

第5章

評価実験

5.1 序言

本章では，COLLAGREE で行われた議論のデータを対象にした提案手法の評価実験について述べる．評価実験では同じテーマのもと行われた複数の議論を用意し，提案手法の有用性を示す．以下に本章の構成を述べる．

5.2 対象データ

5.2.1 議論データ

議論データは COLLAGREE 上で行われた別の実験での議論のものを使用する。
データの概要を以下に示す。

【実験概要】

グループ人数 : 2~3 名

議論時間 : 90 分前後

議論テーマ : 外国人観光客向けの日本旅行プランの決定

議論テーマ説明文 : みなさまに、外国人観光客向けの日本旅行プランを立てて
いただきます。 想定される旅行者の条件は以下の通りです。

- 英語は話せるが、日本語は話せない
- 初めての日本旅行である
- 日程は 6 泊 7 日
- ホテルは自分たちで手配できる
- 旅行のために貯金したので、金銭的には余裕があり、国内をいろいろとまわることが可能である
- 来日、帰国の際の空港は、どこでもかまわない
- 2 つのプランを比較したいと考えている（プランは 2 つ用意してください）

ファシリテータ : あり

5.2.2 評価データ

5.2.1 節で説明した議論データに対し，次に述べる基準でアノテーションを行ってもらった．基準を満たすと思われる発言に”1”のタグを，満たすと思われない発言に”0”のタグを付ける．

① それまで話題となっていた対象や事態とは異なる，新しい対象や事態への言及する発言

話されている内容が，以前と全く異なる対象や事態へと移行する位置でデータを区切る．

例 1:

(今までの話題:パック旅行はなぜ安いのかについて)

- A:ホテルが宿泊費の一部を出しているから安いのかな？
- B:おそらく。
- A:なるほど。
- B:沖縄行きも安いね。(今まで沖縄の話はされておらず，この後“沖縄行きのパック旅行”に話題が変わる(かもしれない))

例 2:

(今までの話題:外国人のツアー旅行の行き先について)

- A:他は寄らなくてもよいですか?(新しい行き先が出るように仕向けている)

② 既に言及された対象や事態の異なる側面への言及する発言

既に話題として取り上げられることについて，以前とは異なる側面から言及がなされる位置で区切る．

例 3:

(今までの話題:外国人のツアー旅行の行き先について)

- A:広島、長崎はどう？
- B:外国人観光客とか広島、長崎で見かけた覚えがない。
- A:ツアーに英語を話せるスタッフとか付けられるかな？(“ツアー旅行のスタッフ”に話題が変わる(かもしれない))

③ 議論のフェーズを移行させる(かもしれない) 発言

議論のフェーズを今までから移行させる(と思われる) 発言の位置で区切る．

例 4:

(今までの話題:外国人のツアー旅行の行き先について)

- A:八坂神社や清水寺など有名どころがたくさんありますし、魅力的だと思います
- B:京都周辺ツアー清水寺、金閣寺、銀閣寺、伏見稲荷大社、嵐山、など日本の建物や食べ物など広島長崎ツアー広島、長崎の戦争の地を見る事と、それぞれの場所で食べ物建造物を見るツアー(地名を挙げる段階から、各地点を結ぶツアープランへの作成段階に話題が変わる(かもしれない).)

例 5:

(今までの話題:外国人のツアー旅行のプランについて)

- A:京都周辺ツアー清水寺、清水焼体験、抹茶・和菓子など体験、きもの体験、金閣寺、嵐山、伏見稲荷大社その中で乗れそうなら屋形船などはどうでしょうか?
- B:屋形船、風情があって良いと思います。
- (途中省略)
- C: まとめると、・京都周辺ツアー京都周辺（八坂神社、清水寺、金閣寺、銀閣寺、伏見稲荷大社、嵐山、有馬温泉）、おいしい料理（豆腐など）、温泉、6泊7日ツアー
・広島長崎ツアー広島（3日）：広島原爆ドーム、平和記念公園、厳島神社、もみじまんじゅう、牡蠣、広島筆（メイクや書道なので使用する）、お好み焼き、呉の戦艦、アナゴ（移動1日）長崎（3日）：ハウステンボス、グラバー園、眼鏡橋、大浦天主堂、軍艦島、長崎ちゃんぽん、佐世保バーガー
この2プランで問題ないでしょうか?（初めて、2つのツアーの内容をまとめ、議論の収束に近づけた。）

また、ファシリテーターによる議論をコントロールするような発言も含む。

例 6:

- F:もし現在の旅先候補でよろしければ、具体的なプランづくりに移行したい
と思います。よろしいでしょうか？
- F:残り 20 分を切りました。皆様、いかがでしょうか？

以上の基準に沿ってタグを付けてもらい、”1”のタグが過半数以上付けられた発言を正解値=1，他を正解値=0 とした．

5.3 実験設定

5.3.1 パラメーター

本実験ではパラメーターは次の通りに設定した．前処理にて用いる okapiBM25 のパラメーターは $k1=2$, $b=0.75$ とし，LexRank のパラメーターは $n=50$, $threshold=0.7$ とした．分散表現として用いる fastText は次元数を 100 次元とし，学習データには wikipedia ダンプデータを用いた．総合類似度の計算に用いるパラメーターは $maxTime=5400(90 \text{ 分})$, $tWeight = 0.5$ とし，総合類似度の閾値は 0.8 とした．表 5.1 に実験の設定をまとめる．

5.3.2 比較手法

① 常時通知

最も単純かつ分かりやすい比較手法として，発言の内容に関係なく常に通知を行う手法を用いる．

okapiBM25	k1	2
	b	0.75
LexRank	n	50
	threshold	0.7
fastText	次元	10
	学習データ	wikipedia ダンプデータ
maxTime		5400
tWeight		0.5
類似度閾値		0.8

表 5.1: パラメーターの設定

② TF-IDF ベクトル

単語の意味は考慮せず出現頻度に基づく比較手法として、分散表現の代わりに TF-IDF で発言をベクトル化する手法を用いる。Algorithm??の??行目で okapiBM25 の代わりに TF-IDF を用いて連想配列を求め、重みのベクトルに変換する。発言内容の類似度計算は提案手法と同じで Cosine 類似度を用い、以降も同じである。

5.3.3 評価指標

本実験では評価指標として適合率 (Precision)、再現率 (Recall)、F 値 (F-measure) の 3 種類の指標を用いる。

適合率、再現率、F 値はそれぞれ次のようにして求める。まず、発言の通知を行うと判定した時を予測値=1、通知を行わないと判定した時を予測値=0 とおく。次に、予測値=1 かつ正解値=1 であるものの個数を *hits*(的中数)、予測値=0 かつ正解値=1 であるものの個数を *misses*(見逃し数)、予測値=1 かつ正解値=0 であるものの個数を *falseAlarms*(誤警報数) として数える。そして、式 5.1、式 5.2 及び式

5.3 に従って適合率，再現率，F 値を計算する．

$$Precision = \frac{hits}{hits + falseAlarms} \quad (5.1)$$

$$Recall = \frac{hits}{hits + misses} \quad (5.2)$$

$$F - measure = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (5.3)$$

3 つの値はどれも値が高いほど判定精度が高いことを示す．

5.4 実験結果

実験結果を表 5.2 に示す．

手法	平均評価指標		
	Pre	Rec	F
比較手法 1	0.3	0.5	0.4
比較手法 2	0.3	0.5	0.4
比較手法 3	0.3	0.5	0.4

表 5.2: 実験結果

5.5 結言

本章では本研究で提案する話題変化の判定手法が有用であることを実験により確認した．COLLAGREE にて行われた議論データに対して基準を満たすと思われるものにタグを付けてもらい評価データとした．評価実験では発言の内容に関係

なく常に通知する手法と分散表現の代わりに TF-IDF を用いて発言をベクトル化する手法を比較手法として用いた.

実験の結果, 提案手法は適合率と再現率の両方でバランス良く高い結果を示すことが分かり, 比較手法よりも良い結果を出すことがわかった.