

パートナー企業との連携強化に向けた マッチング技術の研究

Matching Technology for Strengthening Cooperation with Partner Companies

ものづくり業界の企業にとって、パートナー企業との連携強化は経営上の重要な課題である。これを達成するために、高い技術力を持つパートナーを全社で活用していくことが望ましいが、部門ごとに顧客属性・仕事の進め方・要求スキル等が異なるため、他部門のパートナーと自部門のプロジェクトの適合度を判断することは難しい。本稿では、以上の課題を解決するために現在研究している、パートナーとプロジェクトの適合度を予測するマッチング技術を紹介する。提案手法に対して、マッチング実績の実データを用いた評価を行い、同一部門内のペアの適合度を高精度に予測できることを確認した。今後、実用化に向けて、異なる部門に属するペアの予測精度の改善を行っていく。

清藤 駿成	Seito Takanari
星 魁人	Hoshi Kaito
宗形 聡	Munakata Satoshi
鈴木 麻子	Suzuki Asako
鈴木 馨	Suzuki Kaoru

1. はじめに

ソフトウェア開発・建築・製造業などのものづくり業界では、人手不足の解消・コスト削減・自社で保持していない技術の調達などのために、社外の企業（パートナー企業）と協力して事業を進めることが多い¹⁾。さまざまな企業と協力することで自社の会社規模や技術範囲では実施できない事業に取り組むことができるため、パートナー企業との連携の強化は経営上の重要な課題となっている。

（株）日立ソリューションズ東日本（HSE）も多くのパートナー企業と連携して事業を行っている。HSE では、パートナー企業との連携を強化するために、高い技術力を持つパートナーを、特定のプロジェクトでの一時的な利用にとどまらず、別のプロジェクトで継続して利用することを推奨している。しかし、現状では、高い技術力を持つパートナーであっても、所属している部門で利用できる案件がない場合、契約が終了してしまうことが多い。パートナー企業との連携強化に取り組んでいる調達部では、このような場合であっても、別の部門でパートナーを継続して利用できるように、各部門に契約が終了してしまうパートナーの利用を促している。しかし、部門ごとにパートナーの評価の観点は異なっており、他部門が実施した評価では自部門のどの

ようなプロジェクトと適合しているのかを判断できないことから、部門を越えたパートナーの利用は進んでいない。

そこで、HSE では、部門を越えたパートナーの利用を促進するために、パートナーとプロジェクトの適合度を予測するマッチング技術の研究を行っている。部門間のパートナーとプロジェクトの適合度を正確に予測できるようになれば、利用者が他部門のパートナーを活用するための判断材料の一つになり、部門を越えたパートナーの利用を促進できると考えている。本稿では、研究中の技術の内容・評価結果・今後の展望について紹介する。

2. 提案手法

2.1 概要

本稿では、近年の主流となっている、深層ニューラルネットワーク（DNN）を用いたマッチング技術²⁾をベースとしたモデルを構築した。提案手法の全体像を図1に示す。提案手法の入力は、習得しているプログラミング言語やデータベースなどのパートナーの属性値と、プロジェクト概要や使用する技術などのプロジェクトの属性値である。属性値を直接比較して適合度を予測することは難しいため、特徴量抽出部で複

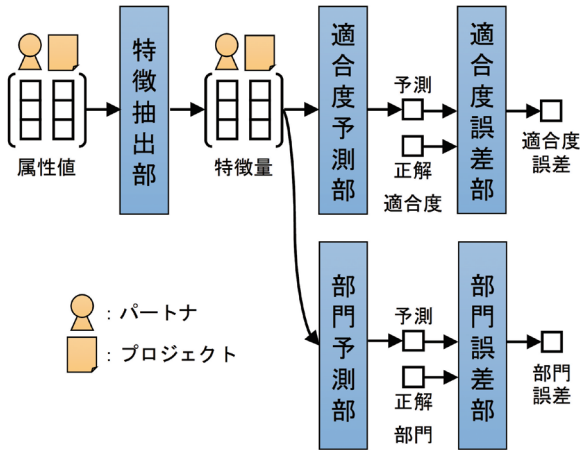


図 1 提案手法の全体像

【特徴抽出部】

$$y_i^0 = A_t^0 x_i + a_t^0 \quad (t \in \{ptn, prj\}, i \in S_t) \quad (1)$$

$$y_i^1 = A_t^1 \text{LeakyReLU}(y_i^{n-1}) + a_t^1 \quad (2)$$

【適合度予測部】

$$p_{ij} = 1 - \tanh(\text{MSE}(y_i^N, y_j^N)) \quad (i \in S_{ptn}, j \in S_{prj}) \quad (3)$$

【適合度誤差部】

$$e_{ij}^{evl} = \text{MSE}(p_{ij}, \hat{p}_{ij}) \quad (4)$$

【部門予測部】

$$z_i^m = y_i^N \quad (5)$$

$$z_i^m = B_t^m \text{LeakyReLU}(z_i^{m-1}) + b_t^m \quad (6)$$

$$z_i^{M+1} = B_t^{M+1} \text{LeakyReLU}(z_i^M) + b_t^{M+1} \quad (7)$$

$$q_i = \text{Softmax}(z_i^{M+1}) \quad (8)$$

【部門誤差部】

$$e_i^{dpt} = w \text{CrossEntropy}(q_i, \hat{q}_i) \quad (9)$$

数の属性値を組み合わせて適合度を予測しやすい情報(特徴量)に変換する。適合度予測部では、パートナとプロジェクトの特徴量を比較することで、それらのペアの適合度を計算する。適合度誤差部では、モデルが予測した適合度と実際の適合度を比較し、その誤差を計算する。この誤差を小さくするように特徴抽出部のパラメータを更新することで、適合度の計算方法が学習される。特徴抽出部・適合度予測部・適合度誤差部をまとめて適合度学習機能と呼ぶ。また、本稿で扱うマッチングでは、部門間のマッチング実績が少ない。これにより、部門内のペアの適合度を予測することに特化した特徴量が学習されてしまい、部門間のペアの適合度を正確に予測することが難しい。これを解決するために導入したのが部門予測部と部門誤差部である。特徴抽出器に部門を予測できないような特徴量を学習させることで、部門間に共通の特徴量を抽出する。部門予測部・部門誤差部をま

めてクロスドメイン機能と呼ぶ。各機能の詳細については 2.2 節と 2.3 節で説明する。

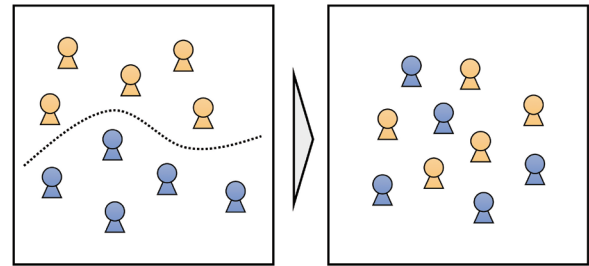


図 2 クロスドメイン機能有無による特徴量分布の違い

2.2 適合度学習機能

適合度学習機能では、マッチング実績のあるペアに対して、DNN が予測した適合度と実際の適合度の差分を小さくするようにパラメータを更新することで、適合度の計算方法を学習する。はじめに、特徴量抽出部(式(1)(2))で、DNN を用いて、マッチング実績のあるパートナ $i \in S_{ptn}$ とプロジェクト $j \in S_{prj}$ の属性値ベクトル $x_i \in R^{D_{ptn}}$, $x_j \in R^{D_{prj}}$ から、それぞれの特徴量ベクトル $y_i^N, y_j^N \in R^{D_{DNN}}$ を計算する。 $t \in \{ptn, prj\}$ はパートナ (ptn) またはプロジェクト (prj) を意味する添え字、 S_t はパートナまたはプロジェクトの集合、 D_t はパートナまたはプロジェクトの属性値数、 D_{DNN} は抽出する特徴量数、 $A_t \in R^{D_{DNN} \times D_t}$, $a_t^0 \in R^{D_{DNN}}$, $A_t \wedge n \in R^{D_{DNN} \times D_{DNN}}$, $a_t^n \in R^{D_{DNN}}$ は DNN のパラメータ、 $n \in \{1, \dots, N\}$ は DNN の層番号、 N は DNN の層数、LeakyReLU は漏洩正規化線形ユニットである。式(2)で全結合処理と活性化処理を繰り返すことで、適合度の予測に必要な特徴が抽出される。次に、適合度予測部(式(3))に、パートナ i とプロジェクト j の特徴量ベクトルの組 (y_i^N, y_j^N) を入力し、適合度 $p_{ij} \in R$ を計算する。MSE は二乗誤差関数、 \tanh は双曲線関数である。最後に、適合度誤差部(式(4))で、予測した適合度 p_{ij} と正解の適合度 \hat{p}_{ij} の誤差 $e_{ij}^{evl} \in R$ を計算し、誤差を小さくするようにパラメータを更新する。これにより、適合度の計算方法が DNN のパラメータとして学習される。

2.3 クロスドメイン機能

クロスドメイン機能は、部門間に共通の特徴量を学習するために導入した機能である^{3, 4)}。クロスドメイン機能の有無による特徴量分布の違いを図 2 に示す。特徴量はパートナの所属する部門ごとに異なる色で記載している。本稿で扱うマッチングでは、部門を越えたパートナの活用が進んでいないため、部門間のマッチング実績が少ない。このとき、適合度学習機能だけでは、図 2 左のように、部門内のペアだけを正確に予

測できるような部門ごとに固有の特徴（部門ごとに分布が異なる特徴）を学習してしまう。部門間のペアの適合度を予測するためには、部門に依存しない共通の特徴が必要であり、部門に固有の特徴では部門間のペアの適合度を正確に予測することは難しい。そこで、本稿では、各部門内のマッチング実績データから部門間に共通の特徴を学習するクロスドメイン機能を適用した。クロスドメイン機能では、図 2 右のように、各パートナー（プロジェクト）に対して、特徴量を用いてもどの部門に属するのか判別できないように抽出することで、部門間に共通の特徴を学習する。特徴量により部門を分類できないということは、その特徴量の近くに別の部門の特徴量もあることを意味している。したがって、抽出された特徴量には部門に共通した特徴が学習されていると言える。このように、部門間に共通の特徴を用いることで、部門間のペアの適合度を正確に予測できるようになる。

クロスドメイン機能の処理を説明する。はじめに、部門予測部（式 (5) ~ (8)）で、DNN を用いて、パートナー i とプロジェクト j の特徴量ベクトル y_i^N, y_j^N から、それぞれが属する部門の onehot ベクトル $q_i, q_j \in R^{D_{dpt}}$ を予測する。 D_{dpt} は部門の数、 $B_t^m \in R^{D_{DNN} \times D_{DNN}}, b_t^m \in R^{D_{DNN}}, B_t^{M+1} \in R^{D_{dpt} \times D_{DNN}}, b_t^{M+1} \in R^{D_{dpt}}$ は DNN のパラメータ、 $m \in \{1, \dots, M\}$ は DNN の層番号、 M は DNN の層数、 Softmax はソフトマックス関数である。次に、部門誤差部（式 (9)）で、予測した部門 q_i, q_j と正解の部門 \hat{q}_i, \hat{q}_j の誤差 e_i^{dpt}, e_j^{dpt} を計算し、部門予測部では誤差を小さくするように、特徴抽出部では誤差を大きくするようにパラメータを更新する。 CrossEntropy は交差エントロピー誤差関数である。なお、特徴量抽出部のパラメータは適合度予測機能とクロスドメイン機能の両方で更新されるため、クロスドメイン機能の影響度を調整するために、部門誤差 e_i^{dpt}, e_j^{dpt} には重み $w \in R$ を乗じた。

3. 評価

3.1 目的

提案手法を使って、マッチング実績データから部門間に共通の特徴を学習し、部門間のペアの適合度を高精度に予測できるかを評価した。

表 1 モデルのハイパーパラメータ

D_{ptn}	92	w	0.001
D_{pri}	147	最適化手法	Adam
D_{DNN}	32	バッチサイズ	フルバッチ
D_{dpt}	6	エポック数	4,000
N, M	8		

3.2 データ

評価には HSE で収集しているパートナーとプロジェクトのデータを用いた。パートナーの属性値には習得しているプログラミング言語やデータベースなどの IT スキルに関する項目を設定した。プロジェクトの属性値にはプロジェクトの内容（概要、使用する技術、など）やリスク（類似案件の経験有無、顧客との関係性、など）に関する項目を設定した。マッチング実績データにはパートナーが過去に従事したプロジェクトの 5 段階評価を 0-1 に変換した値を設定した。評価に用いたパートナーは 2,168 件、プロジェクトは 1,677 件、マッチング実績データは 2,732 件であった。また、マッチング実績データのうち、部門内のペアは 2,574 件 (94%)、部門間のペアは 158 件 (6%) であった。なお、評価で扱うデータを検討するにあたり、調達部門の実務面からの意見を取り入れている。

3.3 方法

マッチング実績のない部門間のペアに対して適合度を高精度に予測できるかを評価するために、部門内のペアだけを用いてモデルの学習を行い、学習したモデルで部門間のペアの適合度を予想して、正解の適合率と比較した。モデルのハイパーパラメータを表 1 に示す。

適合度の予測精度を評価するための評価指標には決定係数 R^2 を用いた。決定係数はモデルが予測した適合度と正解の適合度の差分から計算される値であり、差分が小さいほど値は 1 に近くなる。また、モデルが部門間に共通の特徴を学習していることを確認するために、部門正解率も評価した。部門正解率はモデルが予測した部門と正解の部門が一致している割合である。正解率が低い場合、モデルが部門間に共通の特徴を学習したと判断できる。

クロスドメイン機能の効果を評価するために、クロスドメイン機能をオフにしたケースと比較した。ここで、機能をオフにしたケースについても、部門に固有の特徴を学習していることを確認するために、部門予測部と部門誤差部を別途学習させ、部門の正解率を評価した。

クロスドメイン機能の有無の有意差を検証するために、学習データを分割して複数回評価を行い、得られた一連の結果に対して t 検定を行った。具体的には、学習データを 10 分割し、そのうち 1 区画を除いたデータで行う評価を、区画を変えながら合計 10 回行った。

3.4 結果と考察

決定係数と部門正解率の評価結果を表 2, 3 に示す。それぞれ 10 回の評価の平均と標準偏差（括弧内）が記載さ

れている。

表 2 決定係数の評価結果

	オン	オフ
部門内	0.9980 (0.0007)	0.9985 (0.0005)
部門間	0.3948 (0.0288)	0.3795 (0.0257)

表 3 部門正解率の評価結果

	オン	オフ
部門内	0.3635 (0.0434)	0.9743 (0.0037)
部門間	0.1955 (0.0045)	0.7347 (0.0146)

はじめに、提案手法の各機能が期待通りの学習をしたことを確認する。適合度学習機能については、表 2 のドメイン内ペアに対する決定係数に着目すると、クロスドメイン機能の有無に関わらず 1 に近い値であり、高い精度で予測できていることが分かる。

クロスドメイン機能は部門を予測できないように学習させることで、部門固有の特徴量ではなく、複数部門間に共通の特徴量を抽出する機能である。表 3 より、この機能を有効にすると部門正解率が低くなることから、部門固有の特徴量ではなく共通の特徴量を学習していると考えられる。

部門間のペアに対して適合度を高精度に予測できたかを確認する。表 2 の部門間ペアに対する決定係数は、クロスドメイン機能がオンの場合は 0.3948、オフの場合は 0.3795 であった。したがって、現状のモデルでは部門間のペアに対して適合度を高精度に予測することはできなかった。また、有意水準 5% の t 検定でクロスドメイン機能の有無の有意差は無く (p 値: 0.0872)、クロスドメイン機能による精度の改善効果は得られなかった。しかし、表 3 に示す通り、クロスドメイン機能の効果として、部門間に共通の特徴量は抽出できていると考えられるため、学習方法のさらなる改良により精度を改善できる可能性は高い。改善に向けた現状モデルの課題を特定するために、誤差の大きいペアの傾向について分析を進めていく。また、本稿で紹介した提案手法では、学習に用いていないデータの精度を改善する一般的な工夫 (ミニバッチ学習、正則化項、ドロップアウト、など) を適用していない。これらの工夫を適用することで精度が改善する可能性がある。

4. おわりに

本稿では、部門間でのパートナ活用を促進するための技術として、パートナとプロジェクトの適合度を予測するマッチング技術を紹介した。本稿のモデルの特徴として、部門間のペアの予測精度を改善するために、クロスドメイン機能を導入した。実データを用いた評価で、部門内のペアの適合度を高精度に予測できることを確認した。今後、クロスドメイン機能

の改良など、実用化に向けた精度の改善を行っていく。

参考文献

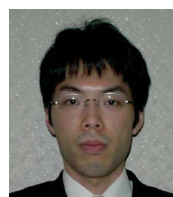
- 1) 日本アイ・ビー・エム株式会社: パートナ活用戦略の重要性, <https://www.ibm.com/downloads/cas/YNJLKBYP> (2021/11/12 閲覧)
- 2) Zhang, S., et al.: Deep Learning Based Recommender System: A Survey and New Perspectives, ACM Computing Surveys, Vol. 52(1), pp. 1-38 (2019)
- 3) Cantador, I., et al.: Cross-Domain Recommender Systems, Recommender Systems Handbook, Springer, pp. 919-959 (2015)
- 4) Kanagawa, H., et al.: Cross-domain Recommendation via Deep Domain Adaptation, European Conference on Information Retrieval, pp. 20-29 (2019)



清藤 駿成 2017 年入社
研究開発部
機械学習・AI を用いた意思決定支援技術、マッチング技術の研究開発



星 魁人 2020 年入社
研究開発部
機械学習・AI を用いた意思決定支援技術、マッチング技術の研究開発



宗形 聡 2003 年入社
研究開発部
機械学習・AI を用いた意思決定支援技術、データ分析技術の研究開発



鈴木 麻子 2009 年入社
調達部
パートナ企業等への委託契約に係る業務、開発人材のマッチング



鈴木 馨 2000 年入社
調達部
パートナ企業等への委託契約に係る業務、開発人材のマッチング