

定性データに基づく企業構造の境界値分析

独立行政法人統計センター 白川清美

1. はじめに

企業構造層化のための境界値設定では、法令、各業界の景気や個々の企業の経理状況等の情報が有効である。特に、財務諸表等のパネルデータ化（時系列化）により、企業構造の層化が容易に出来る。しかしながら、公的経済調査において、詳細な経理状況を経常的に把握することには困難が多い。また、調査項目のみでは、企業構造の層化に限界がある。そこで、企業構造の層化に必要な情報を補うため、企業の業績、事業内容や企業が属する業界の景気に関する定性データ（文字情報）を活用する。さらに、表 1 より、売上営業利益率（売上高に対する営業利益の割合）による産業別格差が明確である。

本研究では、企業構造の一つの境界値として「産業分類別利益の有無」による層化を試行する。なお、企業構造層化の目的は、売上高の欠測値の補定や外れ値の検出である[1][2][3]。

表 1 産業大分類別売上営業利益率 平成 24 年経済センサス速報集計結果

産業大分類	*1	A~B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
売上営業利益率(%)	*2	3.5	25.7	2.6	3.9	2.2	9.7	4.6	2.8	-	9.6	11.4	6.5	4.9	9.3	5.0	2.7	5.0

*1 A~Rは産業大分類を表す記号、 *2 売上利益率=営業利益（売上高-売上原価-販売管理費）÷売上高×100

2. 境界値分析とその評価

EDINET データ（有価証券報告書）から取得した企業の名称や事業内容等の定性データを、R 言語を用いた形態素解析により名詞等の単語を抽出した。その後、それらの単語を基に、産業細分類符号付のための辞書と照合し、当該企業が属する業界に分類した。その業界の景気動向といくつかの指標により、企業が属する業界別境界値（利益の有無）の分析を行った。以下の図 1 に、産業分類と利益に関する指標の一例を示した。

産業分類	企業名：分類 ← RMeCabC(“O×薬品株式会社”)，事業内容：分類 ← RMeCabC(“医薬品の製造をしている。”)
指標	①損益分岐点 (BEP) = 固定費 ÷ (1 - (変動費 ÷ 売上高))，②安全余裕率 = [(売上高 - 損益分岐点売上高) ÷ 売上高] × 100 ③限界利益率 (%) = (限界利益 ÷ 売上高) × 100 *限界利益 = 売上高 - 売上高 × 変動比率

図 1 産業分類と利益の有無を判別するために有効な指標の一例

最後に、EDINET データにある損益計算書との照合により、今回の試行の検証を行った。この結果、層別境界値による各企業の判別ができたことを確認した。

3. 今後の取組み

EDINET データの分析の結果、法令等に基づいた企業構造の層化[1]以外に、経理項目からの層化へのアプローチが有効であると確認した。それゆえ、公的経済調査の調査項目（主に、「資本金」、「売上高」、「従業者数」、「企業の経営組織」、「産業分類」）以外の項目を、インターネット等から取得した定性データを用いて企業の層化を行う。なお、詳細な層化の境界値分析は、平成 25 年 8 月公表（予定）の「平成 24 年経済センサス-活動調査速報集計結果」を用いる予定である。

参考文献

- [1] 伊藤孝之, 白川清美. (2013). 「経理項目に基づく企業の構造化分析-構造的境界値検定-」, 2013 年度統計関連学会連合大会講演報告集, 大阪大学豊中キャンパス.
- [2] Shirakawa, Kiyomi. (2013). “A Post-Aggregation Error Record Extraction Based on Naive Bayes for Statistical Survey Enumeration” *The 59th World Statistics Congress, Hong Kong, China, 25-30 August 2013.*
- [3] 白川清美. (2012). 「経済センサス-基礎調査集計結果に基づく「企業グループ」に関する考察」, 2012 年度統計関連学会連合大会講演報告集, 北海道大学.

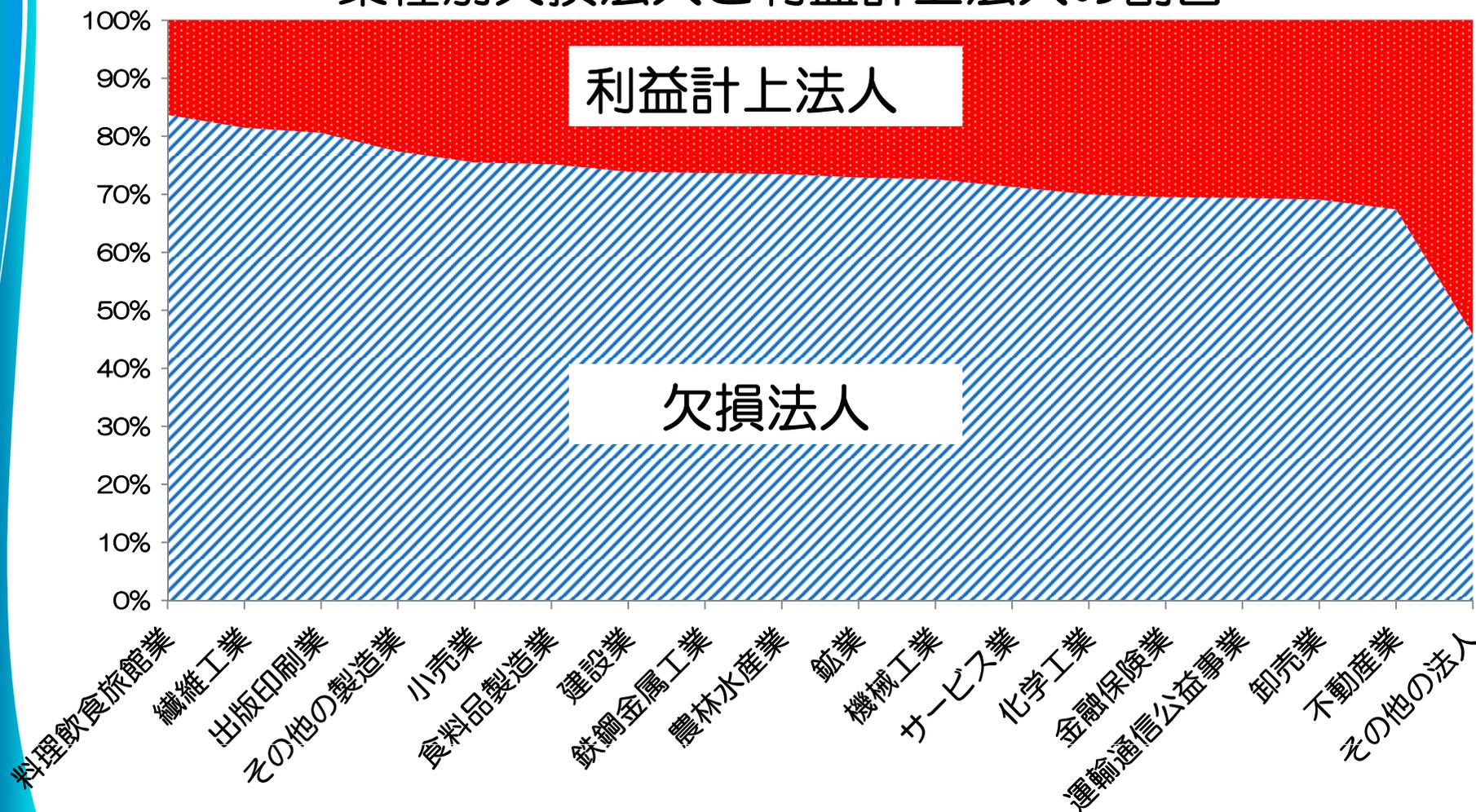
定性データに基づく企業構造の 境界値分析

2013年度統計関連学会連合大会
大会日程：2013年9月9日（月）
場 所：大阪大学 豊中キャンパス

独立行政法人統計センター 白川清美
日本行動計量学会 正会員

なぜ境界値分析が必要なのか？

業種別欠損法人と利益計上法人の割合



出所：会社標本調査結果（税務統計から見た法人企業の実態），国税庁

目次

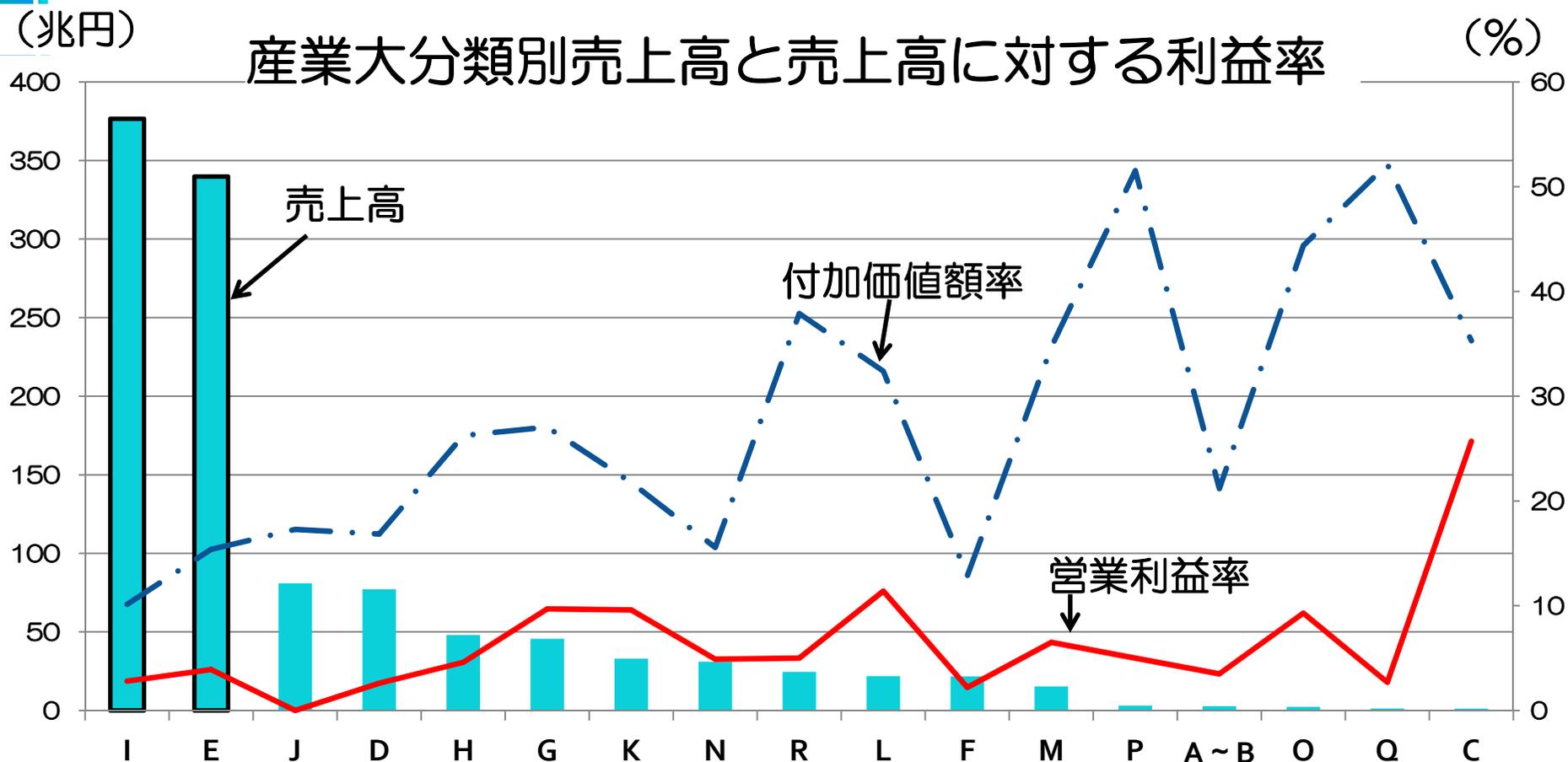
- 目的
- 背景
- 定性データによる境界値の検証
- 検証の結果
- 検証のまとめ
- 今後の課題
- 「経済センサス」への期待

目的：企業に関する定性データ に基づいた最適「層」への分類

「産業分類別営業利益の有無」を境界値とする層化の試行

1. 産業分類により異なる利益率の活用
 - ▶ 利益率の高い産業
 - ✓ **医薬品製造業**，ネット関連企業
 - ▶ 利益率の低い産業
 - ✓ 卸売業，鉄鋼，海運，電力，繊維，建設
2. 産業別事業内容や業績の定性データの活用
 - ▶ 産業別コーパス（Corpus）の作成
 - ✓ 「営業利益」別の頻出単語

売上高に対する境界値の指標： 産業分類の「利益率」による層化が有効



*1 A~Rは産業大分類を表す記号,

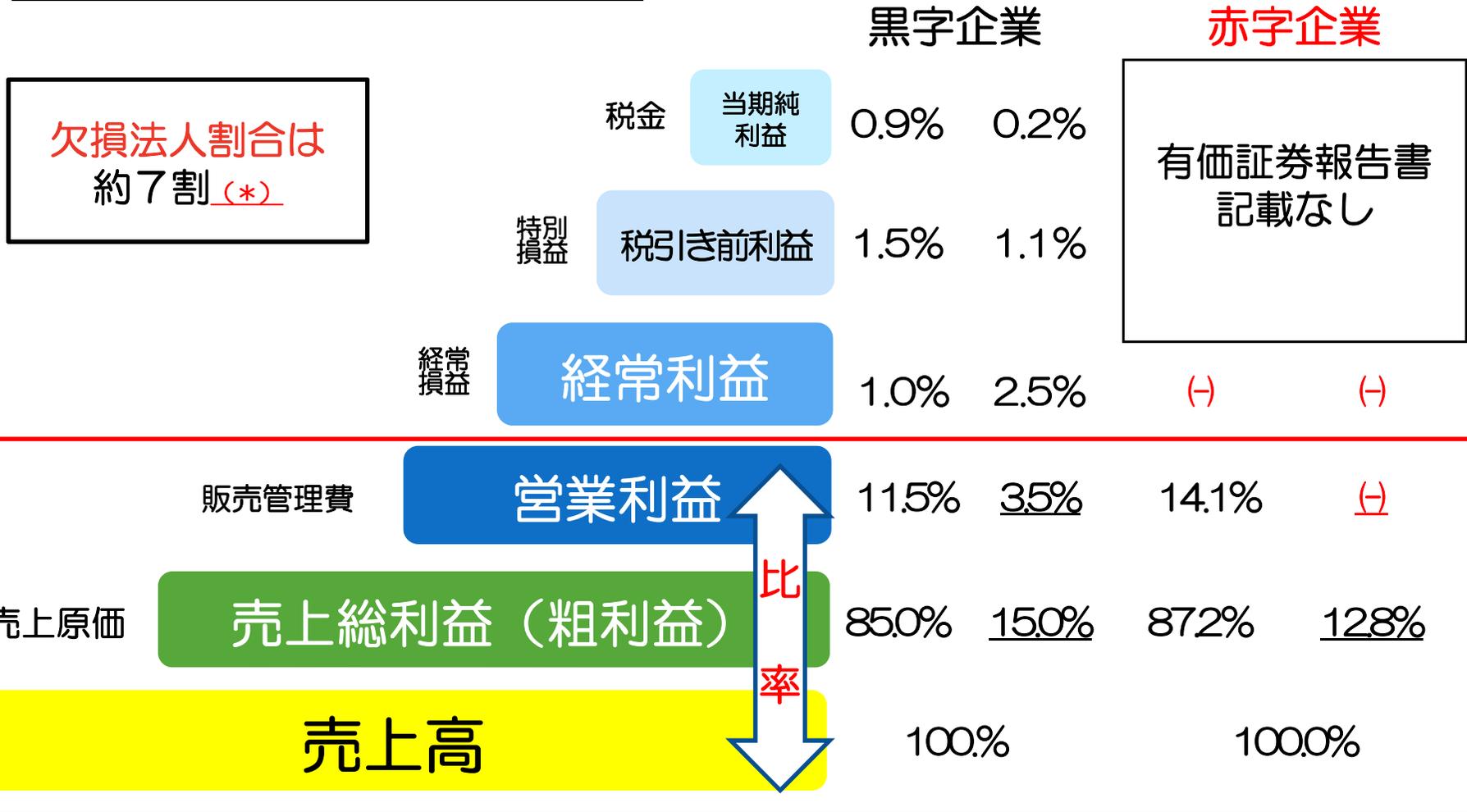
*2 売上利益率 = 営業利益 (売上高 - 売上原価 - 販売管理費) ÷ 売上高 × 100

*3 付加価値額率 = (売上高 - 費用総額 + 給与総額 + 租税公課) ÷ 売上高 × 100

出所：総務省統計局 平成24年経済センサス速報集計結果 企業編第8表

背景： 損益により異なる売上高に対する構成比

黒字企業と赤字企業の相違



(*) 会社標本調査結果 (税務統計から見た法人企業の実態), 国税庁

定性データによる境界値の検証

1. 企業に関するデータ

- EDINET (Electronic Disclosure for Investors' NETwork)
 - ✓ 上場企業 3,558社 (H23年4月~24年3月決算)
- 利用箇所
 - ✓ 企業名, 事業内容, 業績, 損益計算書

2. プログラム言語等

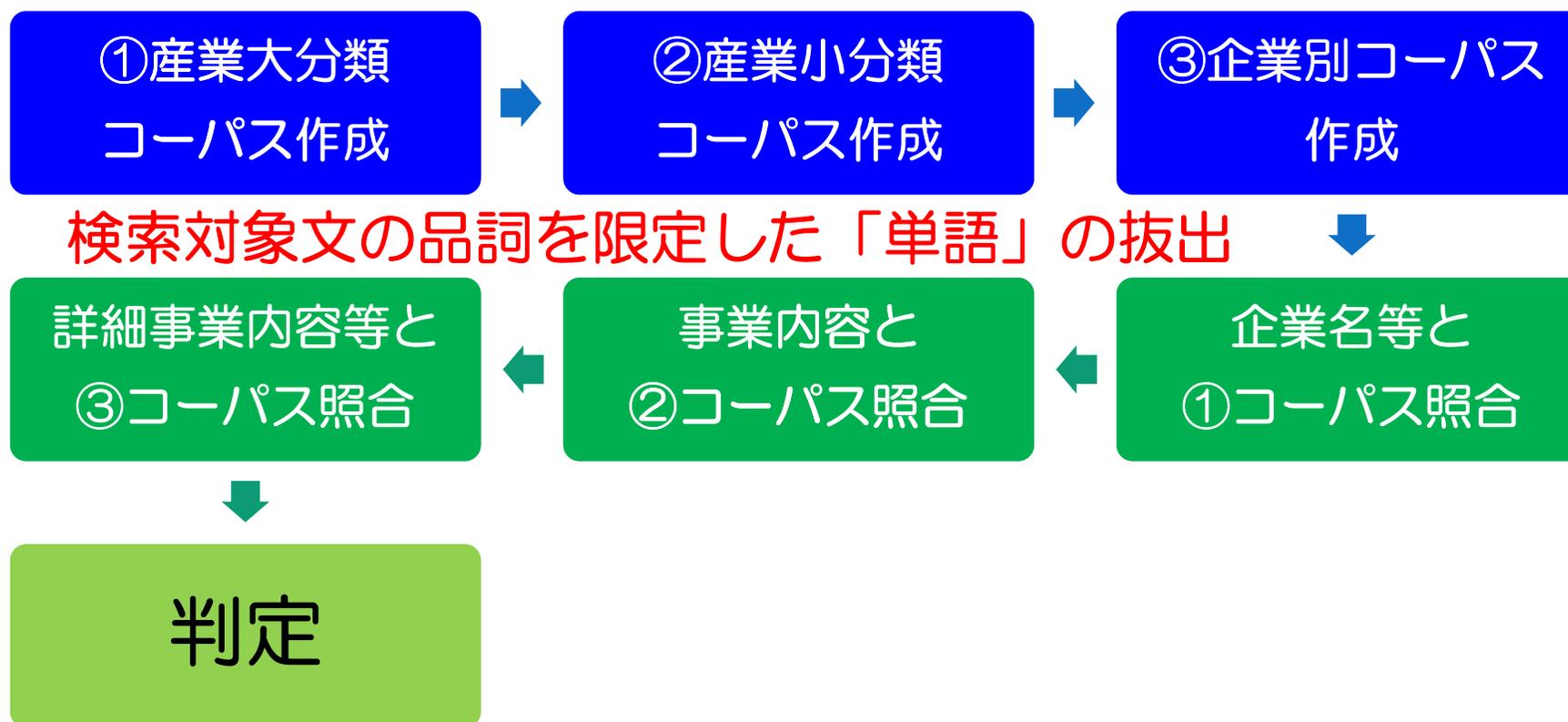
- 汎用統計ソフトウェア「R」, MeCab
- Rパッケージ「RMeCab」, 潜在意味解析プログラム^{*1}
^{*1} 豊田秀樹『データマイニング入門』東京図書 (2008)

3. 検証対象「層」

- 「産業小分類別営業利益の有無」を境界値とする分類

各種コーパス作成から判定までの流れ

形態素解析によるコーパスの作成



①産業大分類コーパス： 産業分類の項目名を活用

産業分類「E製造業」の抜粋

中分類	09	食料品製造業	
	090	管理，補助的経済活動を行う事業所 (09食料品製造業)	特徴がない 情報
	0900	<u>主として管理事務を行う本社等</u>	
	0909	<u>その他の管理，補助的経済活動を行う事業所</u>	
	091	畜産食料品製造業	否定形はな い方がよい
	0911	部分肉・冷凍肉製造業	
	0912	肉加工品製造業	
	0913	処理牛乳・乳飲料製造業	
	0914	乳製品製造業 <u>(処理牛乳，乳飲料を除く)</u>	
	0919	その他の畜産食料品製造業	

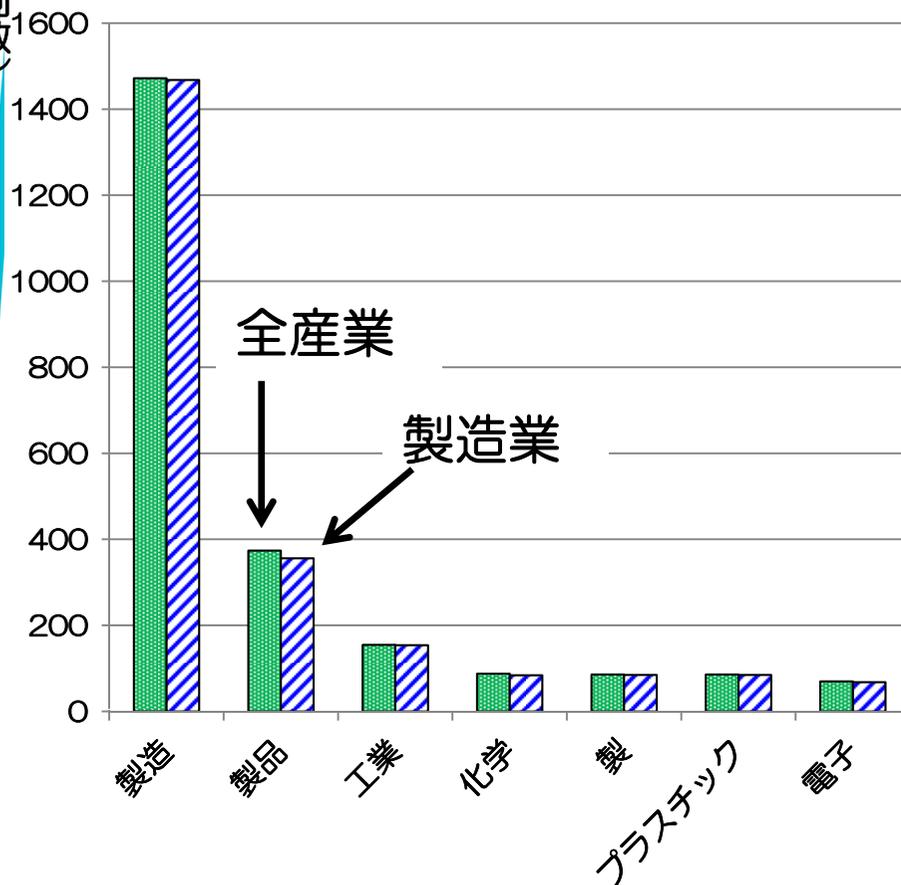
）

出所:産業分類，「平成24年経済センサス - 活動調査 産業分類一覧」，総務省統計局
<http://www.stat.go.jp/data/e-census/2012/sokuho/bunrui.htm>

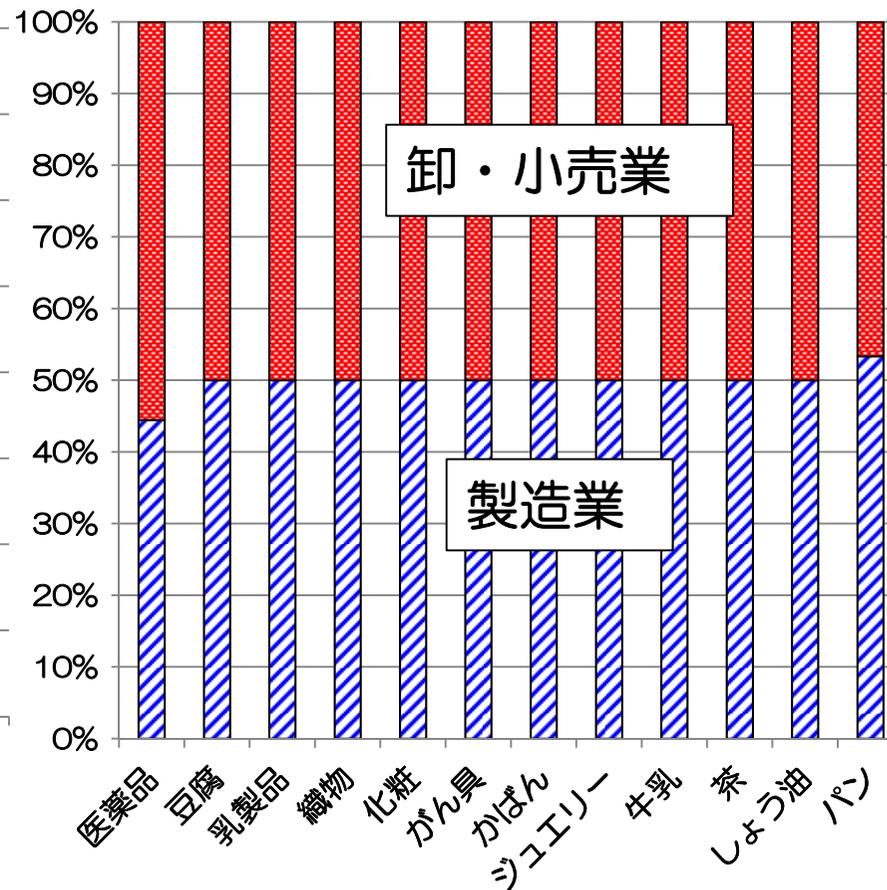
産業大分類に特化した単語

(出現回数)

製造業に特化した単語



製造業、卸・小売業に特化した単語



出所:産業分類, 「平成24年経済センサス - 活動調査 産業分類一覧」, 総務省統計局
<http://www.stat.go.jp/data/e-census/2012/sokuho/bunrui.htm>

「産業大分類」の絞り込み

1. 産業大分類のコーパス作成
産業小分類の項目名を活用
2. 対象企業の名称または事業内容との照合
 - ① bunrui ← RMeCabC(“〇×**薬品**株式会社”)
 - ② bunrui ← RMeCabC(“**医薬品**の**製造**をしている。”)
3. Rでの実行結果 (名詞, 形容詞を抽出)
 - ①> bunrui “薬品” (名詞) “会社” (名詞)
 - ②> bunrui “医薬品” (名詞) “製造” (名詞)
4. 判定結果
潜在意味解析による「産業大分類」の格付け
 - ①, ②のどちらも, 全産業の「E製造業」に分類

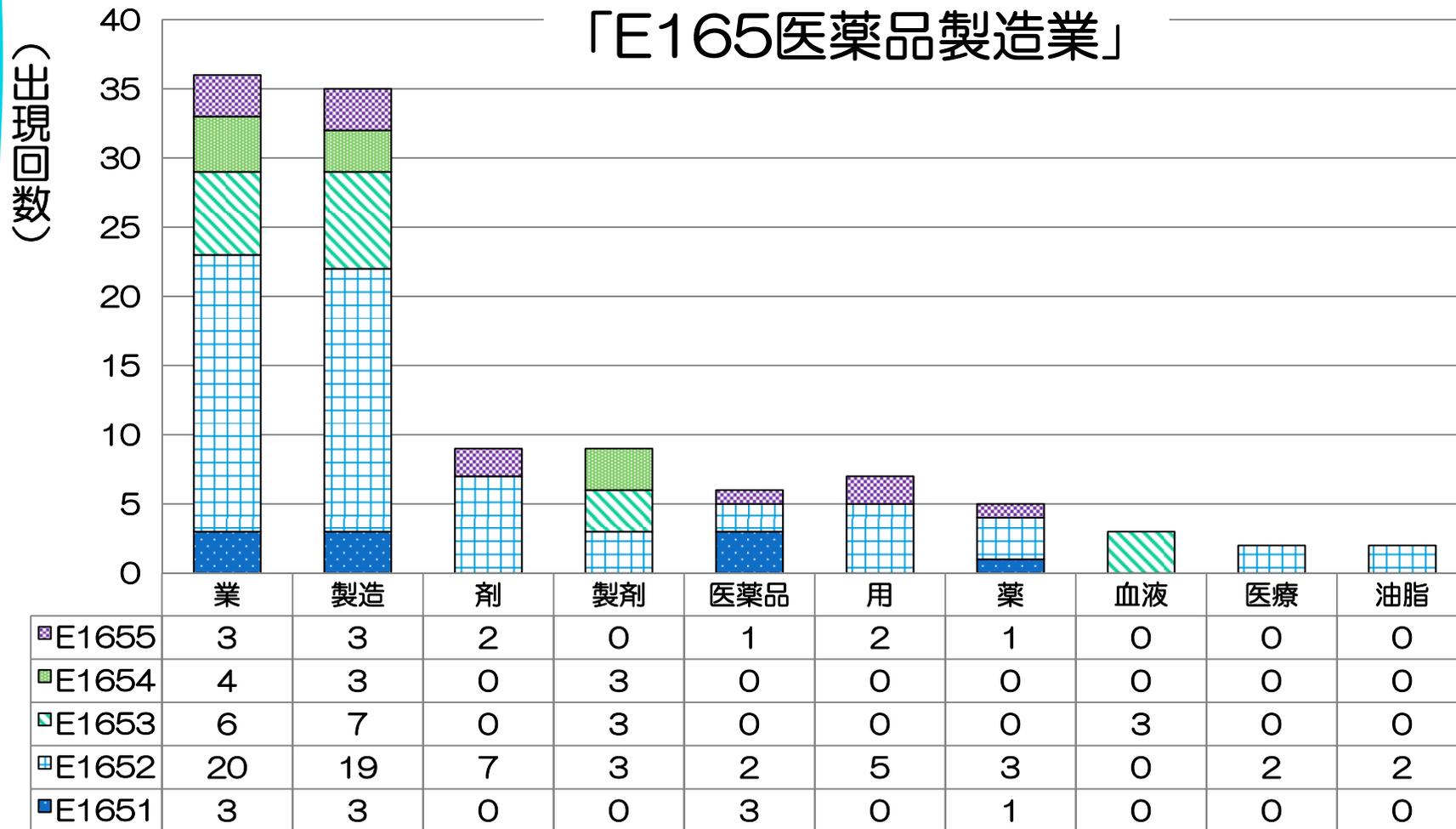
②産業小分類コーパス： 事業の内容例示を活用

産業小分類「E165 医薬品製造業」の内容例示

- 1651 医薬品原薬製造業
医薬品原末製造業；医薬品原液製造業
- 1652 医薬品製剤製造業
内服薬製造業；注射剤製造業；外用薬製造業；殺虫・殺そ
（鼠）剤製造業；蚊取り線香製造業；殺菌・消毒剤製造業；診
断用試薬製造業；医療用植物油脂製造業；医療用動物油脂製造
業；薬用酵母剤製造業
- 1653 生物学的製剤製造業
ワクチン製造業；血液製剤製造業
- 1654 生薬・漢方製剤製造業
生薬製造業；漢方製剤製造業；生薬小分け業
- 1655 動物用医薬品製造業
繁殖用薬製造業；飼料添加剤製造業（成長促進剤など）

不要

頻出単語と産業小分類の内訳



出所:産業分類「平成24年経済センサス - 活動調査産業分類一覧」総務省統計局
<http://www.stat.go.jp/index/seido/sangyo/pdf/19san3e.pdf>

③ 企業別コーパス： 企業の詳細な事業内容や業績を活用

【事業内容の例示】

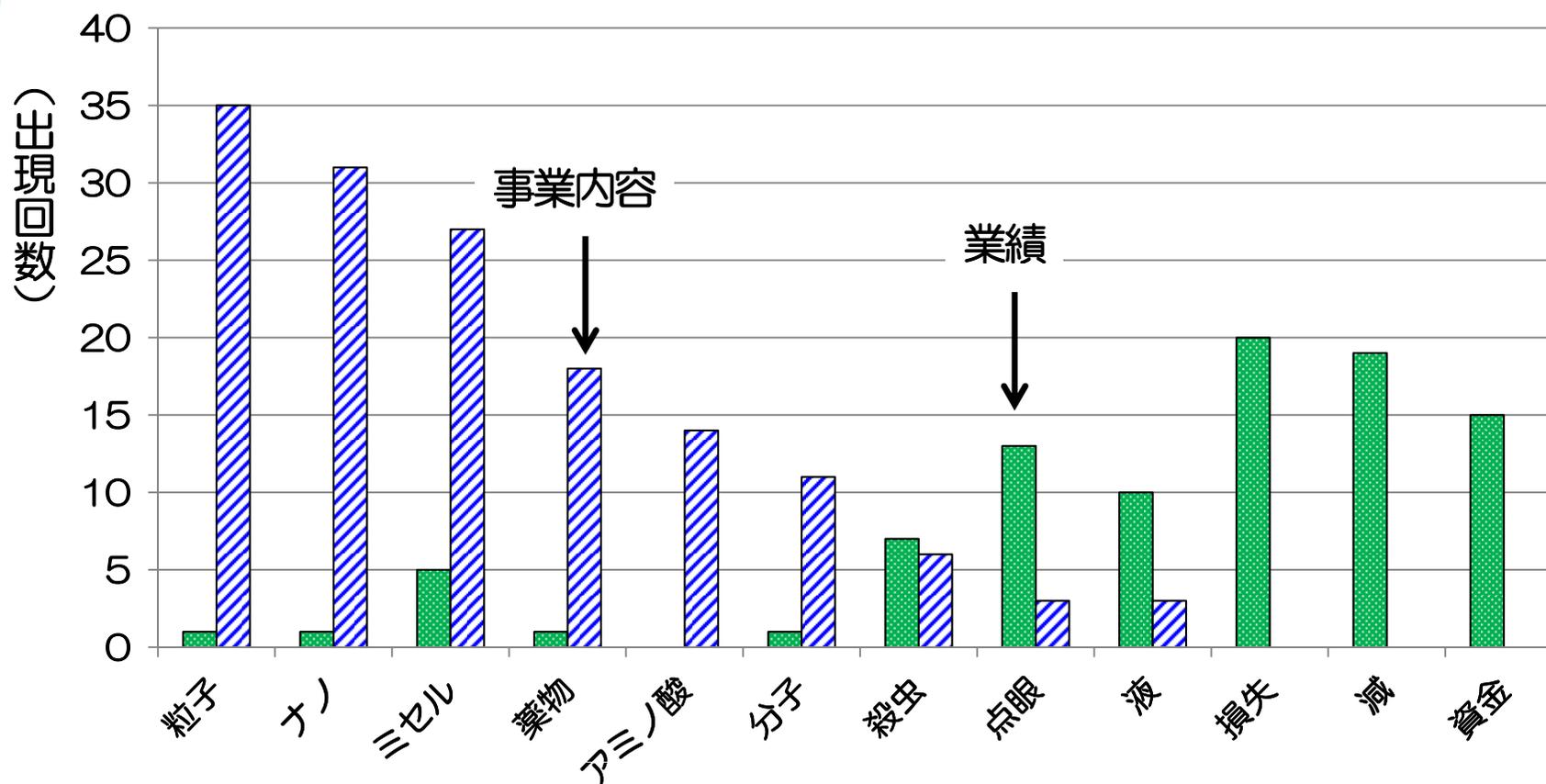
当社グループは、当社，連結子会社6社，関連会社2社及びその他の関係会社1社で構成され，殺虫剤，園芸用品，防疫用剤，家庭用品の製造販売を主な事業の内容とし，その製品はあらゆる種類にわたっております。（以降，省略）

【業績等の例示】

当連結会計年度（平成23年4月1日から平成24年3月31日まで）におけるわが国の経済は，東日本大震災の影響からは回復しつつあるものの，長期化する円高，タイ洪水の影響に加え，信用不安を抱える欧州の経済減速の影響もあり，依然として景気の先行きは不透明な状況になっております。（以降，省略）

判定： 産業別の特性を活用

営業利益がマイナスの企業のみにも出現する単語



検証の結果

1. 産業別各種コーパスの作成により，的確な分類が出来た。
2. 産業別事業内容と業績により，産業ごとに特性コーパスの作成が出来た。
3. 産業別欠損法人割合などの特性を活かし，定性データの適用範囲の絞り込みが出来た。

その結果，営業利益の有無による層化が出来た。

検証のまとめ

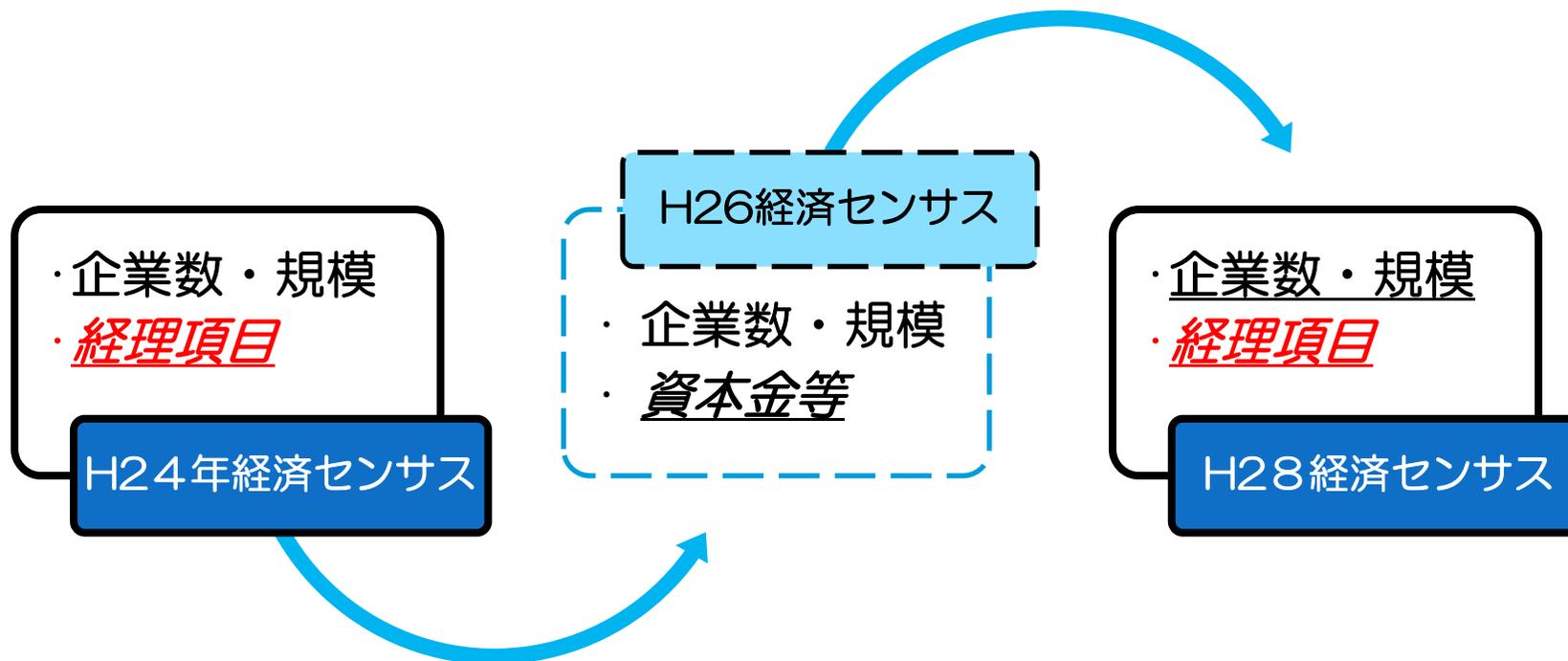
1. 形態素解析によるコーパスの作成
 - 目的に合致したコーパスの作成が重要
 - ✓ 企業構造の層化には産業小分類が必須
2. 産業別出現単語の集約
 - 産業別特性の活用
 - ✓ 検索言語の最適化
3. 「営業利益の有無」別の出現単語の集約
 - 産業別欠損法人割合の活用
 - ✓ プラス, またはマイナスのどちらか一方を検証

今後の課題

1. 作成した産業コーパス等の更新
 - ▶ 産業別の景気変動
 - ✓ 利用目的に適した登録内容の更新
2. 定量・定性データの併用
 - ▶ 文字情報での分類の限界
 - ✓ 数値データを補足するための活用
3. 全産業・企業規模別ベンチマークの作成
 - ▶ 経済センサス-活動調査データの利用
 - ✓ 産業・企業規模別類似企業の分類

「経済センサス」への期待： 国内の事業所・企業の経理状況を網羅

- ▶ H24経済センサスとH28経済センサスの「経理項目の産業分類・規模別集計結果」による**企業構造変化**を分析
 - ✓国内のすべての事業所・企業の経理項目を利用することによる企業構造の明確化



参考文献・参考資料

【参考文献】

1. Shirakawa, Kiyomi. (2013). “A Post-Aggregation Error Record Extraction Based on Naive Bayes for Statistical Survey Enumeration” *The 59th World Statistics Congress*, Hong Kong, China, 25-30 August 2013.
2. Ito, Takayuki, Yutaka Abe, and Tatsuo Noro. (2013). “The Best Stratification to Impute Missing Values of Turnover in Economic Surveys,” *The 59th World Statistics Congress*, Hong Kong, China, 25-30 August 2013.
3. 伊藤孝之, 白川清美. (2013). 「経理項目に基づく企業の構造化分析-構造の境界値検定-」, 2013年度統計関連学会連合大会講演報告集, 大阪大学豊中キャンパス.
4. 豊田秀樹『データマイニング入門』東京図書 (2008)
5. 石田基広『Rによるテキストマイニング入門』森北出版 (2008)
6. Christopher D.Manning『情報検索の基礎』共立出版 (2012)
7. 原田博実『財務諸表“寝かせ読み”速読法』アスキー新書(2010)
8. 森田松太郎『ビジネスゼミナール経営分析入門』日本経済新聞出版社 (2009)
9. 桜井久勝『財務分析諸表分析【第2版】』 (2003)

【参考資料】

1. 会社標本調査結果（税務統計から見た法人企業の実態）, 国税庁
<http://www.nta.go.jp/kohyo/tokei/kokuzeicho/kaishahyohon2011/pdf/kekka.pdf>
2. 産業分類「平成24年経済センサス - 活動調査産業分類一覧」総務省統計局
<http://www.stat.go.jp/index/seido/sangyo/pdf/19san3e>
3. EDINET (Electronic Disclosure for Investors' NETwork)
<http://info.edinet-fsa.go.jp/>

ご清聴ありがとうございました。