トップエスイー修了制作

知見を用いたデータ分析業務への データ中心分析法の適用評価

日本ユニシス株式会社　斉藤　功樹　koki.saito@unisys.co.jp

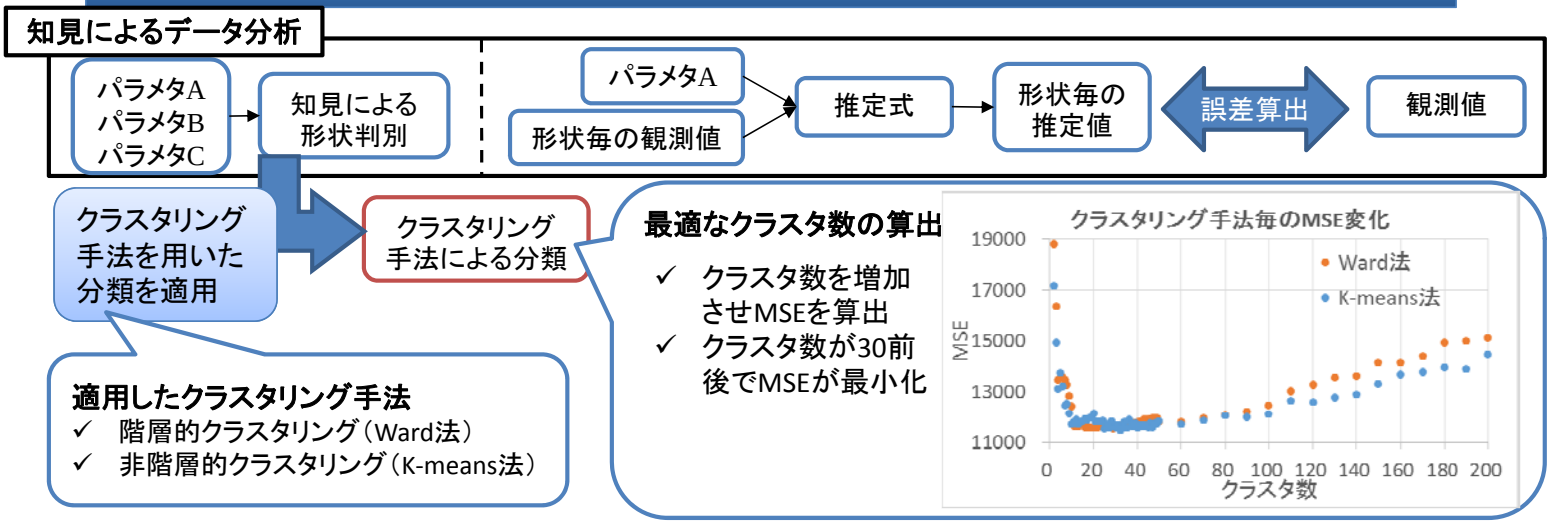
開発における問題点

論文等の知見を用いたデータ分析では単純なルール・式にて表現されるため，複雑さを補えず分析精度の限界がある．知見を活かしたまま精度をあげるためにはパラメタを追加が必要となる．しかし，パラメタの追加にはデータの入手コスト・計算時間のコスト・オーバーフィッティングの可能性が増加する．

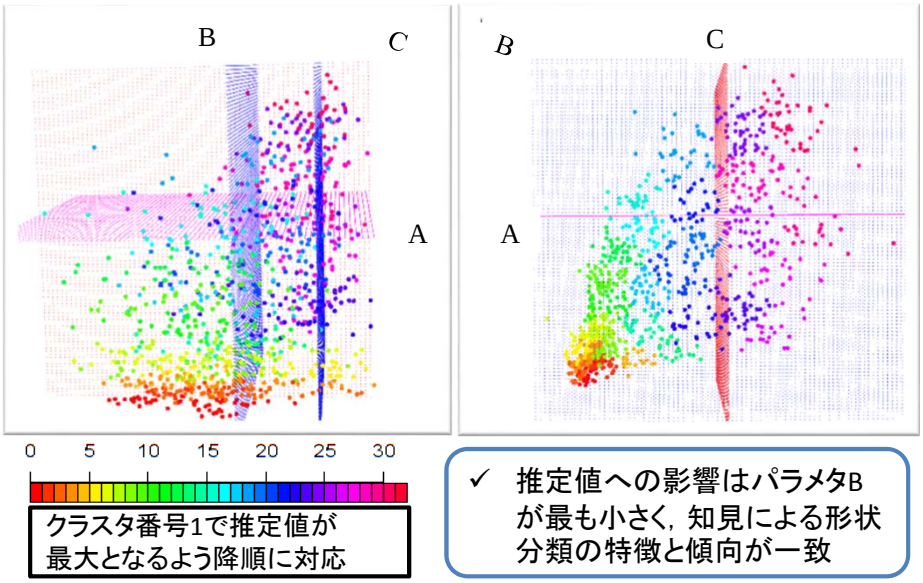
手法・ツールの適用による解決

パラメタの追加ではなく，データ中心とした分析手法を適用する．対象業務では推定値の算出するためにパラメタを利用した形状分類を行っているため，似ているデータ群を分類するクラスタリング手法を適用し，推定精度の向上を目指す．観測値と推定値のMSE（平均二乗誤差）を求め，推定誤差の評価を行う．

クラスタリング手法の適用



クラスタリング結果（K-means法）



✓ 推定値への影響はパラメタBが最も小さく，知見による形状分類の特徴と傾向が一致

評価検証

	知見	クラスタリング手法	
		Ward法	K-means法
クラス数	12	29	32
MSE	13322	11536	11452
RMSE	115	107	107

結果

✓ クラス数は30前後でMSEが最小化

✓ クラスタリング手法の方がMSEが約14%減少

クラスタリング手法の方が特徴を精緻に表現でき，推定精度が向上した