プレゼンテーションにおける ジェスチャーに着目した英文テキストの重要箇所推定手法

A Prediction Method of Important Sentences in English Text based on Gestures in Presentation

尹 泰昌[*1], 鷹野 孝典[*1]
Teachang YUN[*1], Kosuke TAKANO[*1]
[*1] 神奈川工科大学情報学部情報工学科
[*1]Information Engineering, Faculty of Information, Kanagawa Institute of Technology
Email: s1721059@cce.kanagawa-it.ac.jp, takano@ic.kanagawa-it.ac.jp

あらまし：本研究では，プレゼンテーションにおけるジェスチャーに着目した英文テキストの重要箇所推定手法を提案する．提案手法の特徴は，プレゼンテーション中のジェスチャーに対応づく英文テキストを話者が伝えたい重要箇所と捉え，そのテキストに重要度に応じたラベルを付与し作成したコーパスを学習させることで，非言語情報を持たないテキストを対象とした重要テキスト推定モデルを構築する点にある．実験により，提案方式の実現可能性を検証する．
キーワード：プレゼンテーション，ジェスチャー，重要箇所推定，コーパス，重要度

1. はじめに

プレゼンテーションでは，伝えたい重要な内容をジェスチャーや表情などの非言語コミュニケーション(1)を用いて，聴衆が内容を理解しやすくできる．既存の重要箇所推定手法(2)では，語の出現頻度を尺度にしているが，同じ単語でも前後の文脈により非言語コミュニケーションをするとは限らないため，英語プレゼンテーションにおける重要箇所を推定するのは困難である．

2. 提案方式

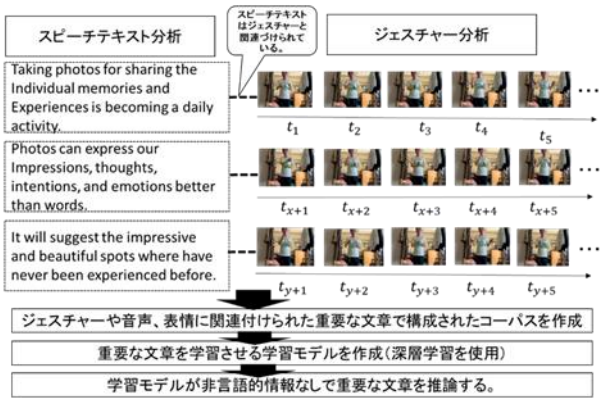

図 1　提案方式の概要図

本研究では，この課題を解決するために，ジェスチャーと関連した英文テキストの情報を学習することにより，英文テキストの重要箇所を推定する方式を提案する(図 1)．提案方式の特徴は，プレゼンテーション中のジェスチャーに対応づく英文テキストを話者が伝えたい重要箇所と捉えることにより，話者が伝えたい内容としての意味情報を各文章に付与し，その意味情報を伴ったテキスト集合（以下，重要テキストコーパス）を学習させることにより，非言語情報を持たないテキストを対象とした重要テキスト推定モデルを構築する点にある．

重要テキストコーパスの生成:

プレゼンテーション動画の各画像に対して，頭，首，腕などの 11 個の姿勢リンク情報を抽出し，姿勢リンク画像を生成する(図 2)．

図 2　元画像(左)と姿勢リンク画像(右)

表 1 重要テキストコーパスの例

テキスト	重要度ラベル
The other is when I had to say goodbye to my father when he was terminally ill.	重要
Families are the heart and lifeblood of the migrant trail.	やや重要
though they don't write your destiny.	普通

姿勢リンク画像をクラスタリングし，クラスタ内画像数C_nと総画像数 S を用いて，クラスタ n における姿勢リンク画像の重要度を$T_n = S/C_n$(式 1)または$T_n = C_n/S$(式 2)で算出する．式 1 はクラスタ内画像数が少ない場合に重要度が高くなり，式 2 は多い場合に重要度が高くなる．重要度T_nの値に対して閾値

を設定し，重要，やや重要，普通などの重要度ラベル L_x を付与する．さらに姿勢推定画像に対応するスピーチ文章 *text* を抽出し，重要度ラベル L_x の付いた文章集合を重要テキストコーパスとして生成する．表 1 に重要テキストコーパスの例を示す．

重要テキスト推定モデルの生成:

重要テキストコーパスを，深層学習を用いて学習し，重要テキスト推定モデルを生成する．重要テキスト推定モデルは，英文テキストを入力として，重要度ラベルを推定結果として出力する．

3. 実験

実験では，実際のプレゼンテーション動画から重要テキストコーパスを生成し，コーパスを学習することにより構築した重要テキスト推定モデルを用いて，入力テキスト中の重要箇所が推定可能であることを確認する．

TED プレゼンテーション動画 10 個を分析し，閾値の異なる 4 種類の重要テキストコーパス C1～C4 を作成した．さらに各コーパスを用いて LSTM[3]により学習した重要テキスト推定モデルの精度を比較する．重要度ラベルは普通，やや重要，重要の 3 種類を設定し，2 つの閾値を設定した．生成したコーパス C1～C4 の詳細を表 1 にそれぞれ示す．

(C1) ノイズ画像を人の判断で除外する前処理を手動で行い，閾値(6, 10)を設定する．
(C2) 前処理: 自動，閾値: (6, 10)
(C3) 前処理: 手動，閾値: 動画ごとに動的設定
(C4) 前処理: 自動，閾値: 動画ごとに動的設定

表 1 重要テキストコーパスの詳細

コーパス	テキスト数			
	重要	やや重要	普通	合計
C1	126	298	876	1,300
C2	194	495	1,375	2,064
C3	268	542	660	1,470
C4	354	888	1,097	2339

重要テキスト推定モデルの評価のためにテスト文章を 80 文用意した．C1～C4 のそれぞれのコーパスを重要テキスト推定モデルに学習させ，テスト文章を推定したときの正解率 a の結果を表 2 に示す．ここで，正解率を a=(推定できたテキスト数)/(判定不能以外の全テキスト数)x100 として算出した．また，「判定不能」のテキストとは，普通，やや重要，重要のどの重要度ラベルにも推定されなかったテキストである．

表 2 各コーパスの正解率比較

コーパス	C1	C2	C3	C4
正解率	38.7%	34.3%	44.8%	43.0%

表 2 の内訳として，各コーパスの重要度ラベルごとの推定数と重要度ラベルごとの正解数を，それぞれ図 3 と図 4 に示す．

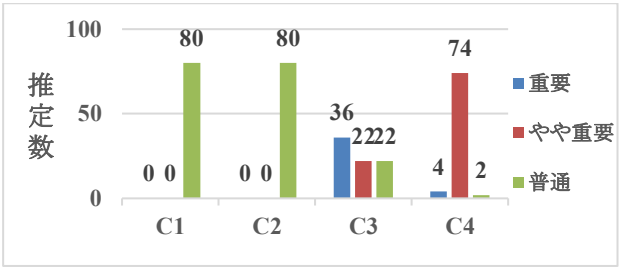

図 3 重要度ラベルごとの推定数

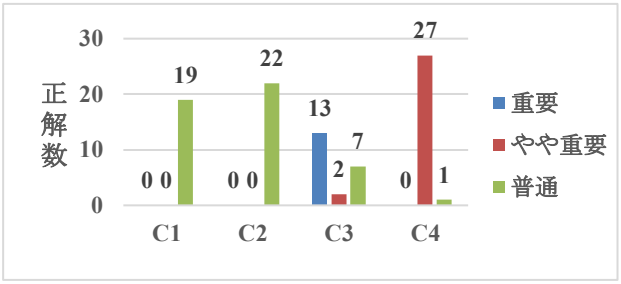

図 4 重要度ラベルごとの正解数

4. 実験考察とまとめ

実験結果より，全動画で共通の閾値により生成したコーパス C1 と C2 では，テキスト数の割合に偏りが生じ，「普通」の重要度ラベルのテキスト数が他と比較して多かった．このため，推定モデルが「普通」らベルのみ推定しており，C1 と C2 から生成した推定モデルの正解率はそれぞれ 38.7%と 34.3%であった．また，各動画で動的に異なる閾値により生成したコーパス C3 と C4 では，テキスト数の割合に大きな偏りはなく，生成した推定モデルの正解率はそれぞれ 44.8%と 43.0%と向上した．

以上から，実際のプレゼンテーション動画から重要テキストコーパスを生成し，重要テキスト推定モデルにより英文重要箇所を推定できる見込みが得られ，提案手法の実現可能性を確認できた．

今後の課題として，姿勢リンク画像の重要度算出式 1 と 2 の改善があげられる．また，重要度分析対象とする動画データ数を増やすとともに，生成した重要テキストコーパスに対して数種類のパターンの閾値で検証する必要がある．

参考文献

(1) 高木：コミュニケーションにおける表情及び身体動作の役割，早稲田大学大学院文学研究科紀要第 1 分冊 51，pp.25-36，(2005)
(2) 岡崎直観，松尾豊，石塚満，”テキストの重要箇所推定のための読み手のモデル”，卒業論文執筆要領”，第 4 回 AI 若手の集い MYCOM2003 予稿集，pp.86-89 (2003)
(3) Seep Hochreiter Jurgen Schmidhuber : LONG SHORT-TERM MEMORY, Neural Computation, Volume9,Issue8, (1997)