

グラフニューラルネットワークを用いた コミュニケーション活性化人材発見

田中 宏昌^{†,††} 實成 優馬[†] 山田 渉[†] 落合 桂一[†] 若宮 翔子^{††}
荒牧 英治^{††}

[†] 株式会社 NTT ドコモ 〒100-6150 東京都千代田区永田町 2-11-1 山王パークタワー 25 階

^{††} 奈良先端科学技術大学院大学 〒630-0192 奈良県生駒市高山町 8916 番地の 5

E-mail: [†]{hiroaki.tanaka.sg,yuuma.jitsunari.rf}@nttdocomo.com, ^{††}{wataruyamada,keiichi.ochiai}@acm.org,
^{†††}{wakamiya,aramaki}@is.naist.jp

あらまし 企業において、チームのパフォーマンスを高めるための一つ的手段として人事異動がある。パフォーマンスをチームの効率性と定義した場合、パフォーマンスはチーム内のコミュニケーションが活発になるほど高くなることが知られている。一方で、誰がどのチームのコミュニケーションを活性化させるかを事前には分からない。そこで本論文では、Slack などのチャットツールのデータを用いて、チームに新しく参加する人がコミュニケーションを活性化させるかを算定するアプローチを提案する。本アプローチにおいて、コミュニケーションを活性化させるかを判定する機械学習モデルを正解率 0.57、適合率 0.58 で構築した。さらに、本アプローチでは、テキストコミュニケーションの内容を利用しないため、プライバシーや秘匿情報の観点から実際のビジネスシーンで活用しやすいという特性がある。
キーワード コミュニケーション活性化, チームパフォーマンス, チャットツール, ソーシャルグラフ, グラフニューラルネットワーク

1 はじめに¹

多くの企業において、利益向上のために組織やチームのパフォーマンスを改善することは重要だとされている。パフォーマンスと呼ばれる概念には、個人に対して定義される創造性や、集団に対して定義されるチームの効率性といった下位概念がある。効率性とは、定められたゴールをチームがいかにか早く達成できるかを示す概念であり、企業におけるチームではゴールが定められている場合が多いため、本稿では効率性という概念を用いてパフォーマンスを扱う。チームのコミュニケーションが活発であるほど効率性が高くなることが知られており [16]、管理職や人事部門の立場は、より活発にコミュニケーションを取るようなチーム構成を望むと考えられる。そのためには、チームのコミュニケーションを活性化させるような人材（以下、コミュニケーションエンハンサーと呼ぶ）を人事異動によってチームに加えるという方法が考えられるが、会社内の誰がコミュニケーションエンハンサーとなるかを事前に知ることは困難だという問題がある。

チームコミュニケーションの活性度は、コミュニケーションの密度や凝集性によって定義されており、多くの既存研究では質問紙調査での計測 [9] やデバイスを用いた計測 [17] が行われていた。しかし、質問紙に定期的に回答することは従業員

にとっては時間的な負担となるため、継続的な活性度の計測が難しいという課題がある。また、Covid-19 のパンデミックによって在宅勤務者が増加したため、デバイスを用いてオフィスでのコミュニケーションをセンシングするアプローチでは十分なデータ量を確保することが難しくなった。

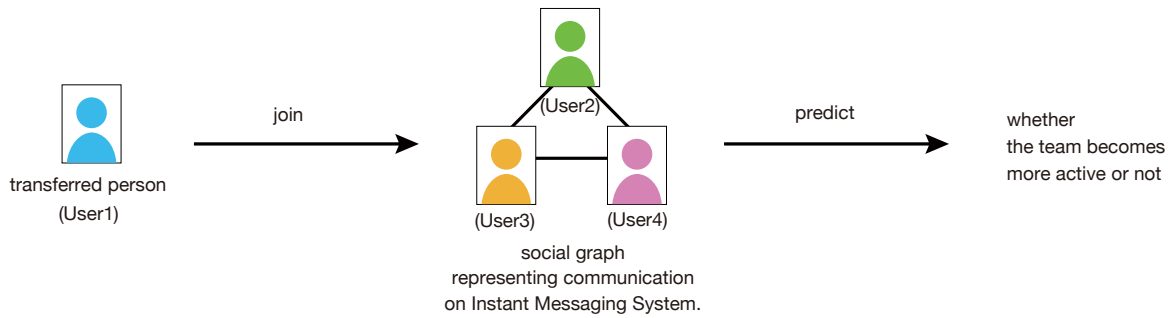
コミュニケーション活性度の計測方法に関する課題に加え、具体的にどうすればコミュニケーションが活性化するのかについては未知だという課題もある。チームのパフォーマンスや効率性とコミュニケーションの活性度の関係性に関する研究は複数存在し [12, 16]、これらの研究は性格特性や人口動態的な多様性がチームのパフォーマンスやコミュニケーション活性度に影響を及ぼす要因の一つであるという重要な知見を与えてくれる。その一方で、コミュニケーション活性化のための支配的な要因や実際に取るべき行動については言及されていない。加えて、それらの研究で利用されている、性格特性を表す Myers-Briggs Type Indicator (MBTI) [14] や Keirsey 気質 [8] は質問紙調査によって計測されるため、多くの要因を継続的に計測するのは面倒だという課題もある。

在宅勤務の増加に伴ってオフィスでのコミュニケーションが減少する一方で、Slack や Teams などの Instant Messaging System (以下、IMS と呼ぶ) 上でのコミュニケーション量が爆発的に増加した。そこで本研究では、IMS のデータを利用してコミュニケーションエンハンサーを発見するアプローチを提案する。本アプローチは

- 質問紙調査や特殊なデバイスを使わずにコミュニケーションエンハンサーを発見できるため、社員への負担が小さい
- IMS 上の個別の投稿文書を必要としないため、プライバ

¹：本研究は、The 6th IEEE Workshop on Human-in-the-Loop Methods and Future of Work in BigData (IEEE HMDData 2022) において「Graph Neural Network Tells Us Who is the Communication Enhancer」として発表したものである。

Problem Configuration



Proposed Approach

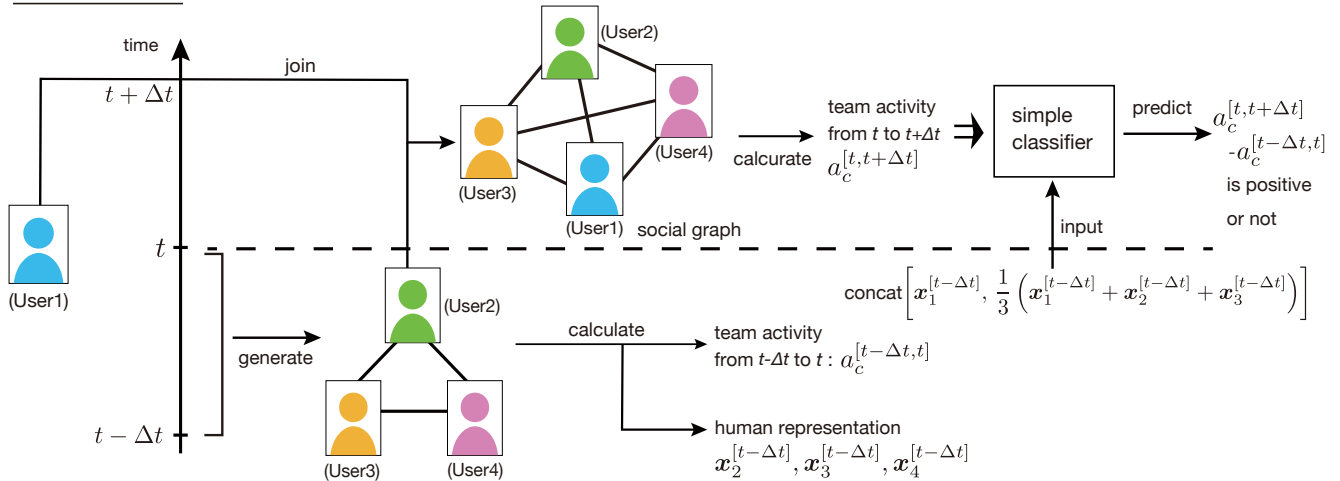


図 1 問題設定と提案アプローチの概要：時刻 t において、ユーザ 2・ユーザ 3・ユーザ 4 で構成されるチームにユーザ 1 を異動させると、時刻 $t + \Delta t$ までの期間でユーザ 1 がコミュニケーションエンハンサーになるのかを予測するという状況を考える。時刻 $t - \Delta t$ までのコミュニケーションログからコミュニケーションネットワークを構築し、個人およびチームのベクトル表現を GNN により獲得する。その後、LightGBM 等の軽量な分類器によって、時刻 $t + \Delta t$ までの間にユーザ 1 がチームのコミュニケーションを活性化させたかを判定する。

シーや秘密保持の観点での懸念も小さい

という利点を持つ。本アプローチの概要を図 1 に示す。提案アプローチでは、異動者がコミュニケーションエンハンサーか否か、すなわち新しくチームに入ってくる候補の人がチームのコミュニケーションを活性化させるか否かを、事前のコミュニケーションログに基づいて判定する。具体的には、社員のコミュニケーションログからコミュニケーションのネットワークを構築して、Graph Neural Network (以下、GNN と呼ぶ) を用いて社員のベクトル表現 (以下、human representation と呼ぶ) を獲得し、その human representation を入力とする機械学習モデルを構築する。そして、この機械学習モデルによって、異動者がコミュニケーションエンハンサーかどうかを判定する。この一連のアプローチによって、チームのコミュニケーション活性化における支配的要因が明らかにならなくてもコミュニケーションエンハンサーを発見することができる。

さらに、本アプローチを評価するために、実際に株式会社 NTT ドコモで利用されている Slack のデータを収集した。デー

タは 2020 年 7 月 1 日から 2022 年 7 月 30 日までの期間で収集されており、データセットの詳細は第 3.1 節に記す。

2 関連研究

2.1 コミュニケーション密度とパフォーマンス

Krackhardt [9] を始めとして、人のコミュニケーション密度が人や集団のパフォーマンスに及ぼす影響に関する研究は多く存在するが、コミュニケーション密度の定義は多岐にわたっている。たとえば、Lucius らの研究 [11] では、コミュニケーション密度をともに働く人の相互の選好と定義している。しかし、近年の実証研究 [3, 10, 20] によって、チームの効率性という意味でのパフォーマンスは、チームの密度または凝集性によって計測されることが明らかになっている (以降、本稿ではパフォーマンスを効率性という意味で用いる)。密度と凝集性は、チームまたは組織のコミュニケーションを表現するコミュニケーショングラフに対して定義される量で、密度や凝集性が増加すると、パフォーマンスも向上することが知られている。上記の関連研

究 [3,9-11,20] は、コミュニケーション密度とパフォーマンスの間の重要な関係性を示しているが、実際のパフォーマンス向上のために、どのような行動をとればコミュニケーション密度が向上するのかという問題までは扱っていない。

2.2 チームパフォーマンスへの影響因子

Deonne らの研究 [4] では、変革型リーダーシップと呼ばれるリーダーシップスタイルが、チームパフォーマンス向上に影響すると報告している。一方で、会社組織においてはリーダーシップを発揮する立場にない人間が大多数であり、その多数派のフォロワーに関する示唆は得られない。

また、Mazni らの研究 [12] では、MBTI やチームメンバーの人口動態の多様性がチームパフォーマンスに影響することを示している。これはすなわち、個人やチームの特性がパフォーマンスに影響を与えるという重要な示唆を提供しており、これらの特性を考慮した機械学習モデルを構築することで、コミュニケーション活性度を予測できる可能性がある。しかし、Mazni らは、MBTI 等の個人特性の計測に質問紙調査を実施しており、同様の手法を継続的に会社組織において実施することは、社員への負担が大きいという課題が残る。

2.3 チームパフォーマンスの予測

Bakker らの研究 [2] では、ジョブ・ディスクリプションに含まれていない役割、たとえば自律性や社会的支援などの職務資源が、パフォーマンスを予測する重要な因子であることが示されている。Bakker らの研究は、職務上で規定された働きだけではなく、ジョブ・ディスクリプションに言語化されていない個人の特性や周囲の状況がチームパフォーマンスの予測に重要であるという重要な示唆を与えている。この研究結果に基づき、本研究で提案するアプローチでは、言語化された情報ではなく、個人が周囲の人間とどのようにコミュニケーションを取っているかという情報を活用して、コミュニケーションエンハンサーを発見するアプローチを採用している。

Wang らの研究 [19] では、チームパフォーマンスを表す指標として論文の投稿数を用いており、チームの論文投稿数を予測する手法を提案している。Wang らは、機械学習モデルを用いて論文投稿数予測を行っており、モデルへの入力としてコミュニケーションスタイルと呼ばれる、Slack チャンネル上での様々な統計量を用いている。これらの統計量は個人に対して定義されるものではなく、チームでのコミュニケーション全体に対して定義される量であるため、特定の個人のパフォーマンスや論文投稿数を予測できない。しかし、Wang らの研究は、チームのコミュニケーションスタイルが論文投稿数といったパフォーマンスに影響するという重要な結論を出している。そのため、本研究においても個人だけでなく集団のコミュニケーション情報を用いるアプローチを提案する。

3 データセットと提案アプローチ

3.1 Slack データの概要

提案アプローチでは、Slack のデータに基づいて、異動者が

表 1 データセット概要：本データセットは株式会社 NTT ドコモの R&D 組織で実際に利用されている Slack のデータから作成した。1 つの WS に部署に所属する社員全員が含まれており、仕事上のコミュニケーションの多くは Slack 上で行われている。

収集期間	2020-07-01 から 2022-07-30
WS の数	1
チャンネルの数	313
ユーザ数	659
チャンネルあたりユーザ数の平均	9.86
チャンネルあたりユーザ数の標準偏差	8.21
ユーザのチャンネル参加イベント数	3253
チャンネルあたりイベント発生回数の平均	12.09
チャンネルあたりイベント発生回数の標準偏差	8.99
ユーザあたりイベント発生回数の平均	3.91
ユーザあたりイベント発生回数の標準偏差	9.84
合計メンション数	488206
チャンネルあたりメンション数の平均	761.63
チャンネルあたりメンション数の標準偏差	1546.33
ユーザあたりメンション数の平均	474.45
ユーザあたりメンション数の標準偏差	1472.89

コミュニケーションエンハンサーかを判別する機械学習モデルを構築する。Slack には、集団に関して 2 つの層がある。第 1 層はワークスペース（以下、WS と呼ぶ）と呼ばれ、第 2 層はチャンネルと呼ばれる。ワークスペースにはメンバー全員が入っており、チャンネルはプロジェクト単位でメンバーが分けられている。チャンネルには、WS のメンバー全員から見る事ができるパブリックチャンネルと、チャンネルに入っている人からしか見ることの出来ないプライベートチャンネルがある。

本研究では、株式会社 NTT ドコモの研究開発部門で実際に利用されている Slack のパブリックチャンネルのデータを収集した。収集期間は 2020 年 7 月 1 日から 2022 年 7 月 30 日までで、2022 年 7 月 30 日時点で WS のメンバーは 659 人、パブリックチャンネルは 313 個存在した。ただし、このパブリックチャンネルは、簡単な事務連絡チャンネルは除外されている。その他のデータセットの統計情報は表 1 に記されている。なお、Slack のデータを収集する前に、本研究の目的や分析内容について全ての参加者からインフォームド・コンセントを得ている。また、本研究では参加者による物理的または精神的な介入を必要とせず、参加者情報は分析前に匿名化されている。

3.2 前処理

Slack のテキストデータに対する前処理を記す。Slack におけるメッセージテキストには、@[ユーザ名] のような相手を指名するマークがあり、このマークの言及相手を会話相手と定める。ただし、マークを付けないユーザも存在しており、マークがない場合は、そのチャンネルのすべてのユーザと会話しているとみなす。また、[ユーザ名] さん や [ユーザ名] くん といった日本語特有の表現についてもマークとみなす。さらに、会議のリンクや別のメッセージへのリンク、さらに絵文字等のコマンドも削除する。

3.3 Slack データの定式化

異動者 $i, i = 1, \dots, N$ が Slack チャンネル $c, c = 1, \dots, C$ において時刻 $t + \Delta t$ においてコミュニケーションを活性化させているかを知りたいとする。ここで、 t は現在の時刻を表す。つまり、本研究のデータではプロジェクト単位でチャンネルが作られているため、そのプロジェクトに新規に人が入ってくる状況を想定し、新規に入ってくる人がコミュニケーションエンハンサーかを判定したいという状況である。

IMS 上のコミュニケーションは、コミュニケーショングラフとして表現され、ノードがユーザに対応し、エッジは会話が発生した人同士を結び、エッジの重み w_{ij} はユーザ i からユーザ j への発言回数とする。より形式的には、チャンネル c のコミュニケーショングラフを $\mathcal{G}_c^{[t-\Delta t, t]} = (\mathcal{V}_c, \mathcal{E}_c)$ で表し、ユーザ i のノードは $v_i \in \mathcal{V}_c$ で、ユーザ i とユーザ $j, j = 1, \dots, N, j \neq i$ を結ぶエッジは $e_{ij} \in \mathcal{E}_c$ で、エッジ e_{ij} の重みは w_{ij} でそれぞれ表されるとする。そして、コミュニケーショングラフ $\mathcal{G}_c^{[t-\Delta t, t]} = (\mathcal{V}_c, \mathcal{E}_c)$ に対して、チームの活性度 $a_c^{[t-\Delta t, t]} \in \mathbb{R}$ を

$$\frac{\sum_{i: v_i \in \mathcal{V}_c} \sum_{j: v_j \in \mathcal{V}_c} \mathbb{1}[w_{ij} \geq 1]}{N_c(N_c - 1)} \quad (1)$$

$$\frac{\sum_{i: v_i \in \mathcal{V}_c} \sum_{j: v_j \notin \mathcal{V}_c} \mathbb{1}[w_{ij} \geq 1]}{N(N - N_c)}$$

によって定義する。ここで、 N_c はチャンネル c におけるユーザ数を、 N は WS 全体でのユーザ数をそれぞれ表す。スコア $a_c^{[t-\Delta t, t]}$ は凝集性と呼ばれており、ネットワーク密度を測るために用いられることが多い。

3.4 アプローチ

提案アプローチは、human representation 獲得・team representation 計測・モデル構築の3つのフェーズに分けられる。具体的には、human representation 獲得フェーズでは、異動者の特徴ベクトルを構築し、team representation 計測フェーズでは、human representation からチームのベクトル表現を構築し、モデル構築フェーズでは $a_c^{[t-\Delta t, t]} - a_c^{[t, t+\Delta t]}$ が0以上の値を取るかを分類する機械学習モデルを構築する。

Human representation 獲得フェーズ： Human representation 獲得フェーズでは、Slack 上のコミュニケーションによって生成されるコミュニケーショングラフから、ユーザの埋め込み表現を獲得する。コミュニケーショングラフには、パーソナリティやリーダーシップなどの様々な個人特性に加え、チームのコミュニケーションスタイルといった、パフォーマンスに影響を与えると思われる様々な特性を内包していると考えられる。したがって、個人レベルの特性とチームレベルの特性の双方を包含するコミュニケーショングラフから個々のベクトル表現を獲得することで、異動者がコミュニケーションエンハンサーであるかを予測可能であると考えられる。コミュニケーショングラフ上の個人をベクトル化するための最も基本的な方法は、ノードに関連した特徴量を設計する手法であるが、本研究では、様々なグラフ構造データに対して高い性能を示す [6, 18] GNN を用いたノード埋め込みを採用する。この手法により、個人レベルの特性やチームレベルのコミュニケーションスタイルの情

報を含んだ human representation を構築することができる。

具体的な GNN のアルゴリズムとして、テキスト内容を使用しない最も基本的なアルゴリズムである DeepWalk [15] を採用する。まず、ユーザ i を含むチャンネルのコミュニケーションのみから、コミュニケーショングラフ $\mathcal{G}_i^{[t-\Delta t, t]}$ を以下のように構築する：

$$\mathcal{G}_i^{[t-\Delta t, t]} = \left(\bigcup_{c: i \text{ is included}} \mathcal{V}_c, \bigcup_{c: i \text{ is included}} \mathcal{E}_c \right). \quad (2)$$

次に、 $\mathcal{G}_i^{[t-\Delta t, t]}$ に対し、DeepWalk を用いてユーザ i の human representation $\mathbf{x}_i^{[t-\Delta t, t]} \in \mathbb{R}^d$ を構築する。DeepWalk では、グラフ $\mathcal{G}_i^{[t-\Delta t, t]}$ 上のランダムウォークを考えてノードの系列を生成し、その系列に対してスキップグラムモデル [13] を用いてノードの埋め込み表現を獲得する。ただし、本研究ではランダムウォークによる系列生成において、ノード v_i からノード v_j の遷移確率を

$$p_{ij} = \frac{w_{ij}}{\sum_{j: v_j \in \mathcal{V}_c, j \neq i} w_{ij}}$$

によって定める。

Team representation 計測フェーズ： Team representation 計測のフェーズでは、個人の human representation からチームのベクトル表現を構築する。まず、時刻 $t - \Delta t$ から t の期間でチャンネル c に所属している人の human representation の集合

$$\mathcal{X}_c^{[t-\Delta t, t]} = \left\{ \mathbf{x}_j^{[t-\Delta t, t]} \mid \text{user } j \text{ is in channel } c \right\}$$

を作成する。そして、チームのベクトル表現を、チームに属する個人の平均とみなし

$$\bar{\mathbf{x}}_c = \frac{1}{|\mathcal{X}_c^{[t-\Delta t, t]}|} \sum_{\mathbf{x} \in \mathcal{X}_c^{[t-\Delta t, t]}} \mathbf{x}$$

によって定める。

モデル構築フェーズ： 最後のモデル構築フェーズでは、 $a_c^{[t, t+\Delta t]} - a_c^{[t-\Delta t, t]}$ の正負を予測する機械学習モデル（分類器）を構築する。このモデルは、ユーザ i の human representation $\mathbf{x}_i^{[t-\Delta t, t]}$ と、ユーザ i が時刻 t 以降に新規に所属するチーム（チャンネル） c の team representation $\bar{\mathbf{x}}$ を結合させたベクトル

$$\begin{bmatrix} \mathbf{x}_i^{[t-\Delta t, t]} & \bar{\mathbf{x}} \end{bmatrix}$$

を入力として、 $a_c^{[t, t+\Delta t]} - a_c^{[t-\Delta t, t]}$ が正のときは1を、非正のときは0を出力する機械学習モデルである。以上の3つのフェーズから、異動者 i がコミュニケーションエンハンサーかを予測する。

4 評価

4.1 評価実験設定

提案アプローチの評価として、構築した機械学習モデルの精度と正確度を確認する。ユーザがチャンネルに参加するイベン

表 2 実験結果：増加ラベルは、 $a_c^{[t,t+\Delta t]} - a_c^{[t-\Delta t,t]} > 0$ 、つまり異動者がコミュニケーションエンハンサーであることを示し、逆に非増加は、 $a_c^{[t,t+\Delta t]} - a_c^{[t-\Delta t,t]} \leq 0$ 、つまり異動者がコミュニケーションエンハンサーでないことを表す。

		予測ラベル	
		増加	非増加
正解ラベル	増加	331	110
	非増加	240	133

トが 3253 ケース存在するため、これら全てのケースにおける $a_c^{[t,t+\Delta t]} - a_c^{[t-\Delta t,t]}$ の正負を測定し、目的変数とする。なお、本稿では Δt を 90 日に設定した。

3253 ケースのイベントのうち、1626 ケースを学習データとし、813 ケースをハイパーパラメータチューニング用の開発データ、814 ケースをテストデータとする。また、分類のための機械学習モデルとして LightGBM [7] を用い、ハイパーパラメータのチューニングには Optuna [1] を用いた。

4.2 結果と議論

実験結果を表 2 に示す。増加ラベルは、 $a_c^{[t,t+\Delta t]} - a_c^{[t-\Delta t,t]} > 0$ 、つまり異動者がコミュニケーションエンハンサーであることを示し、逆に非増加ラベルは、 $a_c^{[t,t+\Delta t]} - a_c^{[t-\Delta t,t]} \leq 0$ 、つまり異動者がコミュニケーションエンハンサーでないことを表す。表 2 から、正解率 0.75、適合率 0.58 でコミュニケーションエンハンサーを発見できたことがわかる。

ただし、本評価結果には注意すべき点がある。まず、評価に用いられたデータは株式会社 NTT ドコモの一研究開発部門のデータのみであるため、安易に一般化できない可能性がある。次の注意点として、予測の安定性への懸念が挙げられる。機械学習モデルの学習及び評価には比較的短期間の Slack データを用いたため、どの程度の期間があれば安定的にコミュニケーションエンハンサーを発見できるのかが不明瞭な点である。

また、提案アプローチの限界として、転職者や新入社員に対しては本アプローチを利用できない点が挙げられる。本アプローチを利用する上で、個人の human representation を構築するには一定量のコミュニケーションログが必要となるが、転職者や新入社員のコミュニケーションログは一般的には活用できないため、本アプローチを転職者や新入社員に適用することができない。

5 おわりに

本稿では、効率性という意味でのチームパフォーマンスを向上させる人材をコミュニケーションエンハンサーと呼び、人事異動等による新規チーム参加者（異動者）がコミュニケーションエンハンサーなのかを予測するアプローチを提案した。先行研究では、MBTI や Keirsey 気質などの性格特性がチームパフォーマンスに影響を与えることが示されているが、支配的な要因は見つかっておらず、良いと考えられている様々な特性を

質問紙調査によって計測することは非常に面倒である。そのため、「チームパフォーマンスを向上させるために、誰をチームに入れるべきか」という問いに現実的に答えることは困難であった。提案アプローチでは、IMS 上でのコミュニケーションネットワークを利用して、この問題を解決することができる。さらに、提案アプローチは IMS 上での会話内容を必要としないため、プライバシーや機密性の懸念が低いという利点がある。具体的には、human representation 獲得・team representation 計測・モデル構築の 3 つのフェーズからなるアプローチを提案した。最初のフェーズでは DeepWalk による埋め込みによって、異動者のベクトル表現である human representation を獲得した。次の team representation 計測フェーズでは、異動者が新規に所属するチームのベクトル表現を構築した。そして最後のモデル構築フェーズにおいて、異動者がチームのコミュニケーションを活性化させるか、すなわちコミュニケーションエンハンサーであるかを予測する機械学習モデルを構築した。提案アプローチを評価するために、株式会社 NTT ドコモで実際に利用されている Slack の大規模なデータセットを収集した。このデータセットを用いて提案アプローチを評価した結果、構築した機械学習モデルは正確度 0.57、精度 0.58 を示した。

本研究で残された課題であるコミュニケーションログ以外の情報を活用できないことへの対策として、human representation 構築のための GNN のアーキテクチャを変更して、専門知識やスキルセットを考慮した human representation を獲得する方法が考えられる。また、本研究では、パフォーマンスを効率性という意味で捉え、チーム内のコミュニケーション活性化を目的としたが、パフォーマンスを創造性という意味で考えると、チーム内コミュニケーション密度に対してチーム外コミュニケーション密度を向上させたほうが良いと言われている [5]。したがって、より実践的なアプローチとして、効率性と創造性のトレードオフを考慮して、複数の意味でのチームパフォーマンスに良い影響を与える人物を発見するアプローチの開発が必要である。

謝 辞

本研究は、JST、未来社会創造事業、JPMJMI21J2 の支援を受けたものである。

文 献

- [1] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A Next-generation Hyperparameter Optimization Framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '19*, pages 2623–2631, New York, NY, USA, 2019. Association for Computing Machinery.
- [2] Arnold B. Bakker, Evangelia Demerouti, and Willem Verbeke. Using the job demands-resources model to predict burnout and performance. *Human Resource Management*, 43(1):83–104, 2004.
- [3] Daniel J. Beal, Robin R. Cohen, Michael J. Burke, and Christy L. McLendon. Cohesion and performance in groups:

- a meta-analytic clarification of construct relations. *The Journal of Applied Psychology*, 88(6):989–1004, December 2003.
- [4] Shelley D. Dionne, Francis J. Yammarino, Leanne E. Atwater, and William D. Spangler. Transformational leadership and team performance. *Journal of Organizational Change Management*, 17(2):177–193, January 2004. Publisher: Emerald Group Publishing Limited.
- [5] Mark S Granovetter. The strength of weak ties. *American journal of sociology*, 78(6):1360–1380, 1973.
- [6] William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, pages 1025–1035, Red Hook, NY, USA, 2017. Curran Associates Inc.
- [7] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: a highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, pages 3149–3157, Red Hook, NY, USA, 2017. Curran Associates Inc.
- [8] David Keirse and Marilyn M Bates. *Please understand me: Character & temperament types*. Prometheus Nemesis Book Company Del Mar, CA, 1984.
- [9] David Krackhardt. The Strength of Strong Ties : The Importance of Philos in Organizations. *Journal of Networks and Organizations: Structure, Form, and Action*, pages 216–239, 1992.
- [10] Paul Leonardi and Noshir Contractor. Better People Analytics. *Harvard Business Review*, 2018.
- [11] Randall H. Lucius and Karl W. Kuhnert. Using Sociometry to Predict Team Performance in the Work Place. *The Journal of Psychology*, 131(1):21–32, January 1997.
- [12] Omar Mazni, Sharifah-Lailee Syed-Abdullah, and Naimah Mohd Hussin. Analyzing personality types to predict team performance. In *2010 International Conference on Science and Social Research (CSSR 2010)*, pages 624–628, December 2010.
- [13] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. In *Proceedings of International Conference on Learning Representations (ICLR 2013)*, January 2013.
- [14] Isabel Briggs Myers and Peter B. Meyers. *Gifts differing: Understanding personality type*. Mountain View, CA: Davies-Black Publishing, 1995.
- [15] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. DeepWalk: online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD ’14*, pages 701–710, New York, NY, USA, 2014. Association for Computing Machinery.
- [16] Ray Reagans, Ezra Zuckerman, and Bill McEvily. How to Make the Team: Social Networks vs. Demography as Criteria for Designing Effective Teams. *Administrative Science Quarterly*, 49(1):101–133, March 2004. Publisher: SAGE Publications Inc.
- [17] Priyamvada Tripathi and Winslow Burleson. Predicting creativity in the wild: experience sample and sociometric modeling of teams. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, CSCW ’12*, pages 1203–1212, New York, NY, USA, 2012. Association for Computing Machinery.
- [18] Petar Veličković, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, and R. Devon Hjelm. Deep Graph Infomax. In *Proceedings of International Conference on Learning Representations (ICLR2018)*, September 2018.
- [19] Dakuo Wang, Haoyu Wang, Mo Yu, Zahra Ashktorab, and Ming Tan. Group Chat Ecology in Enterprise Instant Messaging: How Employees Collaborate Through Multi-User Chat Channels on Slack. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1):1–14, March 2022.
- [20] Sean Wise. Can a team have too much cohesion? The dark side to network density. *European Management Journal*, 32(5):703–711, October 2014.