

経営情報学会 PACIS2018 主催記念特別全国研究発表大会 企業事例セッション

経営情報学会 PACIS2018 主催記念特別全国研究発表大会においても、企業事例セッションを開催しています。ご寄稿を承諾いただいた参加者による内容紹介を掲載いたします。

- 1) ○貝塚 洋（新日鉄住金ソリューションズ）、向 正道（新日鉄住金ソリューションズ）
「画像分類における新しい半教師あり学習深層学習の適応範囲の拡大」
- 2) ○尾崎博子（株式会社 NTT データ数理システム）、岩本圭介（株式会社 NTT データ数理システム）
「テキストマイニング技術とそのビジネス適用への道」

大会プログラム委員長・大会実行委員長

テキストマイニング技術とそのビジネス適用の道

尾崎博子（おざき ひろこ）

岩本圭介（いわもと けいすけ）

株式会社 NTT データ数理システム

1. 企業紹介

（株）NTT データ数理システム（以下当社）は数理学とコンピュータサイエンスの技術を軸に現実世界の問題解決のためのソリューションを提供しています。

AI の技術要素の 1 つであるデータマイニング・テキストマイニング・バイジアンネットワークなどの各種技術のパッケージ製品の開発、コンサルティング、またデータ分析に関する受託分析などに取り組んでいます。

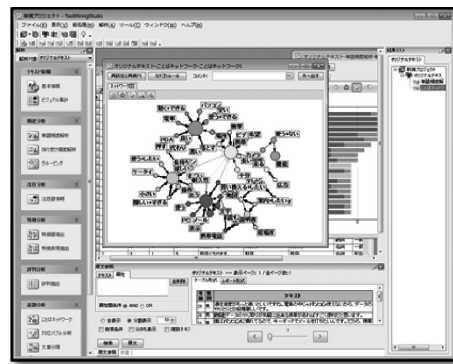


図1 TMS メイン画面

2. 発表概要

本学会の発表では、近年その活用が重要視されているテキストデータの分析に必要な技術であるテキストマイニングと、その分析方法やポイント、ビジネスでの事例 [1] などについて紹介しました。

2.1 テキストマイニングと Text Mining Studio

テキストマイニングとは、文章の情報から有益な知識を発掘することです。売上データやセンサー

データなどの数値データと比べ、テキストデータは定型化されていないことが多いため、数値データと比べ扱い方が難しい場合があります。

当社では、誰にでも高度なテキストマイニングを利用していただくための汎用ツールとして Text Mining Studio（以下 TMS）を開発・販売しています。TMS は単語の頻度集計・属性ごとの特徴的な単語の抽出・共起単語の可視化など、テキストを多角的に分析するための分析機能を数多く持ち、基本

的にマウスのみで操作を行うことができます。

2.2 テキスト分析の目的と手法

TMSを利用するだけでもテキストデータから様々な分析結果を得ることができますが、一般的なテキスト分析の目的としては、

- ✓ 何等かの指針で分類・予測をしたい
- ✓ 傾向把握や集計・要約をしたい
- ✓ 大量データから目的の文書を検索したい

などが考えられます。

分析の指針を決めるヒントとして、当社がこれまで携わってきた案件を元に、テキスト分析の目的別に以下の3タイプに分け、手法や概要を紹介しました。

①分類型

データを何等かの指針により仕分け・分類を行ったり、そのグループを予測したりする場合です。

TMSの「グルーピング」という機能を利用することで、ルールベースの分類を行うことができます。さらに、当社のデータマイニングツール Visual Mining Studio (以下VMS) とTMSを関係することでテキスト情報を数値化し、分類先を予測するモデルを作成する機械学習による分類を行うことができます。

分類型のテキスト分析はコールセンターの対応履歴情報から内容項目の分類を行う場合や、特許や技術論文の仕分けなどを行う場合に利用されます。例えば、コールセンターに問い合わせが多い項目を見つけ、その内容に重点的にFAQを作成することでコール数削減に成功された事例があります。

②サマリー型

データの傾向を把握したり、比較したりする場合です。TMSでは単語や係り受けの頻度集計・共起単語を可視化する「ことばネットワーク」などを利用します。テキストを集約した表やグラフの情報はTMSから得ることができますが、最終的にはその情報を基に分析者が何かしらの考察や解釈を与える必要があります。

サマリー型のテキスト分析はアンケート・日報の内容把握などに利用されます。例えば、ユーザーのアンケートから製品に対する要望・不満点などを探り、新たな製品開発のポイントにつなげることが期

待できます。

③検索型

大量のデータから検索条件を元に、対象となるデータを検索したい場合です。当社ではテキストマイニング技術を用いた類似検索ソリューションを展開しており、検索条件を単語単位だけではなく、文章単位で行うことが可能です。

検索型のテキスト分析は社内文書の検索システムエンジンでの利用などが考えられます。検索条件を元に類似度の高い文章を提示することで、ベテランの経験による文書の探索能力を新人や経験の浅いスタッフに継承することが期待できます。

3. おわりに

本稿ではテキストデータ分析のためのツールと分析手法の概要について紹介しました。

当社ではテキストマイニングだけではなく、本稿でも紹介したデータマイニングを始め、Deep Learning・最適化・シミュレーションのためのツールを開発・販売しています。また、当社の複数製品をシームレスに関係可能な分析基盤である Visual Analytics Platform (以下VAP) を開発しており、VAP上様々な技術を組み合わせて1つの環境で分析する場を提供することができます。

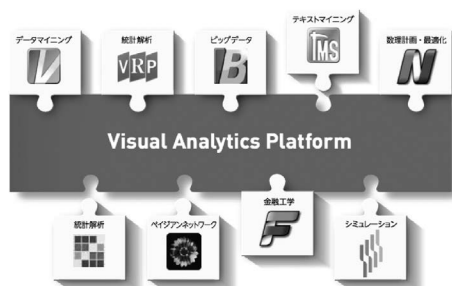


図2 VAP と各関係可能なソフトウェア群

当社ではこれからもデータ分析のためのツールやソリューションの開発を進めていきます。

参考文献

- [1] Text Mining Studio 企業・法人事例, NTT データ数理システムホームページ, <http://www.msi.co.jp/tmstudio/business.html>

略歴

尾崎博子 (おざき ひろこ)

2013年株式会社数理システム(現:株式会社NTTデータ数理システム)入社。データマイニング・テキストマイニングに関わるツール・手法開発及び分析業務に従事。

岩本圭介 (いわもと けいすけ)

1999年株式会社数理システム(現:株式会社NTTデータ数理システム)入社。データマイニング及びディープラーニング関連パッケージ製品の開発リーダーを務める。