視覚化意図を考慮した
データの効果的な視覚化方法の推定

丸田　敦貴

データの視覚化はデータの内容を伝えるのに効果的である．しかしながら，データの効果的な視覚化には専門的な知識やデータの精査が必要であり，特にデータの扱いに慣れていないエンドユーザにとっては大きな労力となり得る．これに加えて，大量のデータの中から必要なデータを探さなければならない場合においては，もしデータが適切に視覚化されていたのであればデータの探索は容易であるが，視覚化がされていない場合の探索は困難であると考えられる．

既存の方法では，表形式データの統計情報，例えば行の数や列の値の分散などを特徴として学習を行い，各データに適した視覚化種類（例えば，円グラフや棒グラフ，折れ線グラフなど）を推定している．しかし，これらの研究では視覚化種類を推定するときにユーザがどのような視覚化を作成したいか，という視覚化意図を考慮していない．また，入力データとして表形式データの中の視覚化に用いる列，すなわち視覚化列しか用いていないため，オープンデータから視覚化を作る際に，表形式データの精査をユーザが行わなくてはならない．

そこで本研究は自動視覚化において，「日本の人口の推移」などといったような視覚化意図を表形式データと同時に考慮し，適切な視覚化方法を推定する方法について提案する．例えば「日本の人口の推移」という視覚化意図が与えられたときに「推移」という単語に注目して折れ線グラフを予測したり，視覚化意図と親和性の高い列に注目して，その統計情報から折れ線グラフを予測することが可能である．この手法を実現するために視覚化意図から表形式データの重要な列を推定し，表形式データから視覚化意図の重要な部分を推定して効果的な予測を行う双方向アテンションを用いたモデルを提案する．具体的には，視覚化意図の各単語と表形式データの各列をベクトル化して各単語と各列の類似度を計算し，その類似度を加味した視覚化意図ベクトルと表形式データベクトルを用いて視覚化種類の予測を行う．さらに追実験として列ごとの類似度から視覚化列の予測を行った．

この実験を行うためのデータセットを入手できなかったため，表形式データと視覚化のペア 183,427 データのデータセットを新たに構築した．実験を行なった結果，双方向アテンションを使ったモデルがベースラインを上回り，提案モデルの中で最も良い性能を示した．また視覚化意図と表形式データの統計情報の両方を使うとより高い性能を示した．さらに，視覚化種類の予測に効果的な視覚化意図の単語や予測の精度が高い視覚化種類を分析した．しかし，視覚化列の予測では提案モデルがうまく機能しなかった．今後の課題としては視覚化方法の予測だけでなく，視覚化列の予測精度が上がるモデルを提案することや，軸やレイアウトなどの予測を行うことが挙げられる．

（指導教員　　加藤　誠）