

Web 上記事タイトルによる AI スタートアップ企業のクラスタリング

溝口 隼人

Edge AI Research Center 研究員
h.mizoguchi@edge-arc.jp

工藤 俊太郎

Edge AI Research Center 研究所長
kudo@edge-arc.jp

1 はじめに

近年の AI ブームもあり、日本においても多くの AI スタートアップ企業が誕生している。しかし、企業数の増加や各企業が AI の事業応用分野を増加させていることなどから、各企業が得意とする領域や、企業間での相対的な類似性・特異性を人手で把握するのは難しくなっていると思われる。それらを把握することは、提携先企業の選定や AI のトレンドの調査などの際に重要な要素である。

そこで本稿では、Web 上で検索可能な各 AI スタートアップ企業に関するニュース記事のタイトルから、各企業のクラスタリングを試みる。Web でのニュース検索結果は企業調査の際に最もよく参照されるものと思われるため、それを基にクラスタリングを行うことで、企業の世間一般的なイメージを明らかに出来ると考えられる。

そして、クラスタリング結果を基に、調査対象企業間の相対的な技術面・事業応用分野面での特異性や類似性が、Web 上ニュース記事のタイトルという一般的かつ限られた情報から抽出可能であるかを考察する。

2 検証手法

2.1 対象企業

国内 AI スタートアップ企業は数多く存在し、そのすべてを解析対象とすることは困難である。そこで本稿では、対象とする AI スタートアップ企業は経済産業省のスタートアップ企業支援プログラムである「J-Startup」選定企業であり、かつ業種が「AI/制御」である 18 企業とした [1]。

2.2 対象記事

各企業のニュース記事の検索には Google ニュース検索を用いた。検索ワードは、類似した名前の企業等の記事が検索でヒットするのを防ぐため、企業名にスペースの後「AI」を加えたものとした。検索結果の上位 50 件の記事を対象とし、明らかに当該企業と関係のない記事は人手で取り除いた。

2.3 文章のベクトル化とクラスタリング手法

各企業の解析対象記事タイトル群は品詞分解し、名詞のみを抽出する。名詞のうち、単独では意味をなさない数詞等はストップワードとして除外する。そのようにして抽出した各企業の記事タイトルに含まれるのキーワード群を、TF-IDF を用いてベクトル化する。そして、各企業の特徴量を k-means 法によりクラスタリングするとともに、主成分分析 (PCA) を用いて 2 次元プロットし、各企業の相対的関係の図示を試みた。

3 実験結果と考察

3.1 実験結果

k-means 法の適用に際し、クラスタ数は 8 とした。各クラスタを構成する企業の一覧を表 1 に示す。また、文書のベクトルを PCA により 2 次元に圧縮し、クラスタ別に色を分けてプロットしたものを図 1 に示す。

3.2 考察

AI-OCR サービスを提供する株式会社シナモン/株式会社アジラ/株式会社 Cogent Labs (Cluster 0) や、エッジ AI・エッジコンピューティングサービスを提供する株式会社アラヤ/株式会社 Idein/株式会社エイシング (Cluster 4)、深層学習の事業応用に取り組む LeapMind 株式会社/株式会社 Preferred Networks (Cluster 3)、また、医療系 AI ベンチャーのアイリス株式会社/メドメイン株式会社 (Cluster 7) が同じクラスタに分類されるなど、各企業の事業内容に反しない企業クラスタリングが、ニュース検索結果キーワードからでも可能であることが示唆された。

4 まとめ

本稿では、Web 上で検索可能なニュース記事タイトルを基に AI スタートアップ企業のクラスタリングを試行した。ニュース記事タイトルにも各企業の事業領域や技術領域が現れており、それに基づいたクラスタリングにより各企業の相対的な技術面・事業応用分野面での特異性や類似性について知見が得られる可能性があることを確認した。

表 1 クラスタを構成する企業一覧

Cluster	Company
0	BizteX 株式会社
1	株式会社 Cogent Labs, 株式会社アジラ, 株式会社シナモン
2	株式会社 ABEJA, 株式会社 PKSHA Technology, 株式会社エクサウィザーズ, 株式会社オルツ
3	LeapMind 株式会社, 株式会社 Preferred Networks
4	Idein 株式会社, 株式会社アラヤ, 株式会社エイシング
5	株式会社 Empath
6	AlpacaJapan 株式会社
7	アイリス株式会社, メドメイン株式会社

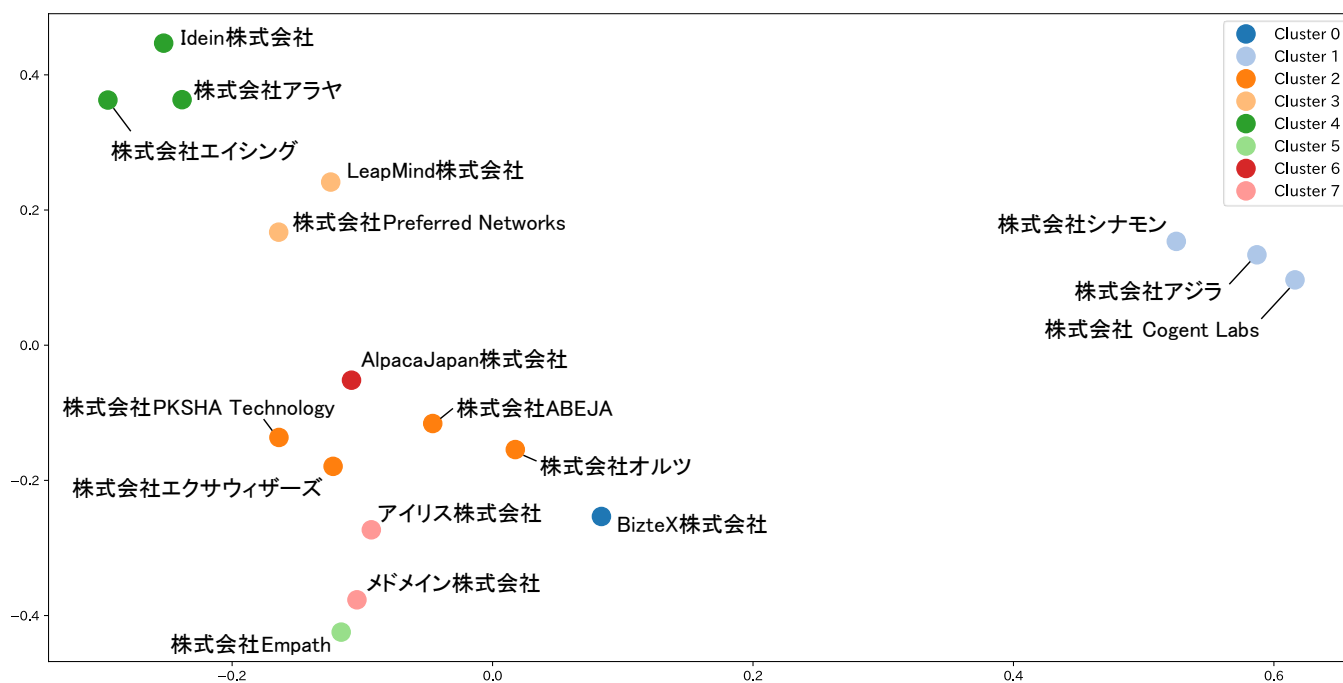


図 1 文書ベクトルの PCA による 2 次元プロット

参考文献

- [1] J-Startup, “スタートアップ企業”, <https://www.j-startup.go.jp/startups/>