

## クロスドメインレコメンドを用いたパートナーとプロジェクトのマッチング

星 魁人† 清藤 駿成† 宗形 聡†

株式会社日立ソリューションズ東日本

### 1. はじめに

プロジェクト型事業を行う企業では、より複雑な需要への対応やイノベーションのために外部パートナーの戦略的な活用が重要である[1]。高いスキルを持つパートナーはプロジェクトの中核を担うことが多く、そのようなパートナーを様々なプロジェクトで活用することは経営上の重要な課題である。しかし、部門ごとに顧客の業種や作業の進め方、要求スキル等が異なるため、ある部門に所属するパートナーが別部門のプロジェクトに適合するかの判断は難しい。そこで本研究では、所属パートナーのプロジェクト参画実績からパートナーとプロジェクトの適性に関する特徴量を学習し、それを用いて両者の適合度を予測するマッチング手法を構築した。クロスドメインレコメンドを適用し、過去の参画実績に含まれない他部門のプロジェクトでも高精度に適合度を予測できるようにした。

### 2. 提案手法

本研究では、深層ニューラルネットワークを用いたマッチングモデルを構築した。提案手法の全体像を図1に示す。また、ネットワークの計算を式(1)~(9)に示す。

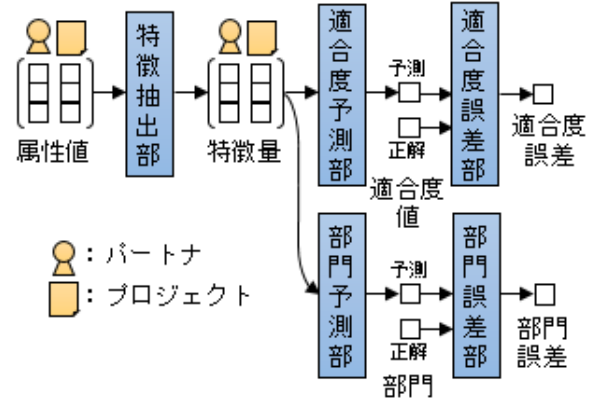
モデルの入力は、習得済みのプログラミング言語や経験年数等のパートナー属性値と、プロジェクト概要や使用技術等のプロジェクト属性値である。パートナーまたはプロジェクトの集合を $S_t$ とし、 $x_s$ を要素 $s(\in S_t)$ の属性値とする。ここで、 $t \in \{\text{ptn}, \text{prj}\}$ はパートナーまたはプロジェクトを示す添え字である。

特徴量抽出部 (式(1)(2)) により属性値 $x_s$ を適合度に関連する特徴量 $y_s^i$  ( $i \in \{0, \dots, n\}$ )に変換する。ここで $n$ は特徴量抽出部のネットワークの層数である。適合度計算部 (式(3)) では、パートナーとプロジェクトの特徴量の差異に基づき、当該ペアの適合度 $\hat{p}$ を予測する。適合度誤差部 (式(4)) では、モデルで予測した適合度 $\hat{p}$ と実際の適合度 $p$ の誤差 $e^{\text{evl}}$ を計算する。この誤差を小さくするように特徴量抽出部のパラメータ $A_t^i$ や $a_t^i$ を更新することで、適合度の計算方法が学習される。

パートナーの別部門プロジェクトへの参画は、それぞれ

#### Matching Partner Engineers to Projects Based on Cross Domain Recommendation

†Kaito Hoshi, †Takanari Seito, and †Satoshi Munakata at Hitachi Solutions East Japan. Ltd.



$$y_s^0 = A_t^0 x_s + a_t^0 \quad (s \in S_t, t \in \{\text{ptn}, \text{prj}\}) \quad (1)$$

$$y_s^i = A_t^i \text{LeakyReLU}(y_s^{i-1}) + a_t^i \quad (i \geq 1) \quad (2)$$

$$\hat{p} = 1 - \tanh(\text{MSE}(y_s^n, y_{s'}^n)) \quad (s \in S_{\text{ptn}}, s' \in S_{\text{prj}}) \quad (3)$$

$$e^{\text{evl}} = \text{MSE}(p, \hat{p}) \quad (4)$$

$$z_s^0 = y_s^n \quad (5)$$

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

$$z_s^m = B_t^m \text{LeakyReLU}(z_s^{m-1}) + b_t^m \quad (j = m, z_s^m \in \mathbb{R}^r) \quad (7)$$

$$\hat{q}_s = \text{Softmax}(z_s^m) \quad (8)$$

$$e_s^{\text{dpt}} = w * \text{CrossEntropy}(q, \hat{q}_s) \quad (9)$$

図1. 提案手法の全体像(上)および計算式(下)

の属性の違いにより適性判断が難しいため、所属部門と比べてその実績は少ない。したがって、学習用に得られるペアのデータは同一部門のものがほとんどとなる。

この場合、同一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、異なる部門間のペアの適合度を正確に予測することができない。この問題を解決するため、本研究ではクロスドメインレコメンド(CDR) [2]を適用し、部門予測部と部門誤差部を構築した。部門予測部 (式(5)~(8)) では、特徴量抽出部で計算したパートナーとプロジェクトの特徴量 $y_s^n$ から、それぞれが所属する部門 $\hat{q}_s$ を予測する。なお、 $z_s^j$  ( $j \in \{0, \dots, m\}$ )は部門予測に用いる隠れベクトル、 $B_t^j$ や $b_t^j$ はパラメータ、 $m$ はドメイン予測部のネットワークの層数、 $r$ は部門数で

ある。部門誤差部（式(9)）では、モデルが予測した部門  $\hat{q}_s$  と実際の部門  $q$  から交差エントロピー誤差  $e_s^{\text{dpt}}$  を計算する。  $w$  は適合度の予測誤差  $e^{\text{evl}}$  に対する部門の予測誤差  $e_s^{\text{dpt}}$  の重みである。

CDRの適用により、  $e^{\text{evl}}$  を小さくし、かつ  $e_s^{\text{dpt}}$  を大きくするようにパラメータ  $A_i^l$ ,  $a_i^l$  を更新することで、適合度を高精度で予測しながらも部門の予測はできない特徴量が学習される。部門の予測ができないことは、同一部門のペアデータから学習されてしまう部門固有の情報が特徴量から除外されることを意味している。異なる部門間のパートナーとプロジェクトのペアであっても、部門に共通した特徴量で適合度を予測できるようになり、結果として予測精度の向上が期待される。

### 3. 評価

#### 3.1. 方法

提案手法を使って、パートナーのプロジェクト参画実績から部門間で共通する適合度の特徴を学習し、異なる部門のペアの適合度を高精度に予測できるかを評価した。評価には自社のパートナーとプロジェクトのデータを用いた。パートナーは2,168人、プロジェクトは1,677件、パートナーがプロジェクトに参画したマッチング実績は2,732件である。マッチング実績のうち同一部門内のペアは2,573件（94%）、異なる部門間のペアは158件（6%）であった。なお、部門数は6である。

マッチング実績のない異なる部門間のペアに対する適合度を高精度に予測できるか評価するため、同一部門内のペアのみを用いてモデルの学習を行い、学習済みモデルで部門間のペアの適合度  $\hat{p}$  を予測した。適合度の評価には  $\hat{p}$  と  $p$  の決定係数  $R^2$  を用いた。CDRの適用あり/なしで  $R^2$  に有意差があるか検証するために、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学習することを、区画を変えながら合計10回行った。

また、CDRによってモデルが部門に共通の特徴を学習しているかを確認するため、CDRの適用あり/なしでそれぞれ部門予測の正解率も比較した。CDRを適用しない場合は、適合度予測部と適合度誤差部のみで学習した特徴量を新たな入力として、部門予測部と部門誤差部を別途独立して学習させた。こちらも学習データを10分割した交差検定を実施した。

#### 3.2. 結果

評価結果を表1, 2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。

学習に用いた同一部門内のペアに関しては、決定係数はCDRの適用有無にかかわらず0.9980以上であり、予測と正解の差分を小さくするように学習されていることが

表1. 決定係数の評価結果

	CDR適用あり	CDR適用なし
部門内	0.9980 (0.0007)	0.9985 (0.0005)
部門間	0.3948 (0.0288)	0.3795 (0.0257)

表2. 部門正解率の評価結果

	CDR適用あり	CDR適用なし
部門内	0.3635 (0.0434)	0.9743 (0.0037)
部門間	0.1955 (0.0045)	0.7347 (0.0146)

分かる。また、部門正解率は、CDRの適用がない場合には97.43%となり、部門固有の特徴が学習されていると考えられる。一方、CDRを適用した場合には36.35%と低く、部門をランダムに予測したときの精度（16.67%）に近い。CDRの適用により部門に共通の特徴を学習していると考えられる。

テストデータとして用いた異なる部門間のペアの決定係数は、CDR適用ありでは0.3948、適用なしでは0.3795であった。どちらの値も部門内のペアに対する結果と比較すると低い値であり、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。有意水準5%のt検定でもCDRの適用有無による有意差はなく（p値：0.0872）、CDRによる精度の改善効果は得られなかった。

### 4. まとめ

本稿では、部門を越えたパートナー活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を構築した。実データを用いた評価の結果、学習に用いた部門内のペアの適合度を高精度に予測できることは確認できたが、テストに用いた部門間のペアの予測精度には課題が残った。しかし、CDRの適用により、部門共通の特徴量は抽出でき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が高い。今後、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：パートナー活用戦略の重要性、<<https://www.ibm.com/downloads/cas/YNJLKBYP>> (参照 2022-01-07).
- [2] Kanagawa, H., et al.: Cross-domain Recommendation via Deep Domain Adaptation, European Conference on Information Retrieval, pp. 20-29 (2019).