

修士論文

介護マッチングサービスにおけるマッチング精度向
上の提案と実装

Proposal and Implementation for Improving Matching Accuracy in Elderly Care Matching
Services

北海道大学 大学院情報科学研究院
メディアネットワーク専攻 情報メディア環境学研究室

星子 祐哉

目次

第 1 章	序論	1
第 2 章	研究の目的	3
第 3 章	提案手法	4
第 4 章	実験	6
第 5 章	考察	9
第 6 章	ユーザースタディ	10
第 7 章	今後の展望	11
第 8 章	結論	12
	謝辞	13
	参考文献	14
	参考文献	14
	研究業績	15

図目次

1.1	介護特化型マッチングアプリの概要	2
-----	----------------------------	---

表目次

4.1	BERT モデル別の分離度スコア	6
4.2	本実験 1：ワーカー別の推薦精度	7
4.3	本実験 2（20 件学習）の上位 5 パターン	7
4.4	本実験 2（30 件学習）の上位 5 パターン	7
4.5	学習データ数による精度推移（Top-5 精度）.	8
4.6	実験 3 の結果（概略）.	8

概要

近年、日本の介護業界は深刻な人材不足に直面している。厚生労働省の資料によると、2040 年までに約 69 万人の介護人材が不足する見込みである。一方で、有資格でありながら介護職に従事していない「潜在介護士」が約 37 万人存在し、供給可能な人材が活用されていない現状がある。さらに、介護事業者は応募者が集まらない、採用コストの高騰、採用後の定着率の低さといった複合的な課題を抱えており、人材のミスマッチが常態化している。

本研究では、介護特化型マッチングアプリにおけるマッチング精度を向上させるための手法を提案し、実装を行った。現状のマッチングシステムは「勤務地」「日程」「資格要件」などの定型条件マッチが中心であり、希望職種や過去の職務経験と求人内容の意味的な一致は評価されていない。そのため、ワーカーと求人の潜在的な親和性が見逃される可能性が存在する。

提案手法では、自然言語処理モデルを利用した距離計算により、職種の距離（希望職種と実際の職種の一致度）とスキルの距離（職務経歴と求められているスキルの類似度）を測定する。また、時間的距離と勤務時間、過去の応募履歴などの要素を組み合わせる重み付けを行う。実験では、日本語 BERT モデルを含む複数の自然言語処理モデルを横断的に評価し、日本語 Sentence-BERT モデルが最も優れた分離度（0.398）を示すことを確認した。さらに、過去の求人応募データから新規求人との類似度を計算する手法を検証し、BERT による意味的類似検索の有効性を示した。

本研究により、介護マッチングサービスにおいて、ワーカーの希望とより適合した求人を効率的に提示できるシステムの基盤が構築された。今後は、実データを用いた検証と、複数の手法を組み合わせたより高精度なマッチングアルゴリズムの開発、実サービスへの実装とユーザーインタビューを通じた改善サイクルの確立が課題である。

第 1 章

序論

本研究は、介護特化型マッチングサービスにおけるマッチング精度の向上を目的とする。背景として、2040 年に向けた介護人材不足が深刻化しており、厚生労働省の推計では約 69 万人の人手不足が見込まれている。一方で、有資格者でありながら介護職に従事していない「潜在介護士」が約 37 万人存在し、供給可能な人材が十分に活用されていない。

介護業界が直面する課題

介護業界は以下に示す複合的な課題を抱えている。

■**人材ミスマッチと経営危機** 介護事業者は、応募者が集まらない、採用コストの高騰、採用後の定着率の低さといった複合的な課題に直面している。人材のミスマッチが常態化しており、人材不足によるサービス低下や廃業に至るケースも増加している。これらの問題は単なる人手不足だけでなく、ワーカーと求人の適切なマッチングが実現できていないことに起因する。

■**北海道の特殊事情** とくに本研究の対象としている北海道は高齢化の進行と人口偏在が顕著であり、2025 年には高齢化率が約 33.5% に達すると予想される。人口の約 3 分の 2 が札幌市付近に集中しており、他の市町村は広大な地域の中で点在・過疎化している。移動距離も長く、他都府県と比較してハンデが大きい状況である。

■**現行マッチングシステムの限界** 現状のマッチングシステムは「勤務地」「日程」「資格要件」などの定型条件マッチが中心であり、希望職種や過去の職務経験と求人内容の意味的な一致は評価されていない。そのため、ワーカーと求人の潜在的な親和性が見逃される可能性が存在する。また、ワーカーが自ら細かい条件を指定して検索する手間が大きく、望む仕事を見つけられないという課題がある。

介護マッチングアプリの仕組み

本研究で対象とする介護特化型マッチングアプリの概要を図 1.1 に示す。このシステムでは、介護事業者が求人情報を登録し、ワーカー（潜在介護士、副業希望者、その他就労希望者等）が希望の勤務地・時間・職種等を入力することで、システムが自動的にマッチングを行う。

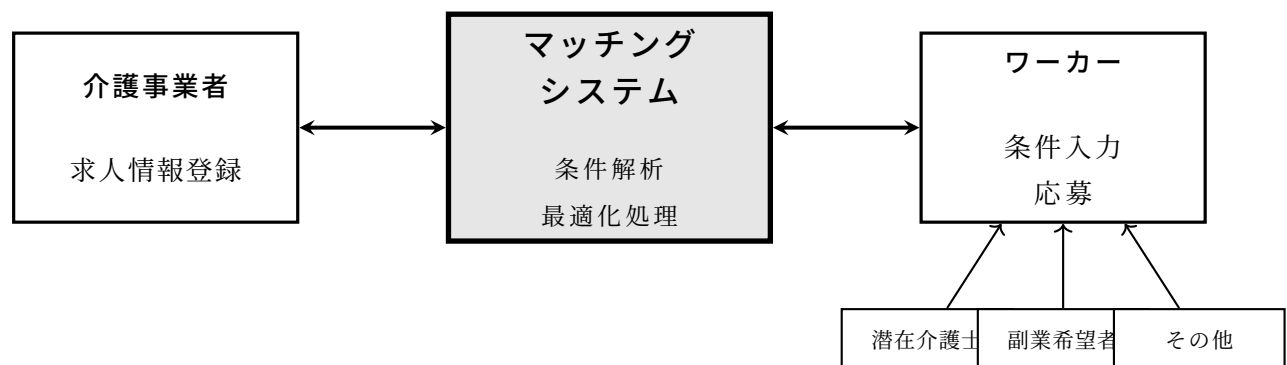


図 1.1: 介護特化型マッチングアプリの概要

これらの課題を踏まえ、本研究では単純な定型条件マッチ（勤務地・日程・資格等）だけでなく、テキストの意味的類似性や履歴情報、時間的要因を組み合わせた重み付けにより、ワーカーと求人の潜在的親和性をより高精度に評価する手法を提案する。

第 2 章

研究の目的

本研究の目的は、ワーカーと求人のマッチングにおける「意味的な親和性」を捉え、ワーカー側・事業者側双方の検索・推薦の手間を最小化することである。従来のマッチングは主に勤務地・日程・資格等の定型条件で行われており、希望職種や過去の職務経験と求人内容の意味的な一致は十分に評価されていない。

その結果、潜在的に適合するワーカーと求人が見逃されることがある。また、ワーカーが自ら細かい条件を指定して検索する手間が大きい点も課題である。本研究では自然言語処理によるベクトル化、履歴情報、時間的重み付けなどを組み合わせることで、これらの課題を解決する手法を検討する。

第 3 章

提案手法

本研究では以下の要素を組み合わせた重み付けを提案する。

自然言語処理モデルを利用した距離計算

求人タイトルや業務内容、職務経歴などのテキストを自然言語処理モデルでベクトル化し、文書間の距離（類似度）を計算する。これにより職種の距離やスキルの距離を定量化できる。

ベクトル化とコサイン類似度

自然言語処理モデル（BERT や TF-IDF 等）を用いて、テキストを高次元ベクトル空間に埋め込む。具体的には、求人タイトルや業務内容などの各テキストフィールドをモデルに入力し、固定長のベクトル表現を得る。

テキスト間の類似度評価には、ベクトル化された表現間のコサイン類似度を用いる。2 つのベクトル \mathbf{a} と \mathbf{b} のコサイン類似度は以下で定義される：

$$\cos(\theta) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|} = \frac{\sum_{i=1}^n a_i b_i}{\sqrt{\sum_{i=1}^n a_i^2} \sqrt{\sum_{i=1}^n b_i^2}}$$

ここで、 $\mathbf{a} = (a_1, a_2, \dots, a_n)$ および $\mathbf{b} = (b_1, b_2, \dots, b_n)$ は n 次元ベクトル、 $\mathbf{a} \cdot \mathbf{b}$ は内積、 $\|\mathbf{a}\|$ および $\|\mathbf{b}\|$ はそれぞれのベクトルのユークリッドノルムを表す。

コサイン類似度は $[-1, 1]$ の範囲の値を取り、1 に近いほど類似度が高く、0 は直交（無関係）、-1 は完全に反対の方向を示す。テキスト解析では通常、非負の値域 $[0, 1]$ で扱われる。

行列間の類似度計算

複数のテキストフィールド（例：求人タイトル、業務内容、介護項目など）を持つ文書全体の類似度を計算する場合、各フィールドをベクトル化した後、それらを行列として扱う。

文書 A のフィールドベクトルを $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ 、文書 B のフィールドベクトルを $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m$ とする。このとき、行列間の類似度は以下の手順で計算される：

1. 各フィールドペア $(\mathbf{a}_i, \mathbf{b}_i)$ のコサイン類似度 s_i を計算
2. 各フィールドに重み w_i を付与し、加重平均を求める：

$$\text{Similarity}(A, B) = \frac{\sum_{i=1}^m w_i \cdot s_i}{\sum_{i=1}^m w_i}$$

この手法により、複数の観点から文書間の類似性を総合的に評価できる。重み w_i は実験的に調整され、本研究では7つの項目（求人タイトル、業務内容、介護項目、施設、勤務時間、時給、勤務日）に対して様々な重み付けパターンを検証した。

時間的距離と勤務条件の差分

勤務地や勤務時間、勤務日などの構造化された条件に基づき時間的・地理的な差分を計算し、ベクトル類似度と組み合わせることで総合スコアを算出する。

過去応募履歴の活用

過去の応募履歴や承認履歴を用いてユーザーの志向性をモデル化する。学習データ（過去 N 件）から平均ベクトルを算出し、新しい求人との類似度を評価する手法を本研究では実験的に採用した。

第 4 章

実験

本節では本研究で実施した予備実験および本実験の設計と結果を示す。

予備実験：BERT モデルの横断評価

複数の BERT 系モデル（日本語 Sentence-BERT、マルチリンガル Sentence-BERT、BERT-large など）を比較し、職種間の分離度を評価した。分離度は類似職種グループと非類似職種グループの平均スコア差で定義する。

表 4.1: BERT モデル別の分離度スコア

モデル	分離度スコア
Sentence-BERT (Japanese)	0.398
Sentence-BERT (Multilingual)	0.349
BERT-large (Mean Pooling)	0.111
BERT-large (CLS)	0.063
BERT (Mean Pooling)	0.039
BERT (CLS)	0.019

本実験 1：過去の応募履歴からの推定

学習データとして過去 30 件、テストデータ 10 件を用い、過去応募の平均ベクトルから次に選択される求人を推測した。評価指標として Top-1/Top-3/Top-5 精度および複合スコアを算出した。

多く働いているユーザー（ワーカー A、B）は 100% の精度で予測できるが、利用頻度が低いユーザー（ワーカー I、J）は精度が大幅に低下する傾向が見られた。

表 4.2: 本実験 1：ワーカー別の推薦精度

順位	ワーカー	Top-1	Top-3	Top-5	複合スコア
1	ワーカー A	100.0%	100.0%	100.0%	1.0000
2	ワーカー B	100.0%	100.0%	100.0%	1.0000
3	ワーカー C	60.0%	80.0%	80.0%	0.7000
4	ワーカー D	60.0%	60.0%	60.0%	0.6000
5	ワーカー E	20.0%	80.0%	80.0%	0.5000
6	ワーカー F	10.0%	80.0%	80.0%	0.4500
7	ワーカー G	20.0%	30.0%	40.0%	0.2700
8	ワーカー H	10.0%	30.0%	60.0%	0.2600
9	ワーカー I	0.0%	30.0%	40.0%	0.1700
10	ワーカー J	10.0%	10.0%	30.0%	0.1400
全体平均 (10 名)		39.0%	60.0%	67.0%	0.5090

本実験 2：複数項目の重み付けによる精度向上

求人タイトル、業務内容、介護項目、施設、勤務時間、時給、勤務日など 7 項目を抽出し、各項目に重みを付与して評価した。以下に一部結果を示す。

表 4.3: 本実験 2 (20 件学習) の上位 5 パターン

順位	パターン名	Top-1	Top-3	Top-5	複合スコア
1	均等 (equal)	57.1%	70.6%	77.1%	0.7112
2	介護・施設重視	58.2%	68.2%	77.6%	0.7094
3	勤務時間重視	52.9%	68.2%	77.1%	0.6959
4	時給 × 勤務時間強化	54.1%	69.4%	75.9%	0.6959
5	時給重視	51.8%	69.4%	76.5%	0.6941

表 4.4: 本実験 2 (30 件学習) の上位 5 パターン

順位	パターン名	Top-1	Top-3	Top-5	複合スコア
1	総合重視	30.0%	56.0%	63.0%	0.5430
2	時給 × 勤務時間強化	38.0%	49.0%	63.0%	0.5380
3	業務 × 介護 × 時給	32.0%	52.0%	61.0%	0.5250
4	求人タイトル重視	38.0%	51.0%	59.0%	0.5240
5	実務重視	26.0%	55.0%	60.0%	0.5170

学習データ数を変化させた場合の精度推移では、学習データが 15 件～20 件の時点で最高値を示し、それを超えると精度が悪化する傾向が観察された。

上記の表から、20 件を超えると全てのパターンで精度が低下することが明確に示された。これは過去の嗜好データが現在の嗜好と乖離するためと考えられる。

表 4.5: 学習データ数による精度推移 (Top-5 精度)

パターン	5 件	10 件	15 件	20 件	25 件	30 件
均等 (equal)	0.69	0.62	0.74	0.77	0.62	0.56
介護・施設重視	0.64	0.64	0.77	0.77	0.61	0.57
時給 × 勤務時間強化	0.68	0.64	0.72	0.76	0.62	0.63
高バランス型	0.70	0.62	0.72	0.74	0.60	0.56
求人タイトル重視	0.68	0.67	0.72	0.70	0.59	0.59

実験 3：TF-IDF と BERT の比較

単語出現頻度に基づく TF-IDF、ルールベース（単語一致）、および BERT による意味的類似度を比較した結果、短文やタイトルレベルの比較では TF-IDF が実用上十分な精度を出す場合があり、計算コストの点で有利であることがわかった。

表 4.6: 実験 3 の結果（概略）

手法	精度	特徴
ルールベース（単語一致）	0.8	シンプルだが限界あり
BERT（意味的類似度）	1.0	最高精度だが計算コスト高
TF-IDF（頻度ベース）	1.0	計算コスト低で有効な場合あり

第 5 章

考察

本研究の実験結果から導かれる主要な考察を以下に示す。

- **利用頻度による精度差**：多く働いているユーザー（データが豊富なユーザー）とそうでないユーザーで応募傾向が大きく異なる。頻繁に利用するユーザー（ワーカー A、B）は 100% の精度で予測できるのに対し、利用頻度が低いユーザー（ワーカー I、J）では複合スコアが 0.14～0.17 と大幅に低下する。これはユーザーの利用頻度に応じた重み付け調整の必要性を示唆する。
- **最適学習データ量**：学習データ量については、15 件～20 件程度が最適帯であることが判明した。20 件学習時の最高複合スコアは 0.7112（均等パターン）であるのに対し、30 件学習時は 0.5430（総合重視パターン）まで低下する。これは約 24% の精度低下に相当する。過度に過去のデータを学習すると、古い嗜好が現在の嗜好を汚染し精度が低下するためと考えられる。したがって、時間的な重み付け（古いデータを低減する等）を導入することが有効である。
- **重み付けパターンの選択**：20 件学習では「均等 (equal)」パターンが最良であったが、30 件学習では「総合重視 (comprehensive)」が最良となる。これは学習データ量によって最適な重み付けが変化することを示しており、動的な重み調整が望ましい。
- **TF-IDF の実用性**：実験 3 において TF-IDF は計算コストが低く、タイトルや短文の比較では BERT に匹敵する実用的な精度 (1.0) を示した。リアルタイム性が要求される UX では、TF-IDF による高速フィルタリング後に BERT で精密ランキングを行うハイブリッド戦略が有効である。

提案：ユーザーの利用頻度に基づく重み付け、時間的な重み付け（直近 15～20 件を重視）、および TF-IDF と BERT を組み合わせたハイブリッド戦略を検討することが望ましい。

第 6 章

ユーザースタディ

本研究では候補者の選定とインタビューを実施し、ユーザースタディから得られた示唆を整理した。候補者はカスタマーサービスと連携して 13 名を選定し、アンケートとインタビュー項目を用意した。現段階では 30 分程度のインタビューを 2 件実施した。

得られた知見として、以下の点が重要である。

1. 過去にその業務で働いたことがあるか（経験の有無）は高い重要度を持つ。
2. 駐車場の有無など通勤の利便性が意思決定に影響する。
3. 特定の業務（例：排泄介助など）を避けたいかどうかは、時給よりも優先される場合がある。

これらの結果は、スコアリングにおける項目別重みの設計に直接反映させるべきである。特に業務の可否や通勤利便性は高い重みを与えるべき項目と考えられる。

第 7 章

今後の展望

今後の研究・実装に関する方向性を以下に示す。

1. 本実験 2 で示唆された重み付けの更なる最適化：ユーザーの利用頻度や応募の時間的な重みを組み込み、学習データの選択基準を改善する。
2. 実験 3 の比較検討の拡充：TF-IDF と BERT の利点を組み合わせたハイブリッド手法や、より多くの日本語モデルを横断的に評価する。
3. 実務実装に向けたインタビューと A/B テスト：実際のサービスに実装し、ユーザー行動の変化を観察し重み付けサイクルを改善する。
4. 実験 2,3 の手法を統合した高精度マッチングアルゴリズムの開発：複数の類似度評価を総合してランキングを生成するフレームワークの確立。

第 8 章

結論

結論

本研究では、介護マッチングにおける意味的類似性の導入と複数項目を組み合わせた重み付けが、実務的に有用なマッチング精度を達成し得ることを示した。予備実験では日本語 Sentence-BERT が高い分離度を示したが、実装上のトレードオフを鑑みると TF-IDF も有力である。

実験では学習データ量や重み付けパターンにより精度が変動し、特に 15 件～20 件程度の学習データで良好な結果が得られた。これは過去データの扱い（時間的重み付け）の重要性を示唆する。

本手法により、ワーカーの希望とより適合した求人を効率的に提示できるシステムの基盤が構築された。今後は、実データを用いたさらなる検証と、複数の手法を組み合わせたより高精度なマッチングアルゴリズムの開発、実サービスへの実装とユーザーフィードバックに基づく改善サイクルの確立が課題である。

謝辞

本研究および本論文の作成に関し、多大なる御指導、御討論を頂きました北海道大学大学院情報科学研究院メディアネットワーク部門、土橋宜典教授に心より感謝いたします。

本研究に対して、御助言、御協力を頂いた北海道大学大学院情報科学研究科メディアネットワーク専攻情報メディア学講座情報メディア環境学研究室諸氏に心より御礼を申し上げます。

参考文献

- [1] 厚生労働省, "2040 年を展望した社会保障・働き方改革本部資料" 厚生労働省 (2019)
- [2] J.Devlin, M.W.Chang, K.Lee and K.Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" *NAACL-HLT* (2019)
- [3] N.Reimers and I.Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks" *EMNLP* (2019)
- [4] 東北大学乾・鈴木研究室, "日本語 BERT モデル" *GitHub* (2020)
- [5] G.Salton and C.Buckley, "Term-weighting approaches in automatic text retrieval" *Information Processing & Management, Volume 24, Issue 5, pp.513-523* (1988)
- [6] A.Singhal, "Modern Information Retrieval: A Brief Overview" *IEEE Data Engineering Bulletin, Volume 24, pp.35-43* (2001)
- [7] M.Kobayashi, Y.Nakamura and T.Suzuki, "Job Matching System using Natural Language Processing" 情報処理学会論文誌, *Volume 62, No.3, pp.789-798* (2021)
- [8] T.Mikolov, K.Chen, G.Corrado and J.Dean, "Efficient Estimation of Word Representations in Vector Space" *ICLR* (2013)

研究業績

- [1] 星子 祐哉, 土橋 宜典 “介護マッチングサービスにおける自然言語処理を用いたマッチング精度向上手法の検討”, 電気・情報関係学会北海道支部連合大会 2024, (November. 2024).
- [2] 星子 祐哉, 土橋 宜典 “BERT を用いた介護職マッチングシステムの提案と評価”, 情報処理学会論文誌, (投稿準備中).