様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　３０　年　　９　月　２８　日現在

機関番号：　３２４２１
研究種目：　基盤研究(B)（一般）
研究期間：　2015～2017
課題番号：　１５Ｈ０３３８３

研究課題名（和文）知識創出ツールに基づくことにより実利用可能となるテキストマイニング手法の開発

研究課題名（英文）Study on Practically Available Text Minning Technique Based on Kowledge Creation Tool

研究代表者

菰田　文男（FUMIO, KOMODA）

埼玉学園大学・経済経営学部・教授

研究者番号：６０１１６７２０

交付決定額（研究期間全体）：（直接経費）　　5,300,000円

研究成果の概要（和文）：　本研究の目的は全体の鳥瞰図を描くのではなく、企業の意思決定に利用可能な信頼性のある知識を持つために、ピンポイントフォーカス型テキストマイニング手法の確立を目指すことである。
　そのためにテキストを５文ごとに切り分けた「テキストブロック」を作成したうえで、（１）テキストブロックを単位とした語の共起関係を解析し、（２）テキストブロックから構文関係を有する基本句を作成し、（３）文書の類似性を基準としてテキストブロック間の結びつけをおこなうことによって、精度の高い知識の獲得が可能となることを論じた。

研究成果の概要（英文）：　In this study, pinpoint focus type of text mining technique in dispensable for decision making by private companies is obtained, in replace for its bird’s eye watching type, which has been studied for a long time. The former method enables acquiring detailed knowledge needed for companies.
　In order to do so, text blocs composed of five sentences are produced from total text data, followed by (1) analyzing co-occurrence relation between words inside each text block, (2) producing “basic phrases” with syntactic relation , (3) integrating text blocks based on the similarity.

研究分野：　技術経営

キーワード：　テキストマイニング　データマイニング　知の構造化　BOPビジネス　国際化　日本企業

１．研究開始当初の背景

近年急速にビッグデータとその解析の意義が注目させるようになっているが、ビッグデータに含まれる多様なデータの中でも最も重要な意味を含んでいるのは、テキストデータと、そのマイニングである。データベースに含まれるメタデータやリレーショナルデータベースに比して、構造化されていないテキストデータの中には遙かに豊かな知識が含まれている。しかし、構造化されていない自然言語の中から知識・意味を獲得することは難しい。

日本におけるテキストマイニング市場の 3/4 を日本企業が占めている現実が示すように、膨大な量のテキストデータから目的の部分を抽出し、価値ある重要な知識・意味を発見する研究は、日本が世界の最先端の水準を維持している。その理由は英語などに比して形態素の抽出が難しいという日本語固有の性質が、その克服のための研究を必要としていることに由来している。このような現実を背景とし、さらには人工知能研究の積み重ねに支えられることによって、多くの研究者が優れた学術研究面での成果をあげてきた。この学術研究の積み重ねの結果、企業の戦略決定、医療・介護の実践の場など、さまざまな社会的な実践の場で応用され、その意義が少しずつ認められ始めており、多くの優れた成果が刊行されるようになっている。

しかし、企業にとって最重要の意思決定(選択と集中、営業戦略、技術開発戦略など)において利用されるに足るだけの精度の高い知識を生み出すほどには、未だテキストマイニングは信頼を獲得するには至っていないというのも否定しがたい事実である。テキストマイニング研究の意義の将来的な潜在的可能性は疑いないとしても、現時点では実践の場で十分に利用されていない。その理由は、現在のテキストマイニング研究がテキストデータから単語と単語との共起関係情報を多変量解析やネットワーク分析などの統計解析手法を適用することによって、知識を発見することに過度に依存しているからであり、その結果として少量のデータを実験的に統計学の手法を適用して解析するという「手段」「手法」の開発のための研究が優先されており、得られた「結果」が実践の場で利用可能であるかどうかにかんする研究は重視されなかったからである。言い換えれば統計学的手法の適用の前提となる、データの収集、データの前処理、データの解釈等の研究が軽視されてきたからである。

また、解析手法についても、従来のテキストマイニングにおいてはテキストデータ全体の特徴を大きな視点から捉えることに主眼を置く、いわば「鳥瞰図描画型テキストマイニング手法」の洗練に力点が置かれたために、テキストの一部にフォーカスした精度の高い知識を得ることができなかったと言える。企業などの実践の場での実利用が可能なだけの精度の高い知識は、テキストデータ全体の中から、必要な箇所を適切に抽出し、その部分に焦点をあてたマイニング、いわば「ピンポイントフォーカス型テキストマイニング手法」が必要である。

このような現状で、新たなマイニング手法の開発は、今日のテキストマイニング研究の焦眉の課題となっている。このような現実が本研究の背景にある。

２．研究の目的

上述から理解されるように、本研究の目的はテキストマイニング研究が社会において利用可能な水準にまで、すなわち例えば企業が自社の将来の成長を決定づける意思決定を委ねるに足るだけの信頼性のある知識・意味を獲得する手法を工夫し、提示することにある。

そのために、

(1)優れたテキストデータの収集
(2)そのデータの解析に適する形式への成形・構造化(＝付加価値の付与するための前処理)
(3) 人間が持つ知識による解析結果の解釈及びそれにもとづくマイニングの適切な繰り返し

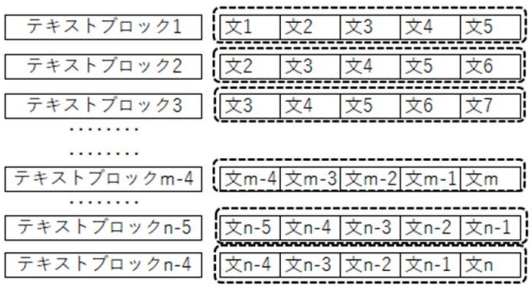
などを重視したマイニング手法の工夫が必要である。言い換えれば、従来のテキストマイニング研究の多くがアルゴリズムの研究に向かっているのに対し、本研究の独自性は「データの収集」「成形・構造化」「統計解析」「結果の解釈」の繰り返しによって精度の高い知識を得るための手法を提示するという点にあり、これによって企業の意思決定という実利用の場で利用できるだけの精度の高い知識・意味を抽出できる解析手法(ピンポイントフォーカス型テキストマイニング手法)の確立を目指す点にある

精度の高い解析手法を可能とする手法を獲得するためには、特許公報ユーや技術論文の中から、求める「解」が叙述されている箇所を適切にピンポイントで抽出することが必要である。そのために本研究では、テキストを 5 文単位のテキストブロック(図 1)に分割するという手法を導入する。このようにデータを成形・加工(前処理)することによって、真に必要な箇所をピンポイントで抽出することを可能にするだけでなく、単語と単語の意味のある共起関係の取りこぼしを可能な限り少なくし、また意味の無い共起関係というノイズを可能な限り削減することによって、知識・意味の精度を高めることも可能になる。

さらに、語と語の共起関係やテキストブロック間の共起関係から知識・意味を獲得するために、急速に注目されているさまざまな機械学習・深層学習を用いたツールが生まれているので、これを導入することによって精度

の向上を目指す。

図１　テキストブロック



このような研究のためには、新たな手法を実験・検証・評価するための分野を適切に定めて進めることが必要である。適切な分野とは、データを大量に収集することが可能であるとともに、本研究が適用したマイニング手法から得られる知識・意味の精度を検証・評価することが可能な分野でなければならないという意味を含んでいる。そのために本研究では、世界的に注目されていながら、期待に反して十分には発展していない BOP(base of the pyramid)ビジネスを選択する。その理由は後述である。

３．研究の方法

本研究は、BOP ビジネスを事例として、精度の高いマイニング手法を獲得することを目的としているので、テキストマイニング手法をマイニング手法にかんする研究と、BOP ビジネスの現状についての研究という 2 つの柱から構成され、それぞれの専門家の共同研究として進めることが必要になる。しかも、研究代表者・共同研究者全員の密度の濃いコミュニケーションが不可欠であるので、過去の長い共同研究の実績のある研究者で構成されることが求められる。したがって研究体制は以下のように組織される。

まず、研究代表者(菰田)、共同研究者(林、中山)は、2008 年から 3 年間に及ぶテキストマイニングにかんする共同研究をおこなった。この研究では、企業現場でのテキストマイニング利用を重視する菰田・林が統計学の専門家である中山と議論することによって、実践的利用においてきわめて重要なデータの時系列解析手法としての個人差多次元尺度構成法等を提案する等の成果をあげた(共著『特許情報のテキストマイニング』2011 年など)。本研究における新たなテキストマイニング手法の獲得は、従来から継続してきた密度の濃いコミュニケーションに支えられた研究体制を継承し、これまでの鳥瞰図描画型の手法を踏まえた上で、ピンポイントフォーカス型テキストマイニングという新たな手法の開発を目指すという方向性で進められる。

また林はテキストマイニングを東南アジアにおける BOP ビジネス戦略の立案に利用するための現地調査研究を共同研究者(井口、荒井、中山)とともにおこない、ＪＥＴＲＯや東南アジア諸国の NPO(Philippine Business for Social Progress など)の協力者を得て、BOP ビジネスの現状と将来動向を評価するための知識基盤を獲得し、テキストマイニングの利用のための基礎を構築してきた。

本研究は、これらの研究を統合するという形で推進されるが、言うまでも前者が重要な位置にある。

本研究の出発点は、データの収集である。BOP 関連のデータとして、新聞記事、業界レポートなどがあるが、とりわけ貴重な価値を持つのは世界銀行のスタッフが同銀行の公式サイトに公開しているブログである。したがって、同ブログの中から BOP ビジネスに関連する記事を中心として収集し、さらに世界の業界レポートや JETRO が刊行している多くの報告書の中からのＢＯＰにかんするレポートや報告書を分析対象データとして収集した。

次のこのテキストデータの中から、(1)単語の出現頻度の時系列分析、(2)単語の共起関係の分析などによって全体を鳥瞰するマイニングをおこない、期待に反して低迷している BOP ビジネスの成長に必要な施策･要件などを解明する。

次に、テキストデータを 5 文単位のテキストブロックに加工する。そしてこの場合の単語の共起関係の分析から得られる結果を、加工しないテキストデータの分析結果と比較することによって、どちらが現状を正しく表現しているかを検証・評価する。

その際に、多次元尺度法やクラスター分析などの手法だけでなく、人工知能研究の成果を自然言語処理に適用したツールとしての word2vec を適用し、従来の手法から得られる分析結果と比較するなどの試行錯誤をおこなう。

さらに以上のような鳥瞰図描画型テキストマイニングにより、全体の本質を捉えた上で、とくに重要なテーマを抽出し、その部分にフォーカスした深い分析をおこなう。そのためには求める知識・意味が叙述されているテキストブロックをできる限り正しく、また取りこぼしなく抽出することが必要であり、そのためにさまざまな試みをおこなう。とくに文書間の文脈の類似性を抽出することを目的として word2vec を発展させた doc2vec を利用することの意義を検証することなどを重視する。

言うまでなく、テキストブロックを可能な限り適切に抽出することができるかどうか、さらにまた関連するテキストブロックを可能な限り適切に結びつけることができるかどうかが、必要な箇所にのみ焦点を当てた精緻で豊かな知識・意味を獲得できるかどうかを左右することになるので、そのための試行錯誤をおこない、最も望ましい手法を発見・

獲得する。
もちろんさまざまな手法の中で、どの手法を用いることが最も精度の高い知識・意味を発見できるかどうかの検証・評価は、BOP ビジネスについて正しい知識を有していることが前提である。そのために、世界の BOP ビジネスの現状と将来動向を調査・研究する。とりわけこのビジネスが十分に成長しない理由、今後のその成長に必要な条件や施策について解明する。そのために、フィリピン、ベトナムなどで同ビジネスに携わっている NGO や大学の研究者を訪問し、意見交換する。

４．研究成果

本研究から、以下のような多くの成果や知見が得られた。
第一に、狭いテーマに焦点を当てて深い知識を得るためには、まず全体を大づかみに捉える鳥瞰図描画型の手法によって全体の傾向や特徴を可能な限り的確に発見し、それに基づいて少しずつ求める「解」の獲得に向けて焦点を絞ってゆくことが、ピンポイントフォーカス型テキストマイニング手法の基本であり、このような手順によって信頼性の高い精緻な知識が得られることが分かった。
第二に、このように特定のテーマに焦点を当てるために、5 文単位のテキストブロックを作成することが有効であることが分かった。サイズの大きい文書の中から、関係の無い部分を削除して、必要な箇所のみを抽出するためにはテキストブロックの作成が不可欠である。
第三に、このようにテキストブロックを作成することによって、単語の共起関係を適切に抽出できることも分かった。すなわち、サイズの大きい一つの文書を単位として共起関係を抽出すれば膨大な量のノイズ(＝意味の無い共起関係)を意味ある関係とみなしてしまうという問題があり、逆に一つの文(sentence)を単位として共起関係を抽出すれば意味のある共起関係を見逃してしまうという問題がある。しかし 5 文単位のテキストブロックを単位とすれば、かなり精度の高い共起関係を抽出できる。
第四に、関連性のあるテキストブロックを結びつけることによって、さらに一層精度が高く精緻で豊かな知識・意味を発見できる。このテキストブロックのを結びつける上で、クラスター分析や深層学習を導入したツールが有効であることも分かった。
また、当初の研究計画の中では想定されていなかった知見も、多く得られた。その中でも最も重要な知見は、共起関係を「単語」を単位としておこなうのではなく、「句」を単位としておこなうことである。単語を文全体の中から切り離して捉えるのではなく、構文関係の中で句として捉えることによって、その意味を厳密にすることができる。したがって「句」を基準としてマイニングすることによって、ピンポイントフォーカス型テキストマイニング手法の意義をさらに高めることが可能になる。

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 16 件）

2017 年

茄田文男,正宗賢,那須川哲哉,大津良司,村垣善治，テキストマイニングに基づく脳神経外科手術ロボット研究の動向分析，埼玉学園大学紀要，17，pp.41-51，査読無．

茄田文男，テキストマイニングに基づく日本企業国際化の研究，政策科学学会年報，7，pp.13-29，査読有．

茄田文男・中山厚穂，ピンポイントフォーカス型テキストマイニング手法の研究，電子情報通信学会技術研究報告，117(207)，pp.1-4，査読無．

Iguchi C., Factors affecting the competitiveness of emerging MNEs: the case of firms in Thailand，明治大学経営論集，65，pp. 59-82，査読無．

林倬史，バリューチェーンとビジネスプロセスアウトソーシング(BPO)の展開過程と国際化:フィリピン BPO 産業との関連から，経営論叢，7(2)，pp.161-189，査読無．

林倬史,鎌田桂輔，新興国の所得の増加とクズネッツの逆Ｕ字仮説の再検討，経営論叢，7(1)，pp.97-127，査読無．

林倬史，アジアの新たな発展経路とルイスの転換点，経営論叢，7(1)，pp.97-126，査読無．

林倬史，BOP 戦略としての BPO 戦略: フィリピン IT-BPO 産業の位置づけを中心として，経営研究所紀要，47，pp.49-69，査読無．

2016 年

林倬史・茄田文男・中山厚穂，テキストデータの分析による BOP ビジネス動向の研究，経営研究所年報(国士舘大学)，46，pp.1-50，査読無．

林倬史，新興国市場と BOP 戦略論の新たな展開: 開発経営学を目指して，経済論叢，6(1)，pp.55-86，査読無．

井口知栄，海外の国際ビジネス関連学会とジャーナルとの関わり方，国際ビジネス研究，8(2)，pp.179-185，査読無．

Sekiguchi T., Froese F., Iguchi C., International human resource management of japanese multinational corporations; challenge and future directions, *Asian Bisiness and Management,* 15(5)，pp.1-27，査読無．

2015 年

井口知栄，日系多国籍企業のグローバル R&D: 在ヨーロッパ多国籍子会社の R&D 拠

点の役割を中心に，三田商学研究，57(2)，pp.141-153 査読有．

林倬史，新興国の台頭とリバース・イノベーションの分析視角，経営研究所紀要，45，pp.1-30，査読無．

林倬史，新興国の BOP と NGO の自律的ビジネス生態系戦略：フィリピンの事例を中心として，アジア経営研究，21，pp.71-82，査読無．

中山厚穂，マーケティングにおけるテキストデータ活用の可能性と限界，マーケティングジャーナル，138，pp.38-54，査読無．

〔学会発表〕（計 18 件）

2017 年

菰田文男・中山厚穂，ピンポイントフォーカス型テキストマイニング手法の研究，電子情報通信学会言語理解とコミュニケーション研究会，2017.09.07，成蹊大学．

菰田文男，日本企業の BOP(Base of the Pyramid)ビジネス推進のために：テキストデータの分析にもとづいて，2017.09.24，埼玉大学．

Nakayama A., Komoda F., Study on globalization of japanese companies based on text mining, *6th German-Japanese Workshop on Advances in Data Analysis and Related New Techniques and Applications,* 2017.08.11, Tama University.

Iguchi C. 多国籍企業の研究開発・技術開発拠点のリロケーション：ホストアジア諸国の事例を中心として，産総研触媒化融合研究センター第 51 回講演会，2017.07.28，産業技術総合研究所．

林倬史，BOP 戦略としての BPO 戦略：フィリピン IT-BPO 産業の位置づけを中心として，アジア経営学会 2017.09.09，東北大学．

2016 年

林倬史，新興国市場の特質と新たな BOP 戦略論，日本経営学会関東部会，2016.11.26，日本大学．

林倬史，アジアにおける BOP 戦略と経営戦略論の再検討，アジア経営学会全国大会，2016.09.06，九州産業大学．

Hayashi T., Iguchi C., Arai M., Base-of-the pyramid business strategies to tackle poverty in emerging countries: strategic management in economic development, *European International Business Academy,* 2016.12.03, Vienna.

Nakayama A., Analysis of trending topics in consumer web communication data, *Abstracts of german-japanese symposium 2016*, 2016.12.12, Schloss Reisensburg.

中山厚穂，マーケティングでターにおける非対称性の分析：Web 上野マーケティングコミュニケーションデータの分析，日本行動計量学会 44 回全国大会，2016.08.31，札幌学院大学．

中山厚穂，出口慎二，鳥谷雅彦，クラスタ中心を再計算しない大規模データのための非階層的クラスタリング，日本行動計量学会 44 回全国大会，2016.08.31，札幌学院大学．

中山厚穂，マーケティングにおける Web コミュニケーションデータ活用の可能性，統計関連学会連合大会，2016.09.05，金沢大学．

Iguchi C., Hayashi T., Nakayama A., The effects of inter-organizational collaborative R&D on MNE's innovation systems, *European International Business Academy,* 2016.12.03, Vienna.

井口知栄，日系企業の研究開発・技術開発のアジア域内でのリロケーション，アジア経営学会全国大会，2016.09.06，九州産業大学．

2015 年

Iguchi C., From local suppliers to malaysian MNEs: effects of national innovation systems on MNEs, *41st EIBA Annual Conference*, 2015.12.01, RUC Rio de Janeiro

井口知栄，日系多国籍企業のイノベーションシステム，工業経営研究学会 2015.08.29，明治大学．

Iguchi C., Local and global innovation by japanese MNEs, *Asia Academy of Management,* 2015.06.22, Chinese University of Hong Kong.

Nakayama A., Classification of production of topics on social media considering temporal variation, *The 9th Conference of the Asian Regional Section of the IASC,* 2015.12.17, National University of Singapore.

Nakayama A., The classification and visualization trending topics considering time series variation, *2015 Conference of the International Federation*, 2015.07.07, University of Bologna.

〔図書〕（計 6 件）

2017 年

Iguchi C., Hayashi T., Nakayama A., The effect of inter-organizational collaborative R&D on MNE's innovation systems, in Sakamoto T., Shoda S. (ed.), *Global, innovative environmental management*, Maruzen Planet, pp.1-192.

2016 年

林倬史，新興国市場の特質と新たな BOP 戦略論，文真堂，pp.1-207.

Nakayama A., The necessity of a triadic distance model, in Withelm A.F.X., Kestler H.A. (ed.), *Analysis of large*

and complex data, Springer-Verlag, pp.1-656.
中山厚穂，調査に従事する人々のための統計学応用講座　第4版，マーケティング・リサーチ協会，pp.1-133.
井口知栄，グローバル化と多国籍企業，関口倫紀，竹内規彦，井口知栄（編），国際人的資源管理，中央経済社，pp.1-264.
2015年
Iguchi C., (Lambregs B., Beerepoot N., Kloosterman R.C. (eds.)), *The local impact of globalization in south and southeast asia: offshore business processes in service industries*, Routledge, pp.1-238.

６．研究組織

(1)研究代表者
菰田文男　(KOMODA, Fumio)
（埼玉学園大学　経済経営学部　教授）
研究者番号：60116720

(2)研究分担者
井口知栄　(IGUCHI, Chie)
（慶応大学　商学部　准教授）
研究者番号： 20411209
研究分担者
林倬史　(HAYASHI, Takabumi)
（国士舘大学　経営学部　客員教授）
研究者番号：50156444
研究分担者
中山厚穂　(NAKAYAMA, Atsuho)
（首都大学東京　社会科学研究科　准教授）
研究者番号：60434198
研究分担者
荒井将志　(ARAI, Masashi)
（亜細亜大学　国際関係学部　准教授）
研究者番号：70549691