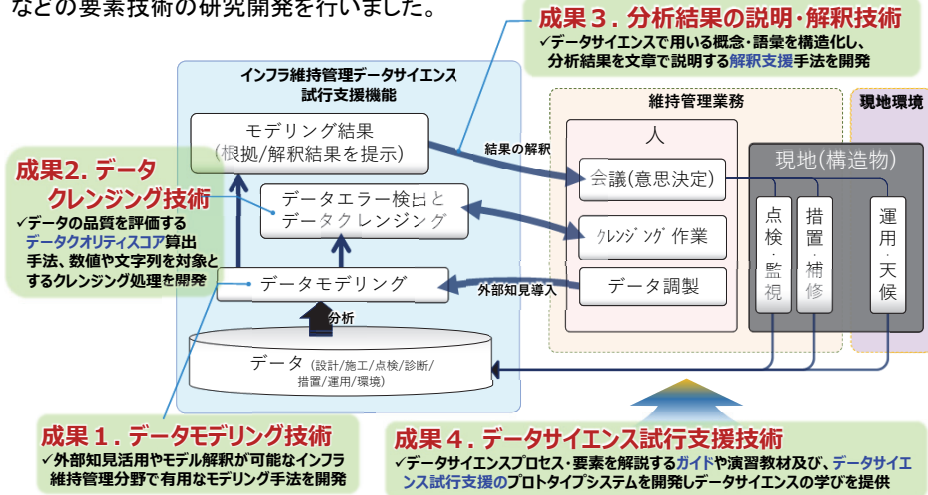


インフラ維持管理データサイエンスの高度化と体系化に関する研究開発

湧田 雄基¹ ¹北海道大学 数理・データサイエンス教育研究センター

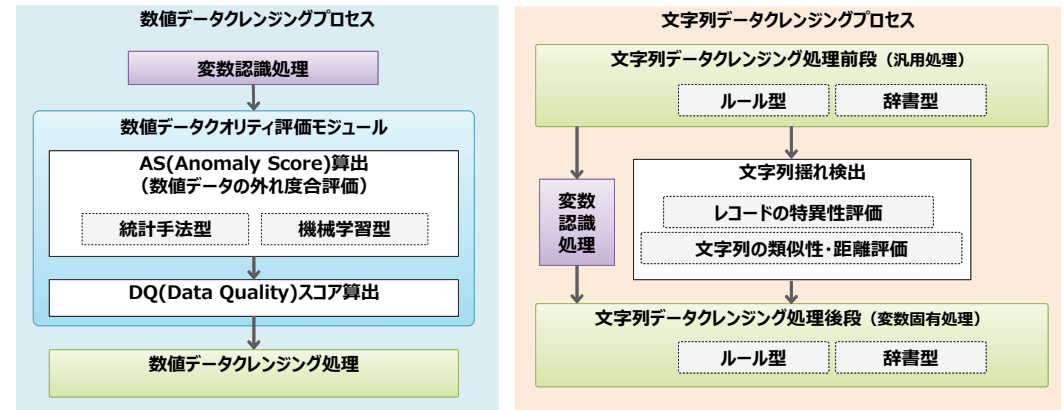
データサイエンスの試行を支援する技術の構築

本研究開発は、データサイエンスの試行を支援し、初学者でも簡単に試行に取り組める技術を構築することを目的としています。特に、インフラ維持管理分野を対象として、データサイエンスの試行に必要な手続き(プロセス)を体系化し、それに必要なデータクレンジング、データモデリング、データ分析結果の説明技術などの要素技術の研究開発を行いました。



データクレンジングに必要なデータクオリティ評価

データサイエンスのプロセスにおいて、最も多くの時間を要するデータクレンジングを支援する技術の研究開発を行いました。データエラーの混入などにより低下する「データ品質(DQ: Data Quality)」を評価する「データクオリティスコア(DQスコア)」の算出技術について、量的データを対象とする手法を開発しました。また、文字列データを対象として、データクオリティ低下の課題となる値の「揺れ」の検出・評価技術を開発しました。これらの技術によって、データクレンジング対象の特定や、クレンジング効果の評価が可能となり、データサイエンスプロセスにおけるデータ理解の深化促進や、データクレンジングの実務効率化を実現しました。



データサイエンス教育・人材育成への展開

本研究開発成果の技術を活用可能なプロトタイプシステムとして、「データサイエンス試行支援システム」を開発しました。このシステムでは、インフラ維持管理分野におけるデータサイエンスに必要なデータクレンジング、データ調製、データ分析の有用な手続きが、ひとまとまりの処理(「レシピ」と呼んでいます)として、蓄積され、利用者間で共有できる構成となっています。データサイエンスの過去の事例などに基づき、有用なレシピを予め蓄積しておくことで、初学者であっても、簡単にデータサイエンスを試行でき、また、その体験からデータサイエンスのプロセスについての学びを得ることができます。さらに、このシステムと連携して利用が可能な教材「インフラ維持管理データサイエンスガイド」の開発も行なったほか、この内容を元にした演習教材への発展にも取り組みました。

これらの成果を、さらに効果的なものとし、我が国のSociety5.0時代を支える人材輩出に貢献するべく、研究代表者が所属する北海道大学 数理・データサイエンス教育研究センターの取組と連携することで、学生や社会人を対象としたセミナー等において、ニュー・ノーマル時代に対応するe-Learningコンテンツとして、積極的な活用の検討を進めています。また、インフラ維持管理分野だけでなく、多様な分野への教育・人材育成の展開に繋げ、本成果技術の社会実装を通じた社会貢献の推進を目指していきます。

