

令和元年6月27日現在

機関番号：32629

研究種目：基盤研究(C) (一般)

研究期間：2015～2018

課題番号：15K00315

研究課題名(和文) 決算短信やWEB情報を活用した就職支援のための企業検索システム

研究課題名(英文) Company search system for employment support by using financial statements and web information

研究代表者

酒井 浩之 (Sakai, Hiroyuki)

成蹊大学・理工学部・准教授

研究者番号：70402659

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：本研究では、決算短信や企業WEBサイトから、企業の事業に関連する情報を自動的に抽出し、その情報を検索対象とした企業検索システムを開発することを目的とした。研究成果として、決算短信から業績要因を抽出する手法、因果関係(原因と結果)を抽出する手法、業績予測文を抽出する手法、関連企業を推定する手法等を開発した。また、本手法により抽出された業績要因等を検索対象とした4つの企業検索システムを開発した。開発した企業検索システムは、WEBサイト(<http://www.ci.seikei.ac.jp/sakai/>)にて公開されている。

研究成果の学術的意義や社会的意義

本研究による企業検索システムでは、学生の研究分野と関連のある企業を容易に検索することができ、学生の就職活動支援に活用できる。これにより、自分の研究分野と関連のある企業を見つけ、就職志望とすることで、学生と企業とのミスマッチを防ぐ。加えて、中小企業にとっては就職情報サイトへの掲載料が高額なので登録できないのが実情であるが、本研究により、中小企業への就職希望学生が増加することが期待できる。本研究における研究成果として、合計で4つの企業検索システムを開発し、それらは全てWEBサイトに公開されており、誰でも利用可能である。

研究成果の概要(英文)：In this research, we aimed to automatically extract the information related to the business of companies from financial statements and web sites of the companies, and to develop the company search systems by using the extracted information. As research results, we developed a method to extract causal information, a method to extract causal correlation (cause and result), a method to extract sentences of forecasted business performance from financial statements, and a method to estimate related companies by using the causal information, respectively. We also developed company search systems by using causal information extracted by our methods. The developed company search systems are published on the website (<http://www.ci.seikei.ac.jp/sakai/>).

研究分野：テキストマイニング, 自然言語処理

キーワード：就職活動支援 企業検索 業績要因 関連企業検索 決算短信 テキストマイニング 自然言語処理

1. 研究開始当初の背景

研究開始当初、(そして現在も)就職活動中の学生は一般に大手企業を志向する傾向にあるといわれている。また、近年では、企業側の求める人材と学生側の志望とのミスマッチが問題として挙げられ、就活中にその企業のことをよく理解しないまま就職してしまい、その結果、早期退職を余儀なくされるなどが問題となっている。これは学生にとってはもちろんのこと、企業にとっても人材確保の面で大きな痛手である。そのため、お互い満足できる就職を実現できるような就職活動支援システムが求められている。

就職活動において企業の情報を調べる場合、就職情報サイトを使用することが考えられる。しかし、就職情報サイトは企業が登録した事業内容やキーワードを検索対象にしているため、専門性が高い語は企業が登録した事業内容に含まれていない可能性が高い。また、多くの事業を行っている企業は、全ての事業に関する情報を詳細に登録することは困難である。そのうえ、このような就職情報サイトへの企業情報の掲載料は高額であり、中小企業にとっては大きな負担となる。そのため、人材は欲しいにもかかわらず、就職情報サイトに掲載していない中小企業も数多く存在する。そして、そのような企業を学生も見つけることはできず、結果、大手企業に志望が集中してしまう。

以上の理由により、就職情報サイトに掲載されているような情報(事業内容など)を決算短信や企業WEBサイトから自動的に抽出し、それを検索対象とする企業検索システムの必要性を考えるに至った。

2. 研究の目的

本研究では、決算短信や企業WEBサイトから、企業の事業に関連する情報を自動的に抽出し、その情報を検索対象とした企業検索システムのための手法の考案を目的とする。本研究成果による企業検索システムでは、学生の研究分野と関連のある企業を容易に検索することができ、学生の就職活動支援に活用できる。これにより、自分の研究分野と関連のある企業を見つけ、就職志望とすることで、学生と企業とのミスマッチを防ぐ。加えて、中小企業にとっては就職情報サイトへの掲載料が高額なので登録できないのが実情であるが、本研究成果により、中小企業への就職希望学生が増加することが期待できる。

3. 研究の方法

決算短信や企業WEBサイトの情報を活用した就職支援のための企業検索システムを実現するために、(1)企業WEBページ、決算短信の収集、(2)企業WEBページ、決算短信からの企業の事業に関連する情報の抽出、(3)抽出した企業の事業に関連する情報を検索対象とした検索システムの開発を、それぞれ行う。(1)では、上場企業の決算短信と、中小企業をも含めたより多くの企業WEBページを収集する。(2)では、決算短信から業績要因や、企業WEBページからその企業の事業に関連するキーワードを抽出する。さらに、抽出したキーワードと企業との関連性を自動的に判別する。(3)は、(2)で抽出した情報をもとに企業検索システムを開発し、一般に公開する。

4. 研究成果

(1) 決算短信からの企業の事業に関連する情報の抽出に関する研究として、決算短信から業績要因文を抽出する手法[1]の改良に取り組んだ。既存手法は、精度はよいが再現率が50%程度であり、多くの業績要因文を抽出することができないという問題があった。そこで、学習データの自動生成と深層学習を組み合わせ、精度を落とすことなく再現率を72.4%まで向上させることができた。具体的には、既存手法は高い精度を達成していることに着目し、既存手法による抽出結果をさらに絞り込むことで、より高い精度の業績要因文の集合を作成した。そして、作成された高精度の業績要因文を深層学習の学習データとすることで学習データを自動生成し、その自動生成された学習データを使用して深層学習を行い、業績要因文を抽出した。その結果、226,438 文もの学習データを自動生成することができ、既存手法では抽出されなかった業績要因文をも抽出することができた。本研究の成果を、企業名を入力するとその企業の決算短信から抽出された業績要因文を検索することができる企業検索システム(図1. <http://hawk.ci.seikei.ac.jp/cees/> にて公開)に組み込み、より多くの企業の業績要因文が検索できるようになった。本研究成果は査読付き学術雑誌論文に採録された[2]。

(2) 決算短信からの企業の事業に関連する情報の抽出に関する研究として、決算短信から因果関係(原因と結果)を抽出する研究を行った。その成果を用いて例えば原因として「猛暑」、結果として「エアコン」を入力すると、原因が「猛暑」、結果が「エアコン」を含む因果関係を記述している決算短信とその企業を検索することができる企業検索システムを開発し、WEBサイトに公開した(図2. <http://hawk.ci.seikei.ac.jp/CS/>)。本研究成果は査読付き学術雑誌論文に採録された[3]。

(3) 企業WEBページから企業の事業に関連する情報の抽出に関する研究として、企業WEBページからその企業にとって重要なキーワードを抽出し、そのキーワードを使用して、ある企業と関連のある企業を抽出する研究を行った。その成果を用いて、企業名を入力すると、その企業と関連のある企業を検索することができる関連企業検索システムを開発し、WEBサイトに公開した(<http://hawk.ci.seikei.ac.jp/LicsPlus/>)。本システムでは、例えば「エプソン」で検索すると、その

関連企業として「ブラザー工業」が検索される。



図1 決算短信から抽出された業績要因を検索対象とした企業検索システム



図2 決算短信から抽出された原因・結果表現を用いた企業検索システム

(4) 業績予測に修正のある決算短信から例えば「音楽分野は音楽制作が好調であることなどにより、分野全体の売上高は 10 月時点の想定を上回る見込みです」のような業績予測文(企業の今後の業績予測を示す文)を抽出する研究を行った。その成果を用いて、企業名を入力すると、その企業の決算短信から抽出された業績予測文を検索することができるように企業検索システム(図1)を拡張した。それにより、業績が期初の予想よりも好調である企業、不振である企業を、その要因も含めて調べることができ、就職活動における企業選択において重要な要素である、業績が好調で成長している企業を調べるための情報源となることが期待できる。本研究成果は査読付き学術雑誌論文に採録された[4]。さらに本研究を発展させ、抽出対象の決算短信の範囲を広げ、期初の決算短信からも抽出できるように手法を改良した。具体的には、抽出された文が業績予測文であるかどうか深層学習を用いて分類し、さらに、分類された文を再び学習データとすることで、より高い分類精度を達成することができた[5]。

(5) 複数の志望企業がある状態から新規の志望企業を探すことを想定し、複数の企業の共通要素をその企業の決算短信における業績要因から推定し、それを使用して新規の志望企業となる可能性のある関連企業を自動的に推定する手法を開発した。そして、その成果を用いて、入力した複数企業における共通要素から、その共通要素と関連のある企業を検索できるシステムをWEBサイトにて公開した(<http://hawk.ci.seikei.ac.jp/RECS/>)。本研究成果は査読付き学術雑誌論文に採録された[6]。

< 引用文献 >

1. 酒井浩之, 西沢裕子, 松並祥吾, 坂地泰紀, “企業の決算短信 PDF からの業績要因の抽出”, 人工知能学会論文誌, vol.30, no.1, pp.172-182, 2015.
2. 酒井浩之, 松下 和暉, 北島 良三, “学習データの自動生成による決算短信からの業績要因文の抽出”, 日本知能情報ファジィ学会誌, vol. 31, no. 2, pp. 653-661, 2019.

3. 坂地泰紀, 酒井浩之, 増山繁, “決算短信 PDF からの原因・結果表現の抽出, 電子情報通信学会論文誌 D”, vol.J98-D, no.5, pp.811-822, 2015.
4. 北森詩織, 酒井浩之, 坂地泰紀, 決算短信 PDF からの業績予測文の抽出, 電子情報通信学会論文誌 D, vol.J100-D, no.2, pp.150-161, 2017.
5. Shiori Kitamori, Hiroyuki Sakai, Hiroki Sakaji, “Extraction of sentences concerning business performance forecast and economic forecast from summaries of financial statements by deep learning”, IEEE Symposium on Computational Intelligence for Financial Engineering & Economics (IEEE CIFE'17), Hawaii, November, 2017.
6. 田中瑞竜, 酒井浩之, 坂地泰紀, 北島良三, 複数企業からの関連企業の抽出と事業内容に基づく分類, 日本知能情報ファジィ学会誌, vol. 31, no. 1, pp. 546-562, 2019.

5. 主な発表論文等

〔雑誌論文〕(計 14 件)

- 酒井浩之, 松下 和暉, 北島 良三, “学習データの自動生成による決算短信からの業績要因文の抽出”, 日本知能情報ファジィ学会誌, vol. 31, no. 2, pp. 653-661, 2019. (査読有り)
- 北島 良三, 酒井 浩之, 上村 龍太郎, “潜在学習による企業トップメッセージと収益性の関係解析”, 日本知能情報ファジィ学会誌, vol. 31, no. 2, pp. 636-644, 2019. (査読有り)
- 田中瑞竜, 酒井浩之, 坂地泰紀, 北島良三, “複数企業からの関連企業の抽出と事業内容に基づく分類”, 日本知能情報ファジィ学会誌, vol. 31, no. 1, pp. 546-562, 2019. (査読有り)
- 坂地泰紀, 和泉潔, 松島裕康, 川瀬和哉, 林寛, “接触履歴を用いた地域景況インデックスの自動生成”, 日本知能情報ファジィ学会誌, vol. 31, no. 2, pp. 626-635, 2019. (査読有り)
- Ito Tomoki, Sakaji Hiroki, Izumi Kiyoshi, Tsubouchi Kota, Yamashita Tatsuo, “GINN: gradient interpretable neural networks for visualizing financial texts”, International Journal of Data Science and Analytics, pp.1-15, 2018. (査読有り)
- 高野海斗, 酒井浩之, 坂地泰紀, 和泉潔, 岡田奈奈, 水内利和, “株主招集通知における議案タイトルとその分類及び開始ページの推定システム”, 自然言語処理, vol.25, no.1, pp.3-32, 2018. (査読有り)
- Zhouhao Wang, Enda Liu, Hiroki Sakaji, Tomoki Ito, Kiyoshi Izumi, Kota Tsubouchi, Tatsuo Yamashita, “Estimation of Cross-Lingual News Similarities Using Text-Mining Methods”, Journal of Risk and Financial Management, Vol.11, No.1, 2018. (査読有り)
- 北森詩織, 酒井浩之, 坂地泰紀, “決算短信 PDF からの業績予測文の抽出”, 電子情報通信学会論文誌 D, vol.J100-D, no.2, pp.150-161, 2017. (査読有り)
- 野中尋史, 河野誠也, 平岡透, 太田貴久, 増山繁, “産業クラスター評価のための特許出願の空間的自己相関を利用した技術集積動向分析方法の基礎的検討”, 日本知財学会誌, Vol.13, No.3, pp.65-78, 2017. (査読有り)
- Ryo Ito, Kiyoshi Izumi, Hiroki Sakaji, Shintaro Suda, “Lexicon Creation for Financial Sentiment Analysis using Network Embedding”, Journal of Mathematical Finance, Vol.7, No.4, PP.896-907, 2017. (査読有り)
- Hirotohi Honma, Yoko Nakajima, Shigeru Masuyama, “An Algorithm for Hinge Vertex Problem on Circular Trapezoid Graphs”, Journal of Information Processing, Vol.25, pp.945-948, 2017. (査読有り)
- Shin-ichi Nakayama, Shigeru Masuyama, “A linear time algorithm for finding a minimum spanning tree with non-terminal set V_{NT} on outer planar graphs”, IEICE Trans. on information and Systems, E100-D, 3, pp.434-444, 2017. (査読有り)
- Hiroki Sakaji, Junya Ishibuchi, Hiroyuki Sakai, “Extraction of Polarity Comments from Nico Nico Douga”, International Journal of Space-Based and Situated Computing, vol.6, no.3, pp.165-172, 2016. (査読有り)
- 坂地泰紀, 酒井浩之, 増山繁, “決算短信 PDF からの原因・結果表現の抽出”, 電子情報通信学会論文誌 D, J98-D, No.5, pp.811-822, 2015. (査読有り)

〔学会発表〕(計 27 件)

- Ryozo Kitajima, Ryotaro Kamimura, Hiroyuki Sakai, “Relationship Between Management Policies and Profitability for Second Section-Listed Manufacturing Companies of the Tokyo Stock Exchange (2016 Results)”, 7th International Conference on Smart Computing and Artificial Intelligence, 2019
- Kaito Takano, Miryu Tanaka, Hiroyuki Sakai, Ryozo Kitajima, Takahisa Ota, Chinatsu Tanabe, Hiroki Sakaji, “Extraction of characteristic terms from patent documents for technical trend analysis”, 7th International Conference on Smart Computing and Artificial Intelligence, 2019.

Hiroki Sakaji, Akio Kobayashi, Masaki Kohana, Yasunao Takano, Kiyoshi Izumi, "Estimation of Tags Using Various Data for Online Videos", The 33-rd International Conference on Advanced Information Networking and Applications, 2019.

高野海斗, 酒井浩之, 北島良三, "有価証券報告書からの事業セグメントごとの業績要因・業績結果文の抽出", 第 21 回 金融情報学研究会, 2018.

Shiori Kitamori, Hiroyuki Sakai, Hiroki Sakaji, "Extraction of sentences concerning business performance forecast and economic forecast from summaries of financial statements by deep learning", IEEE Symposium on Computational Intelligence for Financial Engineering & Economics (IEEE CIFE'17), Hawaii, November, 2017.

Hiroki Sakaji, Risa Murono, Hiroyuki Sakai, Jason Bennett, Kiyoshi Izumi, "Discovery of Rare Causal Knowledge from Financial Statement Summaries", IEEE Symposium on Computational Intelligence for Financial Engineering & Economics (IEEE CIFE'17), Hawaii, November, 2017.

Hiroki Sakaji, Atsuya Miyazaki, Hiroyuki Sakai, Kiyoshi Izumi, "Extracting Laboratory Front Pages from University Websites", The 6th International Workshop on Web Services and Social Media In conjunction with The 20th International Conference on Network-Based Information Systems (NBIS-2017), pp.1117-1125, Toronto, Canada, August, 2017.

Tomoki Ito, Hiroki Sakaji, Kiyoshi Izumi, Kota Tsubouchi, Tatsuo Yamashita, "Development of Sentiment Indicators Using both Unlabeled and Labeled Posts", The 2017 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFE), pp.314-321, Hawaii, USA, November, 2017.

Tomoki Ito, Hiroki Sakaji, Kiyoshi Izumi, Kota Tsubouchi, Tatsuo Yamashita, "Development of an Interpretable Neural Network Model for Creation of Polarity Concept Dictionaries", The 2017 IEEE International Conference on Data Mining Workshops, pp.1122-1131, New Orleans, USA, November, 2017.

Akio Kobayashi, Hiroki Sakaji, Kohana Masaki, "A Method for Extracting Correct Links from Automatic Created Links on Folksonomy", The 6th International Workshop on Web Services and Social Media In conjunction with The 20th International Conference on Network-Based Information Systems (NBIS-2017), pp.1144-1150, Toronto, Canada, August, 2017.

Kohana Masaki, Hiroki Sakaji, Akio Kobayashi, Shusuke Okamoto, "A Topic Trend on P2P Based Social Media", The 6th International Workshop on Web Services and Social Media In conjunction with The 20th International Conference on Network-Based Information Systems (NBIS-2017), pp.1136-1143, Toronto, Canada, August, 2017.

Hokuto Ootake, Hiroki Sakaji, Keiichi Takamaru, Akio Kobayashi, Yuzu Uchida, Yasutomo Kimura, "A Web-Based Visualization System for Interdisciplinary Research Using Japanese Local Political Corpus", The 6th International Workshop on Web Services and Social Media In conjunction with The 20th International Conference on Network-Based Information Systems (NBIS-2017), pp.1076-1085, Toronto, Canada, August, 2017.

Rio Iwabuchi, Yoko Nakajima, Hirotohi Honma, Haruka Aoshima, Akio Kobayashi, Tomoyoshi Akiba, Shigeru Masuyama, "Proposal of the Recommender System Based on User Evaluation and Cosmetic Ingredients", ICAICTA2017, 16-18 Aug. 2017, Denpasar, Indonesia.

Iori Nakaoka, Hayato Fujino, Yousin Park, Yunju Chen, Hirochika Akaoka, Shigeru Masuyama, "A Study on R&D Organization Management of Japanese ICT Companies Using Patent Information", ICAICTA2017, 16-18 Aug. 2017, Denpasar, Indonesia.

田中瑞章, 酒井浩之, 坂地泰紀, "複数の顧客企業からの共通要素と新規関連企業の抽出", 言語処理学会第 23 回年次大会, pp.1192-1195, 2017.

酒井浩之, 松下和暉, "決算短信からの業績要因文の抽出", 第 11 回テキストアナリティクス・シンポジウム, pp.87-91, 2017.

北森詩織, 酒井浩之, 坂地泰紀, "決算短信 PDF からの業績予測文の抽出", 言語処理学会第 22 回年次大会, 2016 年 3 月 8 日 - 3 月 10 日, 東北大学.

Toshitomo Kimura, Keiichi Takamaru, Takuma Tanaka, Akio Kobayashi, Hiroki Sakaji, Yuzu Uchida, Hokuto Ootake, Shigeru Masuyama, "Creating Japanese Political Corpus from Local Assembly Minutes of 47 Prefectures", Proc. of the 12th Workshop on Asian Language resources, Workshop of Coling 2016, pp.78 - 84, 2016, 12.

Hiroki Sakaji, Masaki Kohana, Akio Kobayashi, Hiroyuki Sakai, "Estimation of Tags via Comments on Nico Nico Douga", The 5th International Workshop on Web Services and Social Media In conjunction with the 19th International Conference on NBIS-2016(WSSM-2016), September, 2016.

Yousuke Kawamoto, Yasuki Yasuki Nishiyama, Akio Kobayashi, Shigeru Masuyama, "Investigation and Construction of Dictionary for Analysis of Comments in a Video Sharing Site", ICAICTA2016, p.22, 2016, 8. (マレーシア, ペナン).

- 21 田中瑞竜, 酒井浩之, 坂地泰紀, “関連企業集合からの共通要素と新規関連企業の抽出”, 第9回 テキストマイニング・シンポジウム, pp.19-24, 2016.
- 22 坂地泰紀, 所佳祐, 酒井浩之, “製品名と特許における名称の言い換え知識の抽出”, 2016年度人工知能学会全国大会, 2016.
- 23 Hiroki Sakaji, Junya Ishibuchi, Hiroyuki Sakai, “Extracting Polarity Comments from Nico Nico Douga”, The 4rd International Workshop on Web Services and Social Media In conjunction with the 18th International Conference on NBS-2015(WSSM-2015), Sep. 2-4, Taipei, Taiwan, 2015.
- 24 Akihito Ikeda, Akio Kobayashi, Hiroki Sakaji, Shigeru Masuyama, “Classification of Comments on Nico Nico Douga for Annotation Based on Referred Contents”, 18th International Conference on Network-Based Information Systems, Sep. 2-4, Taipei, Taiwan, 2015.
- 25 酒井浩之, 小林義和, 坂地泰紀, “企業の決算短信PDFから抽出した業績要因への極性付与”, 第15回 金融情報学研究会, 2015年9月26日, 東京大学.
- 26 北森詩織, 酒井浩之, 坂地泰紀, “決算短信PDFからの業績予測を含む文の抽出”, 第15回 金融情報学研究会, 2015年9月26日, 東京大学.
- 27 本間友実子, 酒井浩之, 坂地泰紀, “企業 Web ページを用いた関連企業の抽出”, 第7回 Web インテリジェンスとインタラクション研究会, 2015年11月28日~11月29日, 東京(株式会社リクルートホールディングス本社).

〔図書〕(計 0件)

〔産業財産権〕

出願状況(計 0件)

取得状況(計 0件)

〔その他〕ホームページ等

本研究の成果による企業検索システム(公開済み)

決算短信から業績要因を含む文を自動的に抽出し, その抽出した業績要因を対象にした企業検索システム

URL: <http://hawk.ci.seikei.ac.jp/cees/>

非上場企業も含めた16,461社の企業WEBサイトから重要なキーワードを抽出し, 抽出した重要なキーワードを検索対象とした企業検索システム

URL: <http://hawk.ci.seikei.ac.jp/LicsPlus/>

決算短信から因果関係を抽出し, それを用いて因果関係を検索することができるシステム

URL: <http://hawk.ci.seikei.ac.jp/CS/>

決算短信から抽出した業績要因文を使用し, 入力した複数企業における共通要素を推定して, 共通要素と関連のある企業を自動的に検索できる関連企業検索システム

URL: <http://hawk.ci.seikei.ac.jp/RECS/>

6. 研究組織

(1)研究分担者

研究分担者氏名: 坂地泰紀

ローマ字氏名: Hiroki Sakaji

所属研究機関名: 東京大学

部局名: 工学系研究科 システム創成学専攻

職名: 特任講師

研究者番号(8桁): 70722809

研究分担者氏名: 増山 繁

ローマ字氏名: Shigeru Masuyama

所属研究機関名: 東京理科大学

部局名: 経営学部 経営学科

職名: 教授

研究者番号(8桁): 60173762

科研費による研究は, 研究者の自覚と責任において実施するものです。そのため, 研究の実施や研究成果の公表等については, 国の要請等に基づくものではなく, その研究成果に関する見解や責任は, 研究者個人に帰属されます。