

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 6月 6日現在

機関番号：13904
 研究種目：基盤研究(C)
 研究期間：2010～2012
 課題番号：22500129
 研究課題名(和文) 特許文書からのクロスブートストラップによる表現抽出を用いたパテントマップ自動生成
 研究課題名(英文) Automatic Generation of Patentmaps from Patent Documents by Extraction of Expressions using Bootstrapping
 研究代表者
 増山 繁 (MASUYAMA SHIGERU)
 豊橋技術科学大学・大学院工学研究科・教授
 研究者番号：60173762

研究成果の概要(和文)：特許文書からの技術語・効果語取得はパテントマップの自動生成などに役立つ。そこで、まず、特許文書からのクロスブートストラップによる表現抽出を用いたパテントマップ自動生成法を考案した。次に、手がかり語を用いたブートストラップ法と文法パターンを併用し、これらの語を取得する手法を提案した。特許情報として抽出すべき対象を詳細に分析し、取得する方法を考案した。企業では、製品開発に重要な特許調査で、関連技術分野の特許検索だけでなく、他社の競合特許と他社の競合製品の関係に着目し、製品関連特許の特定、各特徴レベルの関連判定法を考案した。

研究成果の概要(英文)：Technical words and effective words acquisition from the patent document is useful, for example, it can be applied to automatic generation of patent map. Therefore, we first propose an automatic generation method of patentmaps from patent documents by extraction of expressions using cross-bootstrapping. Then we propose a method to a combination of grammar patterns and bootstrap method using the clue word, to obtain these words. We develop a method to analyze in detail the object to be extracted for automatic patent map generation.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	1,300,000	390,000	1,690,000
2011年度	1,000,000	300,000	1,300,000
2012年度	1,000,000	300,000	1,300,000
年度			
年度			
総計	3,300,000	990,000	4,290,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：自然言語処理, テキストマイニング, 特許情報処理

1. 研究開始当初の背景

国内における特許出願は年間40万件近くに達している。特許出願される技術は、出願人である企業にとってきわめて重要なものである場合が多く、その出願動向を調査することは、企業における技術開発戦略、および、知財戦略の策定や国、地方自治体における技術開発推進政策立案に大きく寄与する。図1

のようなパテントマップは、特許出願動向を可視化したものであり、出願動向を容易に把握することができ、上記調査の際に重宝されている。しかしながら、年間の出願件数も多大であり、かつ、特許1件当たりの文書量も膨大であるため、人手によるパテントマップの作成には莫大な作業量がかかるという問題がある。そのため、図1のようなパテント

具体的には、図5に示すように、課題表現と効果表現が獲得されていく。また、クロスブートストラップが適用可能な条件を図6にまとめる。

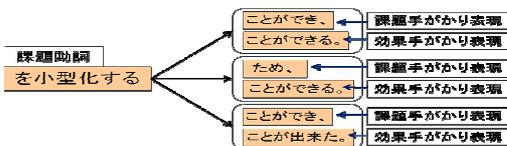


図6 クロスブートストラップが適用できる条件

更に、手がかり語を用いたブートストラップ法と文法パターンを併用し、これらの語を取得する手法を提案した。特許情報として抽出すべき対象を詳細に分析し、取得する方法を考案した。企業では、製品開発に重要な特許調査で、関連技術分野の特許検索だけでなく、他社の競合特許と他社の競合製品の関係に着目し、製品関連特許の特定、各特徴レベルの関連判定法を考案した。

言語の意味的表現獲得の基礎として、以下の研究を行った：

- (1) 自然言語処理における意味処理の基盤となる日本語 Wikipedia のカテゴリ階層に着目した。そして、言語オントロジに対して、上位・下位関係を付加する研究を行った。
- (2) 因果関係を表す5つのパターンがあることを明確にした。文(ないし、連続する2文)を因果関係を表す特有の表現を用いて分類する手法を明確にした。
- (3) 記事の概要文 Wikipedia が持つ記事の分類体系に基づき、多数存在する Infobox テンプレートの種類の中から記事に適した種類を推定する手法を提案した。

更に、ここで研究した語対間の意味的関係の研究に関連して、以下の研究を行った：

- (1) メニュー階層わかりやすさ評価のための語対の連想関係、関連関係自動判定についても成果を得た。心理実験による実験心理値との順位相関により評価した。
- (2) 2次元バブルチャートの見やすい配置を整数計画法を用いて定式化し、被験者に評価してもらう試みを行った。
- (3) 価格.com からトラブル情報を得るために「口コミ掲示板」の情報を利用した製品のかかる動詞に着目する手法を検討した。
- (4) 企業の業績発表記事から前述のブートストラップ法で抽出される業績要因の中で、重要なものを抽出する手法を考案した。具体的には、企業の web サイトからキーワードを抽出し、その際に付与された重みを使用して重要度を付与する

ことで、重要な業績要因の抽出を行う。

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17件)

- ① Hirofumi Nonaka, Akio Kobayashi, Hiroki Sakaji, Yusuke Suzuki, Hiroyuki Sakai, Shigeru Masuyama, Extraction of the Effect and the Technology Terms from a Patent Document, Journal of Japan Industrial Management Association, vol. 63, no. 2E, pp. 105-111, 2012, 査読有
- ② 小林暁雄, 増山繁, 日本語版ウィキペディアのカテゴリ階層に着目した日本語 WordNet 上位下位意味体系の拡張手法, 電子情報通信学会論文誌 D, J95-D, 6, pp. 1356-1368, 2012, 査読有
- ③ 坂地泰紀, 増山繁, 新聞記事からの因果関係を含む文の抽出手法, 電子情報通信学会論文誌 D, J94-D, pp. 1496-1506, 2011, 査読有
- ④ Hirofumi Nonaka, Akio Kobayashi, Hiroki Sakaji, Yusuke Suzuki, Hiroyuki Sakai, Shigeru Masuyama, Extraction of the effect and the technology terms from a patent document, Proc. IEEE 40th International Conference on Computers and Industrial Engineering (CIE40), July, 2010, 査読有

[学会発表] (計33件)

- ① 原田綾花, 太田貴久, 小林暁雄, 増山繁, 野中尋史, 酒井浩之, 特許文書からの発明に関する特徴的技術とその効果の抽出, 言語処理学会第19回年次大会発表論文集, B5-1, 2013年3月5日, 名古屋大学(名古屋市)
- ② 大田仁克, 太田貴久, 酒井浩之, 増山繁, プレスリリースを用いた製品関連特許のIPC推定, 言語処理学会第19回年次大会発表論文集, P4-19 pp. 721 - 724, 2013年3月15日, 名古屋大学(名古屋市)
- ③ 酒井浩之, 増山繁, 製品特徴に基づく製品発表プレスリリースと特許の関連性の判定, 言語処理学会第19回年次大会発表論文集, pp. 725 - 728, 2013年3月15日, 名古屋大学(名古屋市)
- ④ 原圭介, 小林暁雄, 増山繁, Wikipediaにおける記事に適した Infobox テンプレート種類の推定手法, 言語処理学会第19回年次大会発表論文集, P6-1, pp. 818 - 821, 2013年3月15日, 名古屋大学(名古屋市)
- ⑤ 増山繁, 意味の世界に迫るテキストマイニングの拓く未来 -因果関係を例として-, 平成24年度電気関係学会東海支部連

合大会シンポジウム（3）ここまでできる言語処理技術-音声・言語情報処理の最先端-, S3-5, 2012年9月25日, 豊橋技術科学大学（豊橋市）

- ⑥ 高尾美代子, 酒井浩之, 鶴田雅信, 増山繁, 渡邊知早, 梅村祥之, メニュー階層わかりやすさ評価のための語対の連想関係, 関連関係自動判定, 第26回人工知能学会全国大会, 4F1-OS-5-7, 2012年6月15日, 山口県教育会館（山口市）
- ⑦ 太田貴久, 鶴田雅信, 野中尋史, 山本悠二, 酒井浩之, 増山繁, 技術-効果型パテントマップに基づくシーズ知的探索システムの開発, 言語処理学会第18回年次大会発表論文集, pp.875-878, 2012年3月15日, 広島市立大学（広島市）
- ⑧ 坂地泰紀, 増山繁, テキストマイニングによる因果関係抽出, 第54回自動制御連合講演会論文集, 456-459, 2011年11月19日, 豊橋技術科学大学（豊橋市）
- ⑨ 小林 暁雄, 増山繁, ウィキペディアを利用したソースの自動拡張, 第54回自動制御連合講演会論文集, 460-465, 2011年11月19日, 豊橋技術科学大学（豊橋市）
- ⑩ 酒井 浩之, 増山繁, 企業業績発表記事を対象としたテキストマイニング, 第54回自動制御連合講演会論文集, 466-470, 2011年11月19日, 豊橋技術科学大学（豊橋市）
- ⑪ 野中尋史, 酒井浩之, 増山繁, 知財訴訟判例文書からの判例統計情報抽出と知財訴訟分析への応用, 第1回特許情報シンポジウム, pp.29-36, 2010年12月10日, 東京大学本郷キャンパス（東京都）
- ⑫ 増山繁, 野中尋史, 坂地泰紀, 小林暁雄, 鈴木佑輔, 太田貴久, 酒井浩之, 自然言語処理技術を利用した効果-技術型パテントマップの自動生成手法の開発, 第1回特許情報シンポジウム, pp.37-44, 2010年12月10日, 東京大学本郷キャンパス（東京都）
- ⑬ 坂地泰紀, 増山繁, 酒井浩之, 新聞記事中の文が因果関係を含むか否かの判定,

情報処理学会研究報告「自然言語処理」, NL-198, pp.47-50, 2010年7月23日, 東北大学青葉山キャンパス（仙台市）

- ⑭ 野中尋史, 小林暁雄, 坂地泰紀, 鈴木佑輔, 酒井浩之, 増山繁, 効果-技術型パテントマップ作成を目的とした特許明細書中からの効果語及び技術語の自動抽出と意味的統合, 人工知能学会第24回全国大会, 2010年6月11日, 長崎ブリックホール（長崎市）

〔図書〕（計0件）

〔産業財産権〕

○出願状況（計1件）

名称：言語関係判別装置、言語関係判別プログラム、言語関係判別方法

発明者：藤井知早、濱田洋人、増山繁、酒井浩之、鶴田雅信、高尾美代子

権利者：トヨタ自動車株式会社、国立大学法人豊橋技術科学大学

種類：G06F 17/30

番号：TY12-02613

出願年月日：平成24年6月6日

国内外の別：国内

○取得状況（計0件）

〔その他〕

ホームページ等

<http://www.la.cs.tut.ac.jp/>

6. 研究組織

(1) 研究代表者

増山繁 (MASUYAMA SHIGERU)

豊橋技術科学大学大学院工学研究科・教授
研究者番号：60173762

(2) 研究分担者

酒井浩之 (SAKAI HIROYUKI)

成蹊大学・工学部

研究者番号：70402659