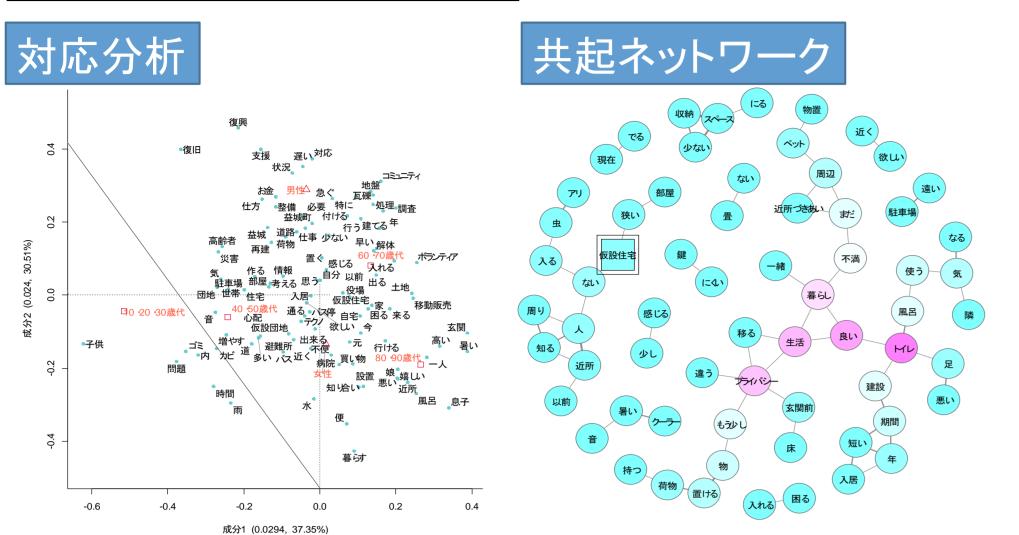
トピックモデルを用いた益城町仮設住宅聞き取り調査の自由回答の基礎分析

熊本大学大学院 〇川野 倫輝 佐藤 嘉洋

▶背景·目的

背景

▼従来のテキスト分析手法



課題

- -部の変量・単語を抽出して分析
- 単語の関係性を表すものが多い 網羅性・俯瞰性に欠ける

目的

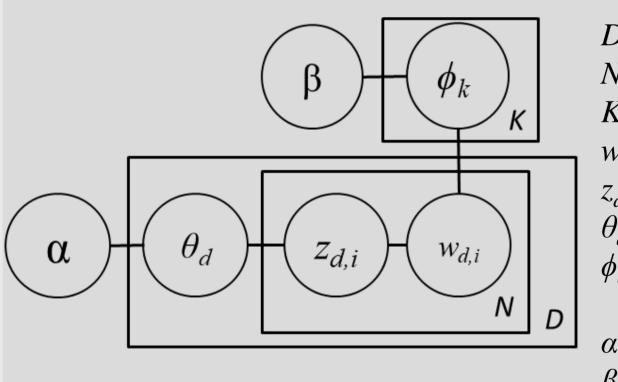
- ① トピックモデルを用いた自由回答文中の トピックの俯瞰的把握
- ② 各トピックあたりの対話時間に着目した分析

▶ 分析アプローチ

▼トピックモデル (LDA) の概要

- ∅ 大量の文書集合から話題となっているトピックを 抽出する文書データの解析手法
- ∅1つの文書が複数トピックから構成されると仮定 トピックの分布を用いた分析が可能

▼LDAのグラフィカルモデル



- D:文書数 N:単語数 *K*:トピック数
- w_{di} : 文書d のi 番目の単語 $Z_{d,i}$: $W_{d,i}$ に対応する潜在変数 θ_d :文書d のトピック分布 ϕ_{dv} :トピックkにおける単語vの出現確率
- α:ハイパーパラメータ β:ハイパーパラメータ

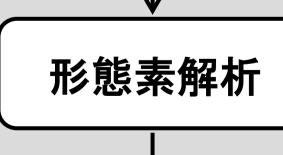
益城町仮設聞き取り調査

概要 益城町内の仮設住宅を対象とした訪問 聞き取り調査

目的 - 住民が希望する災害公営住宅の戸数の把握 - 不自由な点,不安の把握

2016年6月30日~11月20日 調査実施戸数 1,196戸(81.4%)





名詞•動詞•形容動詞 総語数:21,511語 異なり語数:3,591語

特徴量:tf-idf値

トピック抽出

崩壊型ギブスサンプリング サンプリング回数:1000回 $\alpha = 1.0$ $\beta = 1.0$

トピック数:15

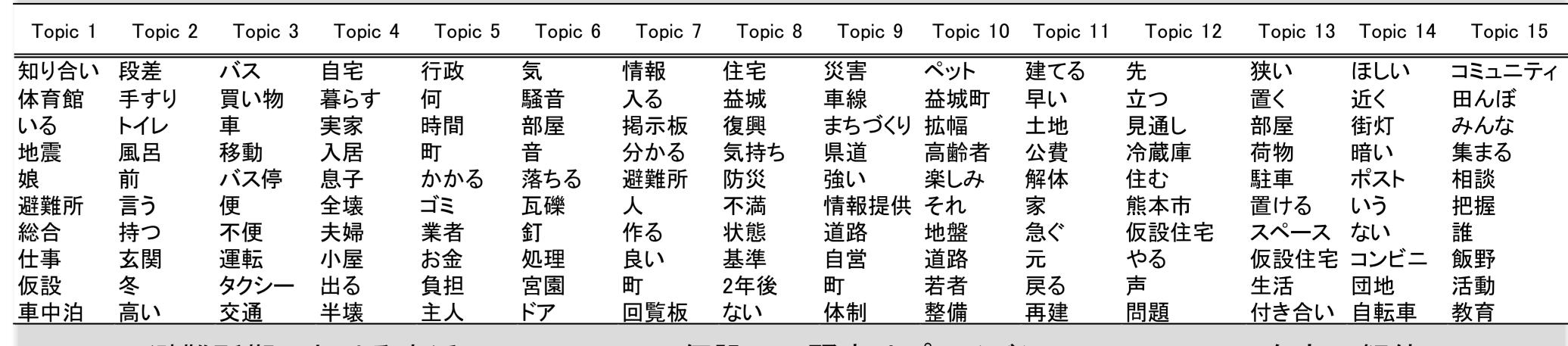
トピック分布

トピックあた りの対話時間

トピック分布と対話時間から トピックあたりの 仮想的な対話時間を推定

トピックモデルを用いた基礎分析

①トピックの抽出結果



Topic 1: 避難所期における生活

Topic 2: 仮設内のバリアフリー環境

Topic 3:仮設の交通環境 Topic 4: 家族の様子

Topic 5: 行政の対応

Topic 6: 仮設での騒音やプライバシー

Topic 7:情報の提供・獲得手段

Topic 8: 仮設後の住まい Topic 9:災害に強いまちづくり

Topic 10: 仮設外の要望

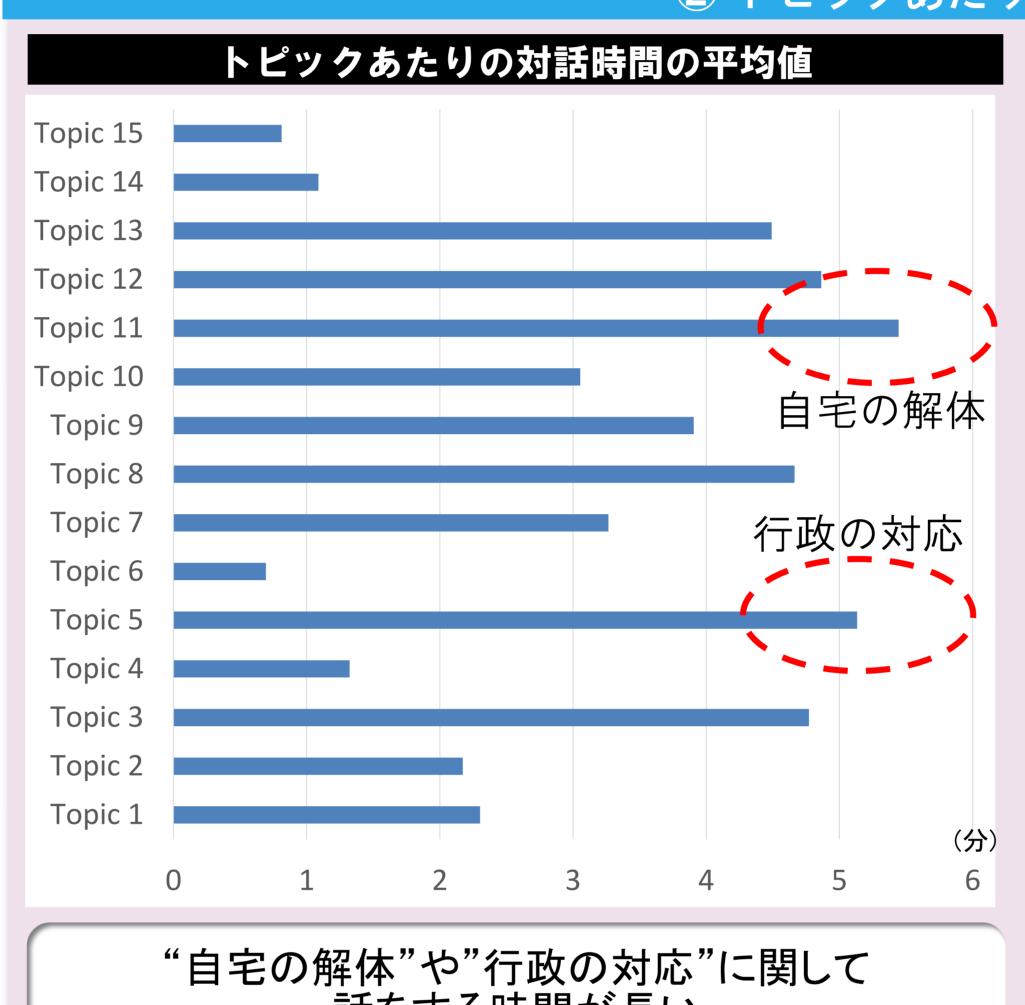
Topic 11: 自宅の解体 Topic 12: 今後の見通し

Topic 13:仮設の用地不足

Topic 14: 仮設周辺の整備要望

Topic 15:コミュニティ

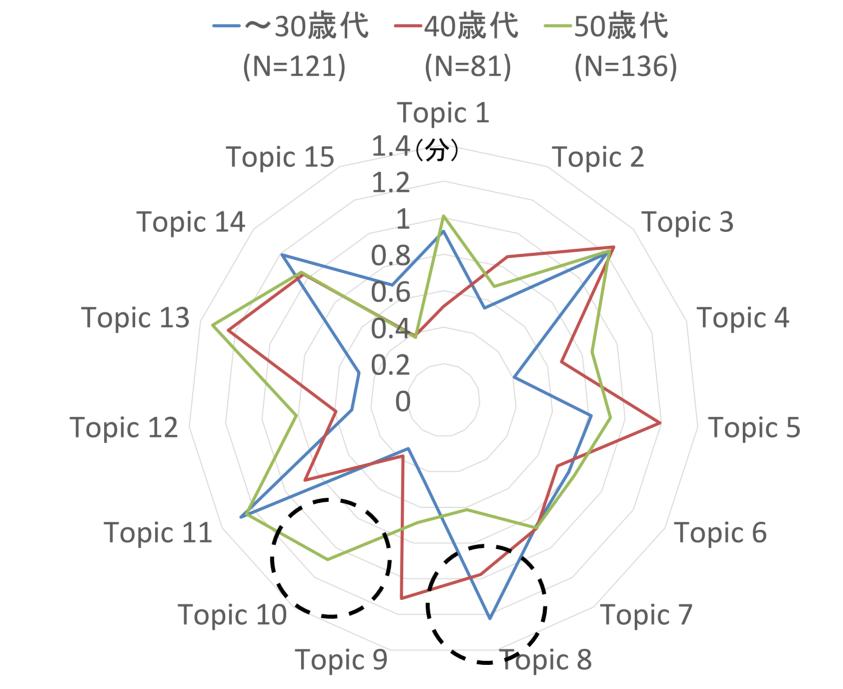
② トピックあたりの対話時間の分析



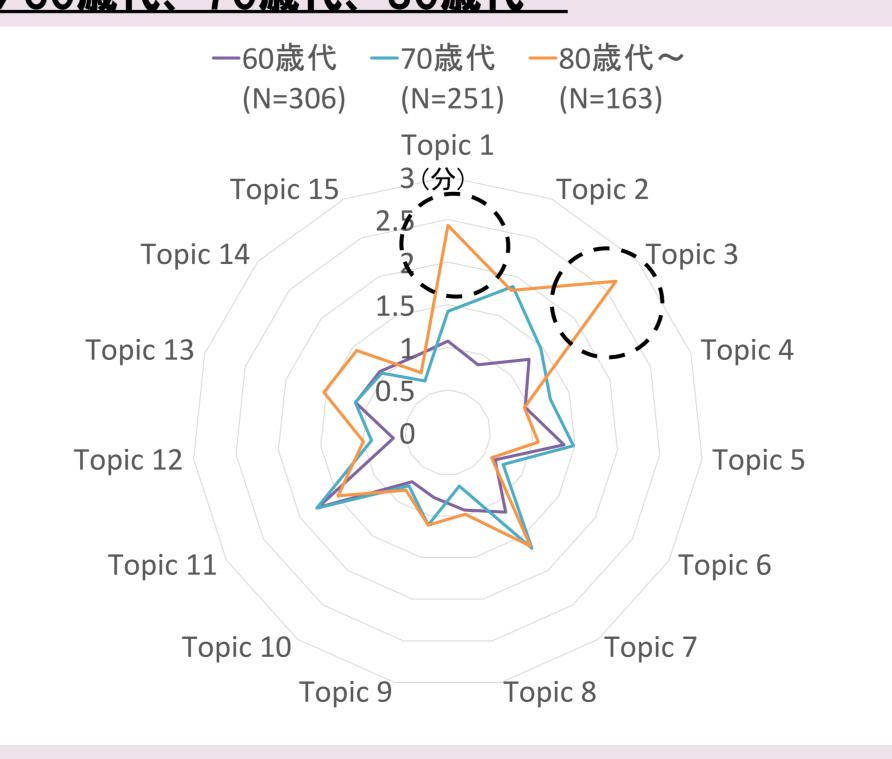
話をする時間が長い

年齢層別のトピックあたりの対話時間

▼~30歳代、40歳代、50歳代



▼60歳代、70歳代、80歳代~

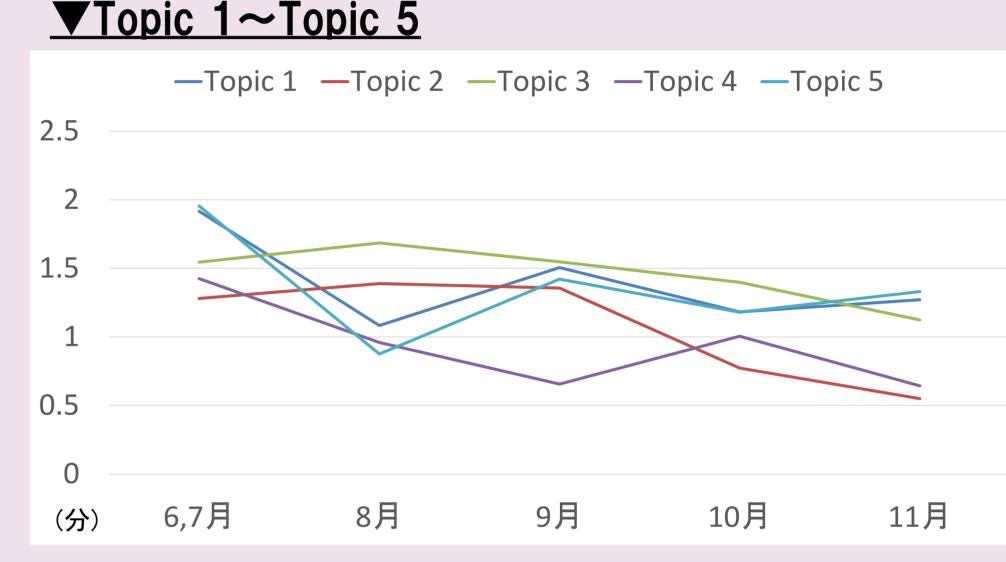


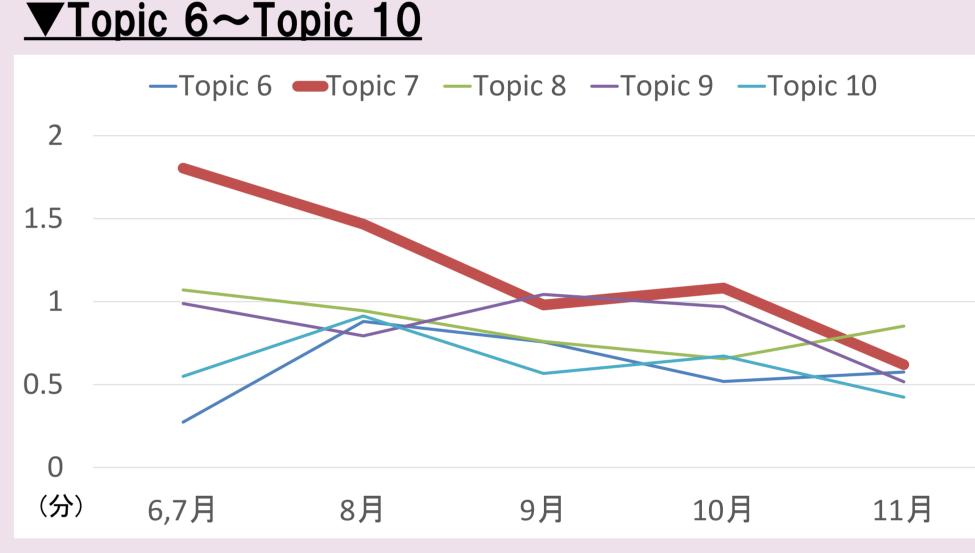
~30歳代 > "仮設後の住まい"

♪ "仮設外での要望"

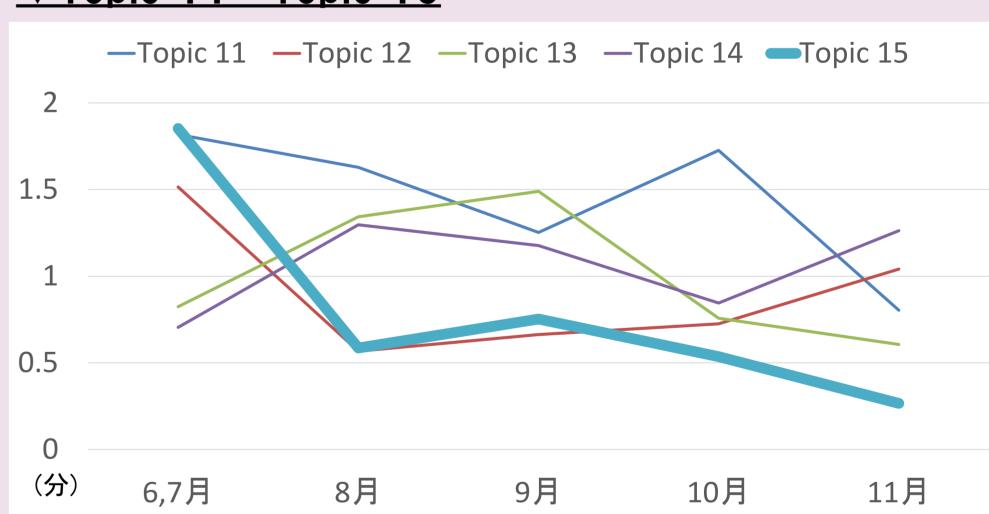
〉"避難所期における生活"&"仮設の 交通環境"

トピックあたりの対話時間の時系列変化





▼Topic 11~Topic 15



♪ "情報の提供・獲得手段"&"コミュニティ"

トまとめ

結論

- ◆トピックモデルを用いて、益城町仮説聞き取り調査の 自由回答から15個のトピック抽出した
- ◆トピックあたりの対話時間に着目した分析から、 80歳代~では避難所期における生活や仮設の交 通環境に関するトピックがよく話され、 時間の経過によって減少するものに情報提供・獲得手 段とコミュニティに関するトピックがあることがわかった

課題と今後の展望

- ◆他調査データとの比較による、より長期のトピック推移 の把握
- ◆自由回答分手法としてのトピックモデルの精度の検証

第56回土木計画学研究発表会・秋大会 2017.11.03 @岩手大学

