

企業のIR情報からの戦略ワードの抽出法の提案

A Proposal of a Method for Extracting Strategy Words from IR Information of Enterprises

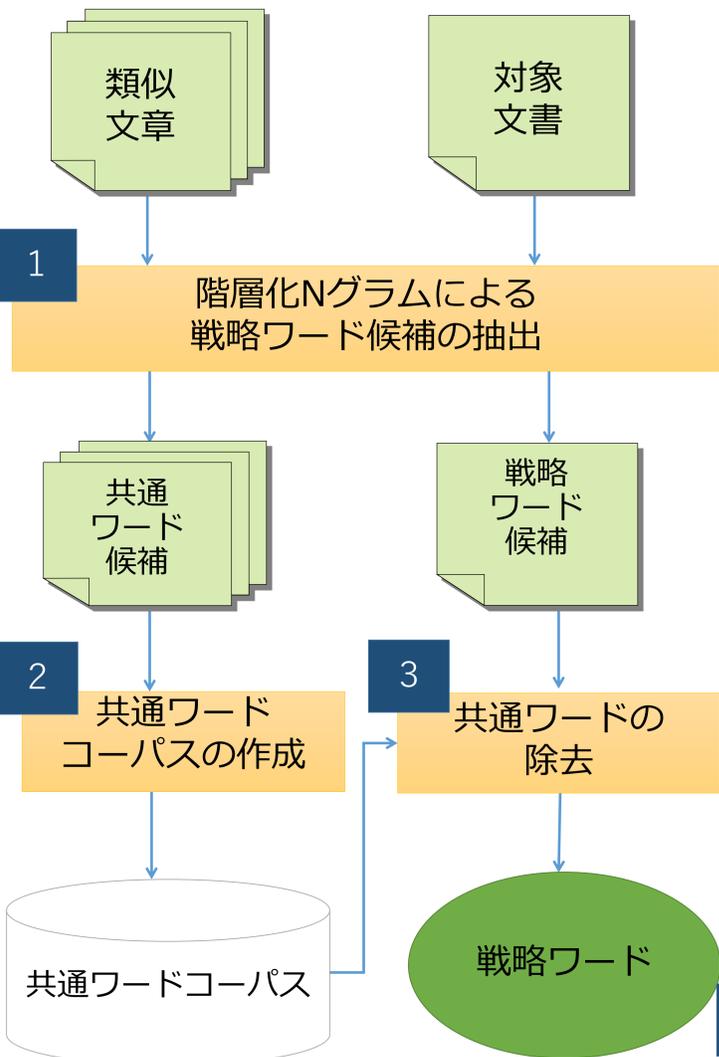
峯田誠也 (東京都市大学大学院) 岡田公治 (東京都市大学)

1. 背景および目的

*戦略ワード = IR文書中に強調され繰り返し記述されるワードやフレーズ

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、企業の投資家向け公開情報 (IR情報) から企業の戦略ワード* を抽出することを試みる。企業の戦略ワードは造語や固有語等、企業が独自に作成したワードが多く、辞書を用いた形態素解析では抽出が困難である。そこで、N-Gramを用いた抽出法を提案する。

2. 提案手法

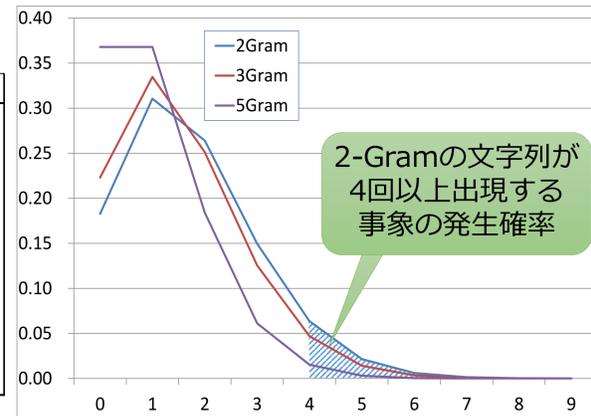


戦略ワード抽出法の概要

1 Step1 文字列の出現確率

Gram数ごとの文字列の出現確率をポアソン分布より求め、実際に出現したワードの出現回数がどの程度起こりにくいかを算出した。

グラム数	2-Gram	3-Gram	...	5-Gram
ワード	企業 業評 評価 を行 う上 で、 財	企業評 業評価 を行 う上 で、 財	...	企業評価 を行 う上 で、 財 務状 況



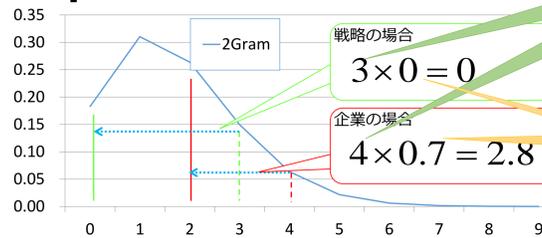
Step2 文字列の独立性係数

N-Gram文字列の階層的な包含関係をもとに独立性係数を求める。

$$\frac{3}{3} \times \frac{1}{2} = 0 \quad \frac{1}{2} \times \frac{4}{1} = 0.7$$

前後の文字列に対して、完全に従属しているものは0, 独立しているものほど1に近くなる。

Step3 文字列の重要度指標

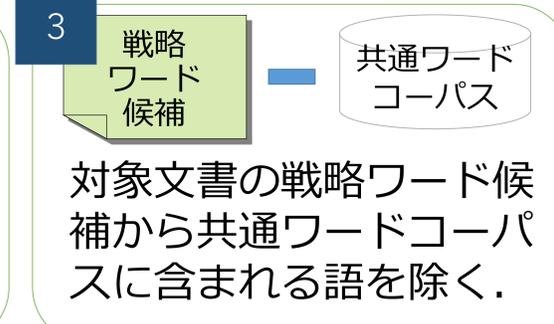
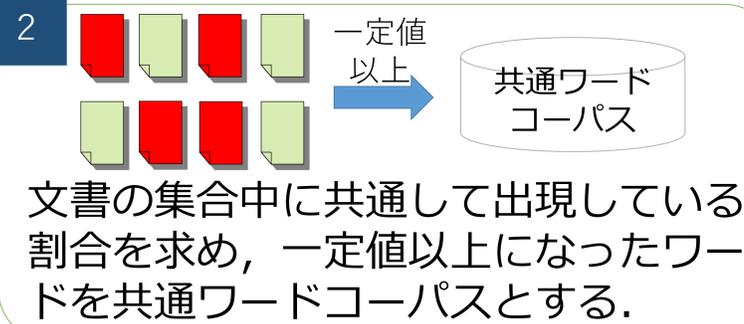


出現回数 q_w

重要度指標

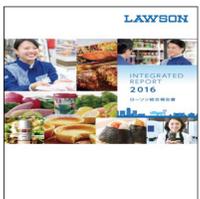
$$S_{d,w} = 1 - P(X \geq q_w I_w) = \sum_{i=0}^{\text{Ceil}(q_w I_w) - 1} P(X = i)$$

独立性係数 I_w



3. 適用事例

2社のアニュアルレポートを適用



戦略ワード	出現頻度	重要度指標
(ローソン)		
健康	69	1.000
1000日全員実行プロジェクト	27	1.000
次世代コンビニエンスストア	21	1.000
小商圏型製造小売業	14	1.000
ローソン型次世代コンビニエンスストア	14	1.000
セミオート発注システム	10	1.000
次世代コンビニ	22	0.999
高度な店舗生産性	6	0.994
まちかど厨房	6	0.992
“みんなと暮らすマチ”を幸せにします。	6	0.979

戦略ワード	出現頻度	重要度指標
(中外製薬)		
個別化医療	25	1.000
7つの強み	22	1.000
新中期経営計画IBI18	21	1.000
「アレセンサ」	17	1.000
すべての革新は患者さんのために	8	0.999
ヒト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
アンメットメディカルニーズを充足する製品カ	7	0.996
パイオニアとしての個別化医療の知見	7	0.995
バイオをはじめとする独自の創薬技術	6	0.979

結果として、企業が打ち出している**戦略施策名**、**製品名**や**システム名**、**企業理念を表すフレーズ**等を戦略ワードとして抽出することができた。これらには形態素解析では抽出が困難な、**接続詞を含むフレーズ**や**固有名詞**を含むものも抽出できている。

戦略ワード	出現頻度	重要度指標
店舗	324	1.0000
医療	114	1.0000
コンビニ	108	1.0000
カル	50	1.0000
につきましては、	43	1.0000
(年度)	41	1.0000
※1	29	1.0000
かり	18	1.0000
mp	34	1.0000

今後の課題
除去しきれしていないノイズや、戦略ワードとしては特徴のないワードも残存。

4. 結論

N-Gramを用いた、IR情報から造語や固有語を含む企業の戦略ワードの抽出法を提案した。今後、本稿の抽出手法を用いて、同一業界のIR情報の比較や同一企業の年度が違うIR情報の比較を行い、不要なワードと戦略ワードの違いを認識する手法の提案も行いたい。