

ジェスチャー情報を利用した対話中の話題遷移誤検出削除

高橋 広大 井上 雅史

山形大学大学院理工学研究科

1 序論

対話における話題遷移の箇所を自動的に検出できれば、対話構造把握の手助けとなる。しかし、書き言葉とは異なる話し言葉の性質のため、テキストセグメンテーション手法による遷移の検出には誤りが多い。対話参加者のジェスチャーの開始点と終了点の情報を利用して誤って検出された話題遷移を削除し、適合率を改善することが可能である [1]。ただし、改善の度合いは、ベースとなるテキストセグメンテーション手法の性能に依存する。本研究では、複数のテキストセグメンテーション手法を比較し、対話中の話題遷移検出における、ジェスチャー情報の利用可能性を検証する。

2 テキストセグメンテーション

文章をセグメントに自動的に分割することを目的としたテキストセグメンテーション手法がいくつか考案されており、Herast による TextTiling は単語分布を利用する代表的な手法である [2]。単語の繰り返しや類義性などの語彙的な結束性をする。TextTiling では文書のある一定範囲における単語頻度ベクトルと、隣接する同一の大きさの一定区間内の単語頻度ベクトル間の距離を類似度とする。セグメントの境界では前後のセグメントの話題が混ざり合うので語彙的結束性は一時的に低下する。そのため、前後のセグメントの \cos 類似度のギャップで分割することができる。文書における連続した類似度が顕著に変化する点を境界として分割する。アルゴリズムの流れを以下に示す。但し、境界は単語と単語の間（境界可能点）に存在するものとする。

1. 境界可能点の前後に一定範囲（以下では窓と呼ぶ）を設け、各窓内の単語列から生成された単語頻度ベクトル間の \cos 類似度を計算する。
2. すべての境界可能点で \cos 類似度を求め、類似度の系列の極小を見つける。
3. 極小と前後近傍点間との \cos 類似度のギャップを合計し、閾値 $\mu - \sigma / 2$ を超えた場合、その箇所をセグメントの境界として出力する。ただし μ と σ は文書全体での \cos 類似度の平均と分散である。

TextTiling は窓サイズによって性能が大きく変動してしまう。そこで内山らはテキスト $W = w_1 w_2 \dots w_n$ が与えられたとき、 m 個の区間からなる分割 $S = S_1 S_2 \dots S_m$ の確率 $Pr(S|W)$ を式 (1) のように定義し、最小記述長原理による最適な分割を動的計画法で求める U00 を考案している [3]。

$$Pr(S|W) = \frac{Pr(W|S)Pr(S)}{Pr(W)} \quad (1)$$

このとき、 $P(W)$ はテキスト W の確率なので定数であり、 $P(S)$ は分割 S の確率である。U00 では最適な分割として $P(S)$ が最大の値になるものを算出して TextTiling より良い性能を実現している。

上記の 2 つの手法では個々の単語の出現頻度が少ない場合に類似度を測定するのが困難であった。そこで、TextTiling を発展させた手法として TopicTiling では LDA (Latent Dirichlet Allocation) を利用して単語の代わりにトピックを用いることで低次元の密なベク

表 1: グランドチャレンジ会話データ概要

会話総時間	10 分 17 秒
会話人数	三人
注釈付与されている モダリティ	発話, 手のジェスチャー, 頷き, 視線

トルを構成し、U00 より評価指標 WindowDiff の点で良い性能を実現している [4]。

3 提案方式

東山らはジェスチャー産出が談話の結束性の高さと関係があることを指摘している [5]。本研究では「対話において、話題の内部ではジェスチャーは継続され、話題の境界ではジェスチャーは終了、または開始される性質がある」という仮定を置き、ジェスチャーの開始・終了時間を利用してフィルタリングを行い、誤検出を削除する。提案手法で扱うデータは、対話書き起こし文とその発話の開始時間と終了時間、各対話参加者のジェスチャーの開始・終了時間がアノテーションされているものとする。具体的なアルゴリズムを下記に示す。但し、テキストセグメンテーションの出力を話題遷移候補とする。

1. 単語と単語の間に定義されている話題遷移候補を発話の開始・終了の情報から時刻に変換し時間軸上にのせる。
2. 話題遷移候補の時刻より t 秒前から t 秒後までを話題遷移区間とする。
3. あるジェスチャーが、話題遷移区間開始以前から終了以降まで継続していた場合、区間を誤検出として削除する。
4. 話題遷移区間内にあるジェスチャーが開始され終了している場合、区間を誤検出として削除する。
5. 話題遷移区間内にあるジェスチャーの開始または終了のどちらかが存在した場合、削除を行わない。
6. 複数のジェスチャーが存在した場合、規則 3 と 4 が規則 5 に優先する。

フィルタリングにより話題遷移候補が話題の内部だった場合は規則 3,4 によって削除され正しい話題遷移だけを残すことで全体的な分割精度を向上させることが可能となる。図 1 に会話開始から終了までのジェスチャーによる誤検出削除の概念図を示す。但し、正しい話題遷移の箇所を矢印で表し、正しい話題遷移区間を黒塗りの長方形、誤った話題遷移区間を中抜き長方形で示す。

4 実験

4.1 実験概要

実験データは国立情報学研究所のグランドチャレンジ研究「情報環境を支える日常的インタラクションデー

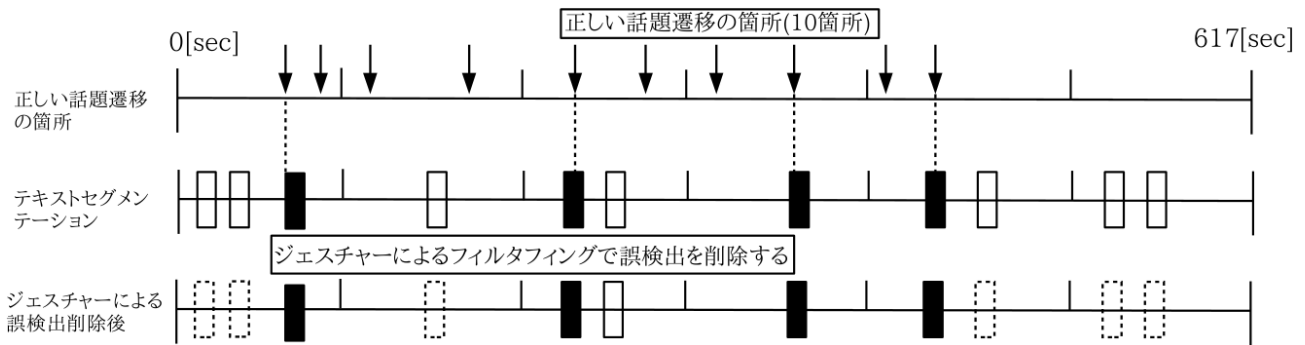


図 1: ジェスチャーによる誤検出削除の概念図

表 2: 各テキストセグメンテーションの実験結果
ジェスチャーによるフィルタリングを Ges として示す

手法	再現率	適合率	F 値
TextTiling	0.20	0.09	0.12
TexTiling+Ges	0.20	0.22	0.21
U00	0.50	0.14	0.22
U00+Ges	0.30	0.23	0.26
TopicTiling	0.26	0.26	0.26
TopicTiling+Ges	0.24	0.30	0.27

タ収録のためのプラットフォーム構築」*において収録された三人会話データを使用する。この対話はアニメーション再生課題と言われるタスクを実行しており、三人のうち二人にのみアニメーションを事前視聴してもらい、視聴していないもう一人に視聴した二人がアニメーションの内容を協力して説明をするものである。さらに、データには事前に各モダリティにアノテーションが施されている。会話データの概要を表 1 に、会話の様子を図 2 に示す。



図 2: テキストの結束性とその関係

正解となる話題遷移の箇所は会話中から 10 箇所を筆者が認定した。ベースとなるテキストセグメンテーション手法に対してフィルタリングによる誤検出削除を行った場合と行わない場合を比較して、正解の箇所に対するシステムの出力の再現率、適合率、その調和平均である F 値を用いて評価する。

*<http://research.nii.ac.jp/bonogrand/index.html>

また TopicTiling で使用する LDA モデルは GibbsLDA++[†]を使用し、 $t = 5$ とした。

4.2 実験結果

3 種類のテキストセグメンテーション手法と、同一のフィルタリングによって誤検出削除を行った結果の再現率、適合率、および F 値を表 2 に示す。TopicTiling は LDA によるトピック推定の結果が実行する毎に異なるので、100 回実行したものを示している。実験結果から全てのテキストセグメンテーションにおいてジェスチャー情報を利用した誤検出削除によって F 値が改善する。再現率の低下を招く場合があるが、適合率が大幅向上する。

5 結論

対話での話題遷移自動検出において、テキストセグメンテーションとジェスチャーによるフィルタリングの併用の有効性を調べるため、複数のテキストセグメンテーション手法を比較した。どのテキストセグメンテーション手法においても発話内容だけでは検出の難しいジェスチャーによる話題の結束性によって誤った検出を削除し、検出精度を向上させることを確認した。しかしフィルタリング後の話題遷移検出の性能は高いと言えない。話題遷移候補の出力や誤検出削除にジェスチャー以外の非言語情報を利用することが今後の課題として考えられる。

参考文献

- [1] 高橋広大, 井上雅史. 発話内容と身振りに基づく談話セグメント推定. 人工知能学会研究会資料 SIG-SLUD-B301, pp. 53-58, Sept 2013.
- [2] Marti A. Hearst. TextTiling: Segmenting text into multi-paragraph subtopic passages. *Comput. Linguist.*, Vol. 23, No. 1, pp. 33-64, March 1997.
- [3] 内山将夫, 井佐原均. 統計的手法による分野非依存のテキスト分割. 自然言語処理, Vol. 8, No. 4, pp. 19-36, Oct 2001.
- [4] Martin Riedl and Chris Biemann. Text Segmentation with Topic Models. *Journal for Language Technology and Computational Linguistics*, Vol. 27, No. 47-69, pp. 13-24, 2012.
- [5] 東山英治, 伝康晴. 談話におけるジェスチャー産出に影響する言語的要因. 認知科学, Vol. 18, No. 3, pp. 508-520, Sept 2011.

[†]<http://gibbslda.sourceforge.net/>