



Title	事業計画策定作業における情報抽出ならびに事業予測に関する研究
Author(s)	鮫島, 正樹
Citation	大阪大学, 2008, 博士論文
Version Type	
URL	<a href="https://hdl.handle.net/11094/49691">https://hdl.handle.net/11094/49691</a>
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、 <a href="https://www.library.osaka-u.ac.jp/thesis/#closed">https://www.library.osaka-u.ac.jp/thesis/#closed</a> 大阪大学の博士論文について

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

## 【9】

氏 名	さめ 鮫 島 正 樹
博士の専攻分野の名称	博 士 (情報科学)
学 位 記 番 号	第 2 2 5 1 3 号
学 位 授 与 年 月 日	平成 20 年 9 月 25 日
学 位 授 与 の 要 件	学位規則第4条第1項該当 情報科学研究科マルチメディア工学専攻
学 位 論 文 名	事業計画策定作業における情報抽出ならびに事業予測に関する研究
論 文 審 査 委 員	(主査) 教授 薦田 憲久 (副査) 教授 藤原 融 教授 西尾章治郎 教授 岸野 文郎 准教授 原 隆浩 准教授 秋吉 政徳

## 論 文 内 容 の 要 旨

本論文は、筆者が2005年から現在まで大阪大学大学院情報科学研究科マルチメディア工学専攻在学中、ならびに2007年から現在まで(株)日立製作所システム開発研究所にて行ってきた、事業計画策定作業における情報抽出ならびに事業予測に関する研究成果をまとめたものである。

事業計画の策定は、環境分析、目標設定、事業計画立案、事業計画評価の4つのフェーズの業務を遂行することで行われる。これらの業務のうち情報技術によって処理可能な業務は、環境分析における情報抽出業務および事業計画評価における事業シナリオ予測業務であり、それぞれ短時間で高い精度の情報抽出、事業シナリオ予測が求められている。環境分析では自社に関する内部環境分析と顧客や競合他社に関する外部環境分析を行う。特に外部環境分析で分析対象となる文書は様々であるため、文書の特徴や抽出すべき情報に合わせた情報抽出方式が必要である。また事業計画の評価においては、将来を予測できない不確実性を含む事業計画を評価する。不確実性の正確な予測に時間をかけられない一方で、不確実性のために評価精度が低いというトレードオフを解消する必要がある。

本研究は上記の観点で実施するものであり、外部環境分析における情報抽出においては、従来取り組みのなされていないプレスリリースからの競合他社に関する情報抽出を扱い、事業計画評価においては、不確実性を定性情報として部分的に含んだ定性・定量融合モデルによ

る事業シナリオ評価を扱う。プレスリリースからの情報抽出においては、(1) プレスリリースからの時系列数値データ抽出方式を提案し、競合他社の動向を効率的に把握することを可能とする。また、(2) プレスリリースからの製品特徴抽出方式を提案し、競合他社の製品特徴を効率的に把握することを可能とする。事業計画評価においては、(3) 定性・定量融合モデルのシミュレーションを用いた事業シナリオ評価方式を提案し、不確実性を含んだ事業計画を正確に評価することを可能とする。

本論文は全5章から構成される。

第1章の序論では、事業計画策定作業における情報抽出ならびに事業予測の重要性、外部環境分析のための情報抽出技術と事業計画評価のためのシミュレーションにおける課題について述べ、関連する従来研究を概観するとともに、本論文の目的と位置づけを明らかにする。

第2章では、プレスリリースからの時系列数値データ抽出方式を提案する。入力したキーワード、禁止ワード、数値データの単語間距離を用いて数値データと時期データの正しいセットの抽出を可能とする。抽出したセットに対し、ユーザが重複データの削除を行い、グラフ化して出力する。ユーザによる重複データの削除回数と、時系列データの再現率、適合率を従来方式を適用した場合と比較し、提案方式の有効性を評価する。

第3章では、プレスリリースからの製品特徴抽出方式を提案する。入力した記述パターンおよびプレスリリースの文章を係り受け関係を含めて比較することで、高い精度で製品特徴を抽出することを可能とする。さらに記述パターンから新たな記述パターンを自動生成する。

再現率、適合率を従来方式を適用した場合と比較し、提案方式の有効性を評価する。

第4章では、定性・定量融合モデルにおける事業シナリオ評価方式を提案する。シナリオの要因と要因間の影響関係を示すモデルにおいて、定性的な関係をもつ要因の挙動を乱数で決定する。影響元の挙動と関係の大小を乱数に反映し、正確な評価値を導出する。定量モデルから導出した評価値と、一部定量化した融合モデルから提案方式で導出した評価値を比較し、正確さを評価する。

第5章では、結論として本研究で得られた成果を要約し、今後の課題を述べる。

#### 論文審査の結果の要旨

事業計画策定作業における、環境分析のための情報抽出業務および事業計画評価のための事業シナリオ予測業務において、短時間で高い精度の情報抽出、事業シナリオ予測を行うことが求められている。しかし、プレスリリースを対象とした時系列数値データおよび製品特徴情報の抽出や、定性・定量情報を含む事業シナリオの評価に関して、十分な取り組みがなされていない。本論文は、これらの課題に対して、事業計画策定作業における情報抽出ならびに事業予測に関する研究成果を纏めたものである。その主要な成果を要約すると次の通りである。

- (1) 特定のキーワードに係る時系列数値データをプレスリリースから抽出するためには、プレスリリースに記載された複数の数値データと時期データの正しい対応付けが必要である。キーワードと数値データ、時期データの間の形態素数を単語間距離とし、単語間距離が最小となる数値データと時期データを対応付ける方式を提案している。本方式を複数企業のプレスリリースに適用し、従来の自然言語処理技術を拡張した方式と比較して、有効性を確認している。
- (2) プレスリリースから製品特徴情報を抽出するためには、修飾語挿入や動詞変化等の様々な表現に対応し、未知の製品特徴キーワードも含めて抽出する必要があり、ユーザ指定の表現と文章比較する方法や、既知のキーワードを参照して抽出する方法では精度良く抽出できない。そこで、名詞と動詞の係り受け関係に着目した記述パターンによる抽出方式と記述パターンを自動生成する方式を提案している。本方式を複数のプレスリリースに適用し、従来の単語出現頻度に着目して抽出する方式と比較して、有効性を確認している。
- (3) 定性・定量情報を含む事業シナリオを評価する方法として、事業要因と因果関係を定量化して扱うシステムダイナミクスと、定量化して扱う定性シミュレーションがあるが、前者は定量化

のための対象分析に膨大な時間を要し、後者は定性的にしか評価できないという問題がある。そこで、事業シナリオを示す定性・定量融合モデルを、モンテカルロシミュレーションを用いて評価する方式を提案している。定性要因と定量要因をランドマークと名づけた境界値を用いて対応づけし、また複数の因果関係を持つときの因果関係間の強弱を寄与度として決定することで、モデルに沿った妥当な因果関係の伝播状況を実現している。本方式を定性・定量融合モデルに適用し、事業シナリオ評価方式の妥当性および有効性を確認している。

以上のように、本論文は事業計画策定作業における情報抽出ならびに事業予測において成果をあげた先駆的研究として、情報科学に寄与するところが大きい。よって、本論文は博士(情報科学)の学位論文として価値あるものと認める。