数値化を行う。この方法は、専門家の知識がなくても作成でき、主観的な単語を高精度で数値化できる。明確な結論を引き出すために大規模なデータセットの検証が必要ですが、これらの数値データは、人が痛みに関連する用語とコンピューター分析を理解するのに役立つ可能性がある。また、医療診断やリスク予測などの後続の自動化されたタスクをサポートするために使用できる。

### 5. 結論

本研究では、痛みに関連する単語を数値化する方法を提案する。この方法は、専門家の知識がなくても作成でき、主観的な単語を高精度で数値化できる。この方法は、特に専門家がいない場合に便利で高速である。

正確性について明確な結論を出すには、さまざまな病気の大規模なデータセットが必要であるが、本研究は、主観的な単語を高精度で定量化し、医療支援システムで役立つようにするのに役立つ可能性がある。

### 参考文献

- 1) Gillian Strudwick, Nicholas R. Hardiker, Understanding the use of standardized nursing terminology and classification systems in published research: A case study using the International Classification for Nursing Practice®, International Journal of Medical Informatics, Volume 94, 2016, ISSN 1386-5056,
- 2) Kang MJ, Dykes PC, Korach TZ, Zhou L, Schnock KO, Thate J, Whalen K, Jia H, Schwartz J, Garcia JP, Knaplund C, Cato KD, Rossetti SC. Identifying nurses' concern concepts about patient deterioration using a standard nursing terminology. Int J Med Inform. 2020 Jan;133:104016. doi: 10.1016/j.ijmedinf.2019.104016. Epub 2019 Oct 31. PMID: 31707264; PMCID: PMC6957124.
- 3) Hashem EM, Mabrouk MS. A Study of Support Vector Machine Algorithm for Liver Disease Diagnosis. Am J Intell Syst. 2014;4(1):9-14.
- 4) Albion Dervishi, Fuzzy risk stratification and risk assessment model for clinical monitoring in the ICU, Computers in Biology and Medicine, Volume 87, 2017, Pages 169-178, ISSN 0010-4825,