

## 平成29年度 情報工学コース卒業研究報告要旨

石川研究室 研究室	氏 名	野田昌太郎
卒業研究題目	大規模データ分析のための可視化手法に関する研究	
<p>近年、コンピュータで扱われるデータサイズが爆発的に増加しており、大規模データ分析の需要が高まっている。インターネット上のサイト、スマートフォンのアプリケーションなどを筆頭に、大規模データを得られる機会が多くなり、それらを分析して見識を得るための分析手法に関する研究も盛んに行われている。</p> <p>大規模データ分析への需要が高まったことで、データベース上で分析を行うインデータベースアナリティクス (in-database analytics) が注目を集めている。インデータベースアナリティクスはデータベースからデータを取り出さずに分析を行う手法である。この手法では、分析処理をデータベースのライブラリとして実装するため、他のソフトウェアにデータを渡す必要がなく、高速な分析処理が可能になる。また、データベース上で分析処理を行うため、選択や射影などのデータベースらしい機能と分析処理を組み合わせることができ、SQLのような記法でデータ分析が可能という点も利点である。</p> <p>また、大規模データ分析の上ではデータの全体像をユーザに提供するような俯瞰的な可視化が重要である。データの可視化には、データの全体像を提示する俯瞰的な可視化、要素の詳細な情報の可視化の2種類があるが、複雑なデータの分析は全体像がわかりづらく、どこを掘り下げれば意味のある情報が得られるのか検討がつかないことも多い。それらの複雑なデータを分かりやすく簡素に表示できれば、ユーザが直感的に複雑なデータを理解しながら分析することが可能となる。</p> <p>本研究ではインデータベースアナリティクスを用いた大規模データ分析の可視化インターフェースを提案する。In-database アナリティクスは大規模データを分析する優れた処理基盤であるが、そのインターフェースはコマンドラインのみであり、使用にはSQLの知識が必要になる。そこで可視化インターフェースをフロントエンド、インデータベースアナリティクスをバックエンドに持つフレームワークを提案し、大規模データの直感的な分析を可能とするシステムの実装を目指す。</p> <p>提案フレームワークの実装はサーバクライアント方式で行い、サーバ側はPython 言語、クライアント側はJavaScript 言語を使用した。クライアント側でユーザが行った操作に応じて、サーバ側でインデータベースアナリティクスを行い、分析結果をブラウザ上に可視化している。実装したグラフは、ヒストグラム、散布図、グラフィカルテーブルである。ヒストグラムは属性の値の分布、散布図は二つの属性の関係、グラフィカルテーブルは記述統計量の表示のためにそれぞれ使用している。また、実装した分析操作は属性のスライス、ドリルダウン、線形回帰モデル、決定木モデルの適用である。スライスとは層ごとのデータの傾向の理解、ドリルダウンはデータの詳細の理解のために実装した。</p>		