

社内データに基づくイノベータ人財のピックアップ

成田 和弥

田内 真惟人

三澤 賢祐

中島 正成

エン・ジャパン株式会社

{kazuya_narita, maito_tsuchi, kensuke_mitsuzawa, masanori_nakashima}@en-japan.com

1 はじめに

エン・ジャパンでは、不満買取センター¹やエン婚活²をはじめとした、新規事業の創出に積極的に取り組んでいる。しかしながら、当社は、主に人材採用・入社後活躍サービスを提供している企業であるため、新規事業に関心が高く、新規事業を任せられる人材(以降、イノベータ人財と呼ぶ)を優先して採用しているわけではない。そのため、イノベータ人財がどれだけ社内存在し、誰がイノベータ人財であるかが明らかでない。社内に潜在するイノベータ人財を見つけ出し、適切に配置することが、新規事業の成功可能性だけでなく、本人のモチベーションの面からも重要である。

社内からイノベータ人財を発見するにあたり、現状では、本人や周囲からのヒアリング等を通して、人手でイノベータ人財のピックアップを行っている。しかしながら、1,000人ほどの社員の中から、人手で網羅的にピックアップを行うことは容易ではない。

そこで本論文では、イノベータ人財を抜け漏れなく効率的に発見するため、社内データを活用したイノベータ人財のピックアップを行った。まず、イノベータ人財をどのような指標でピックアップすべきか、という部分から議論し、9種類の指標を定義した。次に、社内データを元にこれらの指標をスコア化した。スコア化には、スコアが高くなった要因を把握できるように、シンプルで透明性の高い手法を用いた。そして、高いスコアを持つ社員をイノベータ人財としてピックアップした。ピックアップされた社員が抽出すべきイノベータ人財像と合致しているかを人手で評価することで、適切なピックアップが行えていることを示した。

2 関連研究

本論文における社内データとは、企業内に蓄積される、営業、経理、人事等に関する機密性の高いデータを指す。社内データからの人物抽出に関する研究は見られないものの、社内データを用いた何らかの分析、あるいは、社内データ以外のデータを対象とした人物

抽出、といった先行研究は存在している。

社内データを用いた分析に関しては、社内データから種々の傾向を分析することで、業務の効率化等に適用するためのサービスが既に存在している。しかしながら、テキスト情報のような定性的なデータではなく、勤怠記録のような定量的なデータのみを利用していることが多い。社内に蓄積された定性データを用いた研究としては、谷口ら [5] がある。谷口らは、営業日報から顧客の課題を抽出するとともに、課題抽出技術を導入した商談事例検索システムを構築している。

社内データを用いた社員の抽出、分類に関する研究はないものの、それ以外のデータ、例えばソーシャルデータによってそのユーザを分類する研究は存在する。ソーシャルメディアのユーザ属性推定はその最たるものであり、例えば池田ら [7] は性別、年齢、居住地の推定手法を、榊ら [6] は職業の推定手法を、それぞれ提案している。また、Pennacchiottiら [3] は、Twitterのユーザを Political affiliation、Ethnicity、Starbucks fans の3つの観点でそれぞれ分類している。上記の手法はいずれも機械学習に基づく分類手法であるが、本タスクでは十分量の正解データを得ることが困難であるため、これらの手法を応用することは難しい。

3 ピックアップ方針

イノベータ人財のピックアップを行う上で、どのような基準でピックアップを行うかが最も重要となる。そこで、現在新規事業に携わっている社員の意見をもとにして、以下の9種類の指標を定義した。

1. 事業性：新規事業に関心を持っており、ビジネス、収支、利益にシビアである
2. デジタル知見：デジタル領域に対して関心を持ち、その情報を広く把握している(デジタル領域での新規事業も想定しているため)
3. 学習能力：日常の困難の克服の中から、優れた解決策を自ら学ぶことができる
4. 主体性：自ら積極的に責任を持って動くことができる
5. 変革性：現状に満足せず、変化・変革を好む
6. 目標必達志向性：常にチャレンジングな目標を掲げ、達成に強い執着心をもっている
7. 意思伝達力：自分の考えをしっかりと伝えられる
8. 行動性：考えるよりも行動を先にする傾向がある

¹<http://fumankaitori.com/>

²<https://en-konkatsu.com/>

9. 独創性：独創的な発想を持ち、アイデアを多く出すことができる

本論文では、社内データをもとに、上記の指標それぞれに対応するスコアを作成する。そして、そのスコアをもとにランキングを行うことで、イノベータ人財のピックアップを行う。

4 利用するデータ

3節で議論した指標に基づき、社内データを利用してイノベータ人財のピックアップを行う。本論文では、2016年9月時点で所属している正社員935名を対象とし、以下の6種類のデータを利用する。

- 社員各々が記入しているプロフィール
当社では、社員がそれぞれにプロフィールを記述し、社内で公開している。志望動機や仕事内容、趣味など、ある程度のフォーマットは決まっているものの、その記述量は社員によって大きく異なる。
- 社員各々が日々の業務内容を記録した日報データ
2015年9月～2016年9月までに記録された日報、全161,952件。その日の業務内容だけでなく、競合他社の情報等を記録する場合もある。部署や役職によって日報の記述内容や量が異なっており、1社員の日報数は0～246件と幅がある。
- 社員による、日々の業務の改善提案
2016年4月～2016年9月までに社内ですされた改善提案、全8,180件。当社では、最低でも1人1件/月の改善提案を全社員が出しており、多岐に渡る提案が行われている。
- 適性検査「3Eテスト³」の結果
当社で開発している適性検査・適性診断である「3Eテスト」では、性格特性や職務適性などの項目を測定している。社員はこのテストを任意で受験しており、その検査結果を利用する。
- 360度評価結果
社員の自己評価、ならびに、その社員に対する上長および同僚・部下による評価を集めたデータ。評価内容は、目標必達志向性や問題発見力などの27項目。
- デジタル利用についての社内アンケート結果
Webサービスに詳しい社員を発見する目的で、過去に行われた社内アンケート。種々の最新アプリケーションを知っているかどうかを中心に調査している。回答は任意であり、353名の社員が回答。

なお、本論文中で説明のために用いるデータはダミーデータであり、実際のデータとは異なる。

5 各指標のスコア化

3節で述べた指標を、4節で述べた社内データをもとにスコア化する。各指標と、そのスコア化に利用するデータとの対応を表1に示す。

9種類の指標の内、〈主体性〉〈変革性〉〈意思伝達力〉〈行動性〉に関しては、3Eテスト結果に対応する項目が存在している。各項目は既に点数化されているため、その値をそのままスコアとして利用する。

〈目標必達志向性〉に関しては、360度評価結果に対応する項目が存在しており、ここでは該当項目の上

³<http://jinji-test.en-japan.com/>

表1: 各指標とそのスコア化に利用するデータとの対応

指標	社内データ(4節の番号に対応)					
	言語データ			その他のデータ		
	1	2	3	4	5	6
事業性						
デジタル知見						
学習能力						
主体性						
変革性						
目標必達志向性						
意思伝達力						
行動性						
独創性						

長からの評価を利用する。この項目は4段階で点数が付与されているが、上長によって評価の付け方が異なる。そのため、その上長がつけた平均評価点を中心として、全員の点数を付与し直し、〈目標必達志向性〉のスコアとする。

残りの4種類の指標に関しては、社内の言語データをもとにスコア化を行う。以下では、4種類の指標のスコア化手法を述べる。なお、スコア化を行う際に利用する形態素解析には、形態素解析器 MeCab [1]、および、新語辞書 mecab-ipadic-NEologd [4] を用いる。

5.1 事業性

事業性に関して感度が高い社員は、日報や改善提案、社員プロフィールにおいて、「新規」「事業」「利益」「売上」などの単語を多く用いていることが予測される。ここでは、人手で設定したキーワードをシードとして自動拡張し、日報や改善提案とのマッチングを行うことによるスコアリングと、事前に把握している、事業性に関して感度が高い社員数名が、社員プロフィールで特に多く用いている単語を利用したスコアリングの2種類の手法を用いる。それらのスコアの平均を〈事業性〉のスコアとする。

5.1.1 キーワードの自動拡張に基づくスコアリング

人手によってキーワードをいくつか設定し、そのキーワードを多く含む日報や改善提案を抽出することで、スコアリングを行う。しかしながら、人手だけで網羅的にキーワードをあげるのは難しい。そこで、人手であげたキーワードに対して、Word2Vec [2] を適用することで、キーワードの自動拡張を行う。Word2Vecの学習は、社員プロフィール、日報、改善提案の全言語データから行う。

〈事業性〉に関するキーワードのシードとして、「利益」「売上」「損益」「損益分岐点」「上場」「株価」「シェア」「市場」を利用する。Word2Vecにより、「営業利益」「為替」「売り上げ」などの類義語が得られる。本論文では、シードである単語との類似度が0.5以上の類義語のみを、拡張キーワードとして採用する。

これらのキーワードが日報内でどれだけ用いられているかにより、日報によるスコアを算出する。社員 m

の日報によるスコア $S_{\text{daily}}(m)$ は以下の式で算出する。

$$S_{\text{daily}}(m) = \frac{1}{N_{D_m}} \sum_{d \in D_m} \sum_{w \in d} \max_{k \in K} (\text{sim}(w, k))$$

D_m は社員 m による日報の集合、 d は D_m 中の日報、 w は日報 d に含まれる単語、 K はシードとなったキーワードの集合、 k は K に含まれるキーワード、 $\text{sim}(w, k)$ は単語 w とキーワード k との類似度、 N_{D_m} は、社員 m の日報数を表す。改善提案によるスコアも同様に算出し、それらを合わせることで、キーワードに基づくスコアを算出する。

5.1.2 事前に把握している社員数名の傾向に基づくスコアリング

新規事業担当者へのヒアリングによって、事業性に関して感度が高い社員が数名いることを事前に把握している。そこで、特に事業性に関して感度が高い社員 8 名をもとに傾向を分析し、傾向が近い社員のスコアが高くなるようにスコアリングを行う。

ここでは、社員プロフィールを「事業性に関して感度が高い社員」と「それ以外の社員」とに分割し、それぞれに含まれる名詞の頻度差を集計することでスコアを算出する。「事業性に関して感度が高い社員」中での頻度が高く、かつ、「それ以外の社員」との頻度差が大きい名詞ほど、「事業性に関して感度が高い社員」の傾向を表しているといえる。頻度差を用いることで、両方で多く出現する名詞を抑制することができる。例えば「新規」「事業」「マーケティング」が傾向を表す名詞として得られる。このような名詞を多く用いる社員のスコアが高くなるようにスコアを算出する。

社員 m のスコア $S_{\text{profile}}(m)$ は以下の式で算出する。

$$S_{\text{profile}}(m) = \sum_{w \in p_m} (\text{freq}_f(w) - \text{freq}_o(w))$$

p_m は社員 m のプロフィール、 w はプロフィール p_m 中に含まれる名詞、 $\text{freq}_f(w)$ は、「事業性に関して感度が高い社員」中での、単語 w の 1 プロフィール当たりの頻度、 $\text{freq}_o(w)$ は、「それ以外の社員」中での、単語 w の 1 プロフィール当たりの頻度を表す。

5.2 デジタル知見

デジタル領域に対して関心、知見をもっている社員は、日報や改善提案にも、「こういうアプリがあるので、業務に応用可能だ」といった記述をしていることが期待できる。そこで、5.1.1 節の手法と同様に、キーワードの自動拡張に基づくスコアリングを行う。〈デジタル知見〉に関するキーワードのシードとして、「デジタル」「人工知能」「スマートフォン」「スマホ」「Web サービス」「アプリ」「IoT」「ドローン」「AI」を利用することで、「ビッグデータ」「ロボット」「機械学習」といった拡張キーワードが得られる。

それに加えて、デジタル領域に関する知見をもっているかを測る基準として、デジタル利用についての社内アンケート結果における既知のアプリ数をスコアとして利用する。これら 2 種類のスコアの平均を〈デジタル知見〉のスコアとする。

5.3 学習能力

学習能力が高い社員は、自分の長所・強みが似た傾向になることが予想される。そこで、5.1.2 節の手法と同様に、学習能力が高いと評価されている社員 8 名が、社員プロフィール中で多く用いる単語（例えば「知識」「アルバイト」）を利用してスコアリングを行う。

それに加えて、3E テスト結果内の対応項目を利用する。3E テスト結果はすでに数値化されているので、これら 2 種類のスコアの平均をとることにより、〈学習能力〉のスコアを算出する。

5.4 独創性

さまざまな社員が業務に関する改善提案を出している中で、他の社員がしないような独創的な提案を多く出している社員は、独創性が高い社員と考えられる。そこで、改善提案データから独創的な提案の抽出を行うことで、〈独創性〉をスコア化する。

独創的な提案の抽出を行うために、全提案に対して形態素解析を行い、名詞を抽出、集計する。そして、1 つの提案内でのみ出現する名詞が、提案内に含まれる数によって、提案ごとの独創性スコアを作成する。これは、他の提案に出現しない単語を多く用いている提案は独創的な提案である、という仮説に基づいている。

しかしながら、頻度が 1 である名詞の中には、誤記、誤入力に由来する、存在しない名詞が含まれる可能性があり、そうした名詞が含まれる提案を独創的な提案とみなすのは適切ではない。そこで、頻度が 1 である名詞の内、以下の名詞のみを対象とする。

- 意味のある名詞である可能性が高い、Wikipedia に記事がある名詞（ただし、ひらがな 1 文字、カタカナ 1 文字の名詞、数値表現、人名、地名を除く）
- Wikipedia に記事はないが、意味のある造語・新語である可能性のある、カタカナのみの名詞、アルファベットの名詞（例えば「データサイエンティスト」「キュレーションメディア」「GooglePlay」など）

ここで対象とした名詞が提案内に含まれる数を、提案ごとの独創性スコアとする。そして、提案ごとのスコアを社員単位で合算し、その社員が出した提案数で平均をとることで、各社員に対する独創性スコアとする。

6 人手による評価

6.1 各スコアに対する定性的評価

各指標に関して、算出されたスコアが適切なものであるかどうか、定性的評価を行う。そのために、特定の部署に所属する社員のみを対象に、各スコアにおけ

表 2: スコア別ランキングに対するヒアリング結果

指標	所感
事業性	概ね違和感なし
デジタル知見	概ね違和感なし
学習能力	違和感あり
主体性	概ね違和感なし
変革性	違和感なし
目標必達志向性	違和感あり
意思伝達力	概ね違和感なし
行動性	違和感なし
独創性	独創性に富んだ人が 部署内では思いつかないため、 評価が難しい

表 3: 総合的なランキングに対する主観評価結果の分布 ; @ k は上位 k 人に対する評価を表し、各値は社員数、カッコ内はその割合を表す

評価点	本手法 @10	本手法 @50	本手法 @100	ランダム @100
5	3 (30%)	13 (26%)	13 (13%)	0 (0%)
4	6 (60%)	23 (46%)	26 (26%)	4 (4%)
3	0 (0%)	4 (8%)	27 (27%)	8 (8%)
2	0 (0%)	0 (0%)	0 (0%)	8 (8%)
1	0 (0%)	0 (0%)	0 (0%)	8 (8%)
判断不能	1 (10%)	10 (20%)	34 (34%)	72 (72%)

るランキングを作成し、どれだけ感覚と一致しているか、部署長にヒアリングを行った。結果を表 2 に示す。〈学習能力〉〈目標必達志向性〉はスコア化の手法を再考する必要があるものの、全体的に違和感のないランキングが得られていることが確認できた。

6.2 総合的な評価

9 種類の指標から算出したスコアをそれぞれ偏差値化し、それらの平均をとることで、各社員に対する総合的なスコアを算出した。そして、総合的なスコアによってランキングを作成し、上位の社員がイノベータ人財かどうかを手で評価した。ランダムでピックアップした場合と比較することで、本手法によってどれだけ効率的にピックアップが行えたかを検証する。

評価は、各社員に対して、イノベータ人財かどうかを 5 段階で主観判定することで行った。作業は、新規事業の担当者を含めた 2 名によって、判断の目安を議論した上で、新規事業の担当者 1 名が行った。評価者は全社員を把握しているわけではないため、判断できない場合には「判断不能」とした。把握しきれていないイノベータ人財をピックアップすることが目的であるため、ピックアップされた社員をすべて把握しているのでは意味がない。よって、評価点の高い社員が多くピックアップされるとともに、評価者が把握していない社員をある程度含む結果となることが望ましい。

表 3 に評価結果を示す。本手法によってピックアップした上位 10 人、50 人、100 人に対する評価結果に加え、ランダムにピックアップした 100 人に対する評価結果も併記している。「判断不能」とされた社員がいずれにおいても見られるが、その数を比較すると、本手法の方がランダムよりも少ない。本評価者は、新

規事業に関する社内勉強会も主催しているため、新規事業に興味を持った社員を多く把握している。そのため、「判断不能」の社員が少ないというのは、それだけ新規事業に興味を持った社員をピックアップできている、ということだと考えられる。評価点を比較すると、本手法では、高い評価点をもつ社員がピックアップできており、ピックアップする社員数が増えても、評価点が 2 点以下の社員はピックアップされていなかった。このことから、高い精度でイノベータ人財のピックアップができていたことがわかった。

最後に、評価者が「判断不能」とした社員の内、4 名に対して、評価者が実際にヒアリングを行った。その結果、2 名が 5 点、2 名が 4 点と、高い評価点をもつ社員であることが明らかになった。このことから、本手法によって効果的なピックアップができたといえる。

7 おわりに

本論文では、社内からイノベータ人財を抜け漏れなく効率的に発見することを目的として、社内データを活用したイノベータ人財のピックアップを行った。9 種類の指標を定義し、複数のデータ、複数の手法を用いてスコア化することで、適切なイノベータ人財のピックアップを行うことができた。

今後も、設計した指標の改善、新たな指標の追加、他に利用できそうなデータの検討などをすすめ、イノベータ人財をより精緻にピックアップしていく。また、例えばマネジメント人財のような、他の人財ピックアップに関しても、指標を吟味し、適切なデータを選択することで、同様にピックアップが可能だと考えられる。イノベータ人財ピックアップの改善をすすめるとともに、他の人財のピックアップにも取り組む。

参考文献

- [1] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to Japanese morphological analysis. In *Proceedings of EMNLP*, pp. 230–237, 2004.
- [2] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems 26*, pp. 3111–3119. 2013.
- [3] Marco Pennacchiotti and Ana-Maria Popescu. A Machine Learning Approach to Twitter User Classification. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 2011.
- [4] Toshinori Sato. Neologism dictionary based on the language resources on the Web for Mecab, 2015.
- [5] 谷口元樹, 出雲英剛, 杉原大悟, 満岡淳, 長谷川宏, 大熊智子. SFA データ活用に向けた商談事例検索システムの検討. 言語処理学会第 22 回年次大会ワークショップ 言語処理の応用, 2016.
- [6] 榊剛史, 松尾豊. ソーシャルメディアユーザの職業推定手法の提案. 知能と情報, Vol. 26, No. 4, pp. 773–780, 2014.
- [7] 池田和史, 服部元, 松本一則, 小野智弘, 東野輝夫. マーケット分析のための twitter 投稿者プロフィール推定手法. 情報処理学会論文誌コンシューマ・デバイス&システム (CDS), Vol. 2, No. 1, pp. 82–93, 2012.