研究テーマ:国会議事録APIを用いた内閣支持率と答弁の相関の予想

研究内容:

私は国会会議録APIを用いて内閣支持率と大臣の答弁データの相関関係を導きだす研究を行う予定である。国会会議録APIには国会における答弁データが時系列、発言者ごと、会議ごとに整然と保管されている。この優良な言語資源を用いて統計・機械学習を行い、人の目では気づきにくい相関関係を導きたい。視覚メディア研究室では先行研究として、経営トップのメッセージに着目し、経営状態が良好な企業と芳しくない企業のメッセージの特徴を見いだし、経営状態を自然言語処理によって予測するという研究が行われていた。このような切り口で、内閣支持率と国会における大臣の答弁データを比較し、支持率が高い時、低い時にはどのような特徴が出ているか。またどういった発言をすると支持率が上下するか予想したい。具体的には後述する末尾表現を用いた特徴分析やディープラーニング等を用いて単語ベクトルを生成し、時系列分析をする。現在行っている研究は、

「衆議院予算委員会における国会議員の発言の末尾表現から言葉遣いの特徴を抽出する」

というものである。以下に具体的手法と結果を示す。

　まず民主党政権時代の野党と、現在の野党の発言データを二つのカテゴリーとして年代・人物で分けて収集する。次に集めたデータを形態素解析し、文章の末尾４素性のみを取り出す。そして、二つのカテゴリーの中で出現頻度が偏っている言葉を特長語としてとらえる。また、機械学習アルゴリズムにかけ、どの議員が各カテゴリーに分類されるか、カテゴリーを特長づける末尾表現は何かを調べる。現時点の研究成果として、最大エントロピー法という機械学習を用いて、二つのカテゴリーの分類正答率を74%まで達成することに成功している。また、旧野党は「○○であります」「〇〇です」など、事実を確認する末尾表現、現野党は「○○ですよね」「でよろしいですか」など言質を取るような末尾表現が上位素性として出力されている。

　現在行っている研究結果からも確かに、人の目から気づきにくい相関関係を導き出すことに成功している。今後は、国会会議録APIの大量かつ整然と保管されているという特徴を用いて、政治という個人の思想が大きく左右する話題を客観的かつ定量的に調査することで適切な判断をし、投票等の新たな指標になる研究結果を打ち出すことを目指す。