

企業間の相違や時間的な変化を考慮した IR情報からの戦略ワード抽出法の提案

Proposal of a Method for Extracting Strategy Words from IR Information of Enterprises
with Considering Differences among Enterprises and Changes in Time Sequence

鎌田裕嗣 (東京都市大学) 峯田誠也 (東京都市大学大学院) 岡田公治 (東京都市大学)

1. 背景および目的

企業評価を行う上で、財務状況のみでなく、企業の戦略やその変化を捉えることが重要である。本稿では、投資家に積極的に訴求したい今後の戦略や施策が含まれる IR 情報から、形態素解析では抽出困難な造語や固有語も戦略ワードとして抽出可能な階層化 N-Gram手法を基に、企業間の相違や時間的な変化も含めて戦略ワードを抽出する手法を提案する。

2. 提案手法

戦略ワードの出現パターン

戦略ワード：企業が積極的にアピールしたい語
⇒企業間の相違や時間的な変化に特定のパターンがあると考えた。
電機8社、83件のアニュアルレポートを目視で確認し以下の**出現パターン**を特定

- (1)業界動向型戦略ワード (例 クラウド、ビッグデータ)
⇒ある時点から業界全体で使用されるワード
- (2)特定企業型戦略ワード (例 エコ・ポジティブカンパニー)
⇒製品・サービス名等、特定の企業だけで使用される
- (3)特定期間型戦略ワード (例 エコポイント、東日本大震災)
⇒政策や災害対策等、特定の時期に業界内外で使用される

階層化N-Gram手法 (峯田2016)

1. N文字列長とN+1文字列長との包含関係により出現回数を補正

2. 補正された回数文書中に出現する事象発生確率から重要度指標を算出

3. 重要度指標が閾値以上のワードを戦略ワード候補として出力
- 適用例 (中外製薬アニュアルレポート)
- | 戦略ワード | 出現頻度 | 重要度指標 |
|-----------------------|------|-------|
| 個別化医療 | 25 | 1 |
| 7つの強み | 22 | 1 |
| 新中期経営計画IBI18 | 21 | 1 |
| 「アレセンサ」 | 17 | 1 |
| すべての革新は患者さんのために | 8 | 0.999 |
| ヒト化モノクローナル抗体 | 9 | 0.999 |
| ロシュ・グループとの戦略的アライアンス体制 | 7 | 0.996 |
| バイオニアとしての個別化医療の知見 | 7 | 0.996 |
| バイオをはじめとする独自の創薬技術 | 6 | 0.977 |
- 固有名詞・造語も抽出可能
しかし不要なワードも多く含まれる

自動抽出アルゴリズム

複数企業、複数年分のアニュアルレポートから階層化N-Gramにより戦略ワード候補を抽出

①業界動向型

	'10	'11	'12	'13	'14	'15
A社	0	0	0	0.68	1.00	1.00
B社	0	0	0	0.99	1.00	1.00
C社	0	0	0	0	0.62	0
D社	0	0	0.68	0.97	0	0.88
E社	0	0.63	1.00	1.00	1.00	1.00

①-a 拡散期 ①-b

②特定企業型 および ③特定期間型

	'10	'11	'12	'13	'14	'15
A社	0	0	0	0	0	0
B社	0.85	0.63	0.81	0.88	0	0
C社	0	0	0	0	0	0
D社	0	0	0	0	0	0
E社	0	0	0	0	0	0

③-a ③-b ③-a

②-a ②-b ②-a

パターン毎に以下の2群に分け、重要度指標の平均値に有意な差があったワードを抽出

- (1)業界動向型戦略ワード
⇒時系列の両端 (左図 ①-a と ①-b)
- (2)特定企業型戦略ワード
⇒1社とその他の企業 (左図 ②-a と ②-b)
- (3)特定期間型戦略ワード
⇒一部期間とその他の期間 (左図 ③-a と ③-b)

3. 適用結果

IT業界7社、2005年～2015年のアニュアルレポートに適用

(1)業界動向型戦略ワード

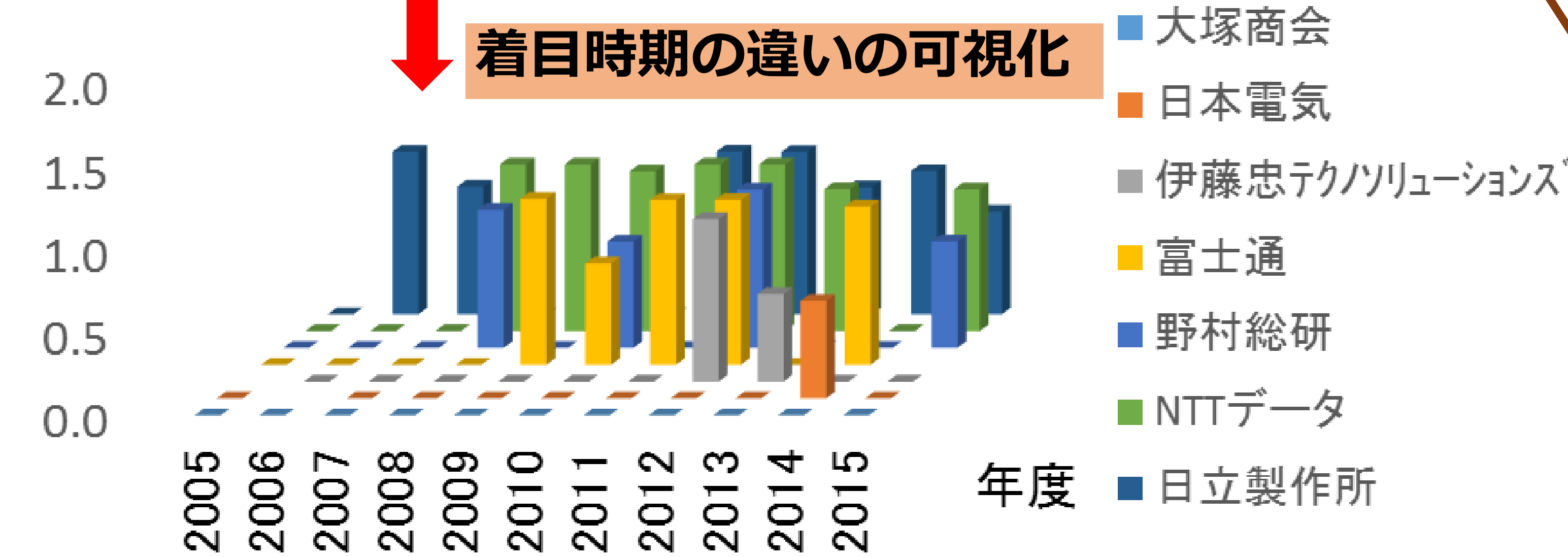
戦略ワード	拡散期
グローバル化	2006～2014
アウトソーシングサービス	2006～2014
クラウド	2009～2014
スマートフォン	2011～2014
ビッグデータ	2012

(2)特定企業型戦略ワード

戦略ワード	企業	出現年度
オフィスサプライ通信販売事業「たのめーる」	大塚商会	2005～2015
NECグループビジョン2017	日本電気	2008～2014
社会イノベーション事業	日立製作所	2007、2009～2015
「らくらくホンシリーズ」	富士通	2007、2010～2013
横浜第二データセンター	野村総合研究所	2007～2011、2013～2015

(3)特定期間型戦略ワード

戦略ワード	出現年度
個人情報保護法	2005
東日本大震災	2011
タイの洪水	2012
障がい者雇用	2015
マイナンバー制度	2015



辞書を前提とした形態素分析では抽出が難しい造語や固有名詞も抽出

法規改定や災害対策等に関するワードも抽出

戦略ワードと見なせない不要なワードの削減

対象としたアニュアルレポート数	74
階層化N-gram実行後 ワード数	264565
プログラム実行後 ワード数	6947
意味のある戦略ワード	およそ約200

不要なワード
約95%削減

4. 結論

時間的な変化点や企業間の相違点に着目することで 3 種類の戦略ワード出現パターンを特定した。階層化 N-gramによって得られた戦略ワード 候補から、戦略ワード出現パターンに該当する戦略ワード候補のみを自動抽出することで、不要なワードを大幅に削減できた。