

MT 法および最適化手法の適用による特徴分析

大阪大学 *村田 真一 MURATA Shinichi
01604524 大阪大学 森田 浩 MORITA Hiroshi

1. はじめに

ICT 技術のさらなる進展により、IoT・AI・ビッグデータの活用環境がますます身近になり、企業での利活用も進んでいる。一方でIoT・AI・ビッグデータは過度な期待のピークを過ぎ、真価や実際に適用する領域等を見極めるタイミングにきている。

加えて、多くの企業ではIoT・AI・ビッグデータの様々な分析が試行されているが、上手く利活用できないケースが多くある。課題の一つにデータ整備状況が挙げられる。実際の企業ではデータ分析で半ば前提となっている完全なデータがそろわないことが多く、不完全な状態での実施となる場合も多い。本研究ではTV録画データの特徴分析を対象とする。このTV録画データでも特定のデータ群がある空間に属する事は明らかになっているが、その他のデータはどの空間に属するかわからないという状況である。他の課題としてデータが益々複雑化・高度化しており、特徴量も増える一方であるため、該当データにおける深い業務認識がないと分析できない状況にある。TV録画関連データでは多いもので数百の特徴が存在しており、簡単に分析できる状況ではない。また、分析できたとしてもAIによる分析はブラックボックス化されているため、結果に対する説明が困難であり、企業における意思決定には不向きという側面もある。

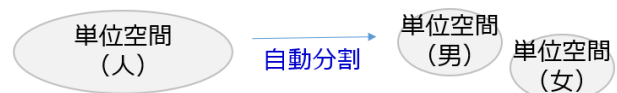
このような課題を抱えているため、実際の企業においてはIoT・AI・ビッグデータをフル活用できていない状況にある。そこで本研究では、これらの問題を解決するために、以下を実現する事を目的とした。

1. 限定されたデータセットでの分析
2. 特徴量選択の自動化
3. 分析結果の見える化

2. MT 法および最適化手法の適用による特徴分析

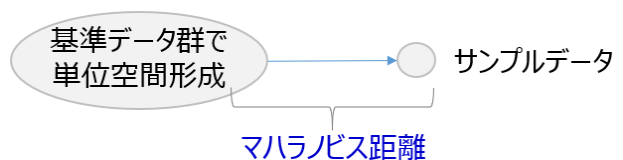
2.1 クラスタリングによる詳細な空間の形成

データが複雑化・高度化することで、我々人間が適切と考える空間でも本当に適切な単位となっていない場合がある。そこで、クラスタリングの手法を利用し、適切な空間へ自動で分割できるようにする。



2.2 MT 法 (マハラノビス距離) の適用

品質工学等の分野で多く利用されている、マハラノビス距離を利用した分析を用いる。MT 法では最初に基準となる 1 つの空間に属するデータ群で単位空間を形成し、検定対象データ(サンプルデータ)との距離でその空間に属するか否かを決定する手法であり、基準データを形成するための 1 空間のデータがあれば分析が可能となる。



2.3 分析に利用する特徴選択の自動化と利用した特徴量の見える化(フィードバック)

データが益々複雑化し特徴量が増大する一方で、企業では各種データに精通した有識者であり、かつデータサイエンティストの素養も具備している人材は中々いないのが実情で、データ分析における有効な特徴量選択が非常に困難である。また、AIによる分析ではモデルがブラックボックスであるため、AIの予測結果がどのような計算過程を経て算出されたのか、どのような特徴量が分析に影響を与えたのかわからず、企業における意思決定の足かせとなっている。これらの問題に対して、最適化手法を利用す

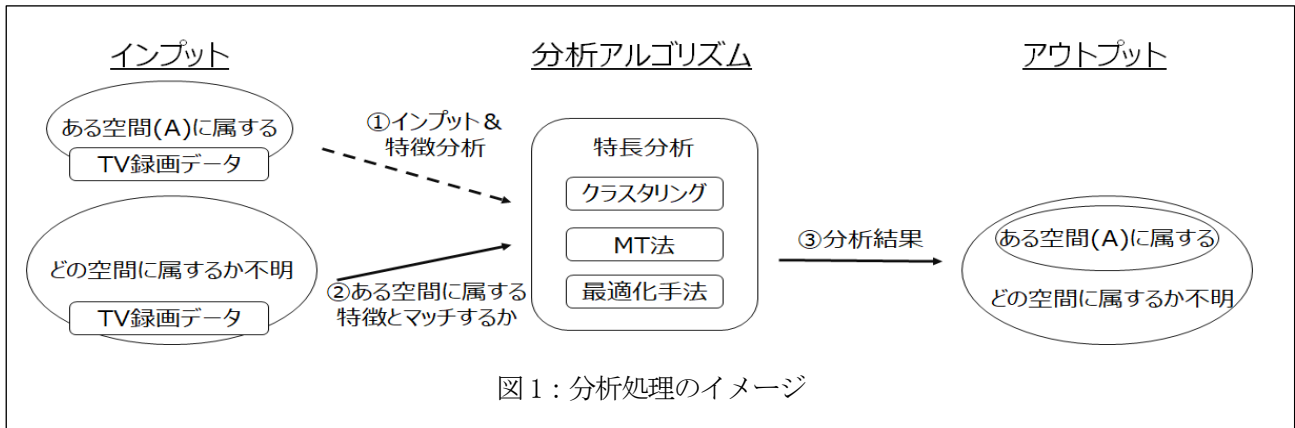


図1：分析処理のイメージ

ることで、分析に必要・有効な特長を自動的に選択するとともに、どの特徴量を使うことで有効な結果が導出されたのかをある程度認識可能とする。

	特徴1	特徴2	特徴3	特徴4	特徴5	特徴6	特徴7	...	特徴M
機器1(個人1)									
機器2(個人2)									
機器3(個人3)									
機器4(個人4)									
機器5(個人5)									
...									
機器N(個人N)									

分析に必要・有効な特長を自動的に選択
→最適化手法を利用

3. TV録画データへの提案手法の適用

本研究では提案概要を検証するために、データ数が多く、特徴が表現されやすいと考えられるTV録画データを用いた実験を実施する。

本実験で利用するTV録画データとは、録画した番組内容、放送日時やチャンネル情報、ジャンル情報等のメタ情報に加えて、視聴有無やどのようなモードで録画したか、どのようなデバイスへ録画したか等の情報で構成されている。

まず、ある空間に属しているユーザの録画データ群をインプットとし、クラスタリング手法を用いて必要に応じ空間を分割する。生成された空間毎に該当録画データの分析に利用する特徴量をランダムに自動的に選択する。そして選択された特徴量を用いてマハラノビス距離を算出して単位空間を生成し、選択した特徴量を用いたモデルの評価を行う。次に最適化手法を用いて特徴量選択、単位空間生成、選択した特徴量を用いたモデル評価のサイクルを繰り返す。これを生成された

空間毎に行う。その後、該当空間に属しているか、属していないか不確定なユーザの録画データをインプットとし、事前に算出された空間毎の最適な特徴量を用いて不確定なユーザの録画データのマハラノビス距離を算出し、該当空間に属しているか否かを判断する。

表1：TV録画データ例(一部抜粋)

項目	補足
録画機器	<ul style="list-style-type: none"> ・本研究では約20~30の特徴量を利用 ・約500~1000件の録画機器に含まれるTV録画データを対象
録画数	
チャンネル情報	
...	

4. おわりに

本研究ではある空間に属するユーザのTV録画データをもとにMT法とクラスタリング手法および最適化手法を適用し単位空間を求めた。その後、該当空間に属するかわからないユーザデータを用いて分析した。今後は色々な最適化手法の適用やTV録画データ以外の様々なデータ種で分析する。

参考文献

- [1] 新村 諭・浜 勉(2018)「MT法と機械学習を用いたプレス加工音による金型摩耗の検知」,『長野県工業技術総合センター研究報告』, 13, pp. 134-137.
- [2] 救仁郷 誠(2001)「解説 マハラノビスの距離 入門」,『品質工学』, 9(1), pp. 13-21.