

# テキストマイニングによる FGI データの分析

## Analysis of Focus Group Interview Data by Text Mining

### 1. はじめに

今プロジェクトでは、テキストマイニングの基礎を学ぶとともに、形態素解析ソフト「茶筌」、Microsoft Excel を用いて、原子力に関する FGI (Focus Group Interview) データの分析を行った。

### 2. テキストマイニング

#### 2.1 テキストマイニングとは

テキスト (データ) を分析し、分析者にとって有益な知識や情報を取り出そうという技術。

##### 2.1.1 テキストマイニングの目的と要素

- (1) 多量のテキストから必要な情報を探し出す。
- (2) 要約して論旨を明確にする。
- (3) 意図していなかった新しい情報の発見。

#### 2.2 「茶筌」とは

テキストを、文法的に意味づけが可能な最小単位にわけて、それぞれの品詞情報を与えてくれるソフト。

### 3. FGI データの分析

#### 3.1 FGI の概要

30~70 代の男女約 40 人 (原子力発電所立地周辺住民 19 人、電力消費地域住民 20 人) をそれぞれ男女だけ 5 人ずつに分けてインタビュー。

##### 3.1.1 FGI の主な質問内容

リスク情報を活用した原子力発電所の検査の見直しについてのインタビュー。以下の内容を具体的に尋ねた。

- (1) リスク情報を活用した原子力発電所の検査見直しに対する印象。
- (2) 関係主体に対する信用。
- (3) 見直しにより節約された資源の適切な用途。
- (4) 非原子力分野へのリスク情報を活用した検査の適用の妥当性など。

#### 3.2 分析 (地域、性別による傾向) の手順

- (1) FGI データを茶筌により形態素解析
- (2) 品詞情報を基にキーワードを取り出す
- (3) キーワードの出現頻度をヒストグラム化
- (4) 分析

#### 3.3 分析結果 (地域、性別による傾向の違い)

意見がわかるであろう、「見直しにより節約された資源の適切な用途」について分析。

##### 3.3.1 電力消費地域住民

- (1) 女性 「お金、安全、研究、情報、安い、還元、料金」といったキーワードが頻度上位にきていることから、
  - 消費者還元、電気料金の値下げ
  - 他の安全対策

- 原子力に関する研究開発、安全研究といった用途を挙げている発言者が多いと予想できる。

- (2) 男性 「還元、地元、安全、開発、研究」といったキーワードが頻度上位

- 立地周辺への還元、振興策
- 原子力に関する研究開発、安全研究といった用途を挙げる発言者多数。

##### 3.3.2 原子力発電所立地周辺住民

- (1) 女性 「教育、地元、還元、住民、コミュニケーション」といったキーワードが頻度上位

- 立地周辺への還元、振興策
  - 広報、情報提供、市民への原子力教育
  - 従業員教育、専門家育成
- といった用途を挙げる発言者多数。

##### 3.3.3 考察

キーワードの頻度で、大まかにインタビューの中身が掴めることがわかった。

#### 3.4 発言者 (電力消費地域住民) ごとの傾向の分析

発言者 (電力消費地域住民のみ) ごとにキーワードをベクトルにして、発言者同士の余弦を計算。それぞれの類似度を数値化する。

##### 3.4.1 分析結果

- (1) 発言数が少なく、質問に沿った発言をしている人は、比較的他の発言者との類似度が高い。
- (2) 発言数が多く、質問に対する答え以外にも自分独自の意見を多く述べている人は、比較的他の発言者との類似度が低い。

#### 4. 結論

- (1) 今回の手法は、発言者やグループの意見を大まかに示すことにはそれなりに有用。
- (2) 強烈に自分の意見を持っている発言者をみつけることが可能。
- (3) 今回の手法では、細かなニュアンスの違いや、少数だが貴重な意見を反映できない。
- (4) 可視化が不十分。

#### 参考文献

- 1) 林俊克 著, Excel で学ぶテキストマイニング入門, オーム社 (2002).
- 2) 徳永健伸 著, 辻井潤一 編, 言語と計算 5 情報検索と言語処理, 東京大学出版会 (1999).