

目的指向型対話の脱線検出を目指した不要発話の検出

吉田 理貴 Graham Neubig Sakriani Sakti 戸田 智基 中村 哲

奈良先端科学技術大学院大学 情報科学研究科

{riki-y, neubig, ssakti, tomoki, s-nakamura}@is.naist.jp

1 はじめに

複数人が集まり、目的に沿って、協調的に物事を決定する“話し合い”という行為は意思や情報の共有を行う為に非常に重要な役割を持っており、社会的活動を円滑に進める上で必要不可欠である。一方で、話し合う目的とは異なる方向に話が進む“対話の脱線”によって、無意味な時間を過ごすことも日常的に発生している。仮に、人間の会議にシステムを参加させ、対話の脱線を回避させることができれば、回避された時間でより有意義な話し合いができると期待される。本稿では、その初期目標として対話の脱線検出を行う。

対話の脱線を検出する先行研究は著者の知る限り存在しないが、考え方が類似するものとして会議の自動要約 [1] の研究がある。要約は効率的に有益な情報にアクセスする一つ的手段として提案されているものであり、会議の中で重要であった発話を抽出する抽出型要約が主流である。対話の脱線を回避する問題と対話を要約する為に重要でない発話を削除するという問題は性質が異なる。例えば、従来の会議要約は決定した内容を抽出することを重要視するのに対し、本研究では決定した内容に到るまでのプロセスも重要視する。

本稿では、二人が協調して問題解決にあたる設定の対話コーパスを収録し、発話の必要・不要を自動分類する問題に取り組む。分類の素性として言語的特徴量や対話遷移を利用した特徴量を使用し、対話の脱線の分類率の調査と検討を実施する。

2 対話コーパスの収録とタグ付与

2.1 対話コーパスの収録内容

対話コーパスを収録するにあたり、対話者が協調して問題解決にあたることのできる題材として、本研究ではコンセンサスゲーム [2] を採用する。コンセンサスゲームとは、出された問題をグループで話し合い、グループ内で問題に対して一つの解を導き出すことを目的としたゲームのことを指す。具体的な問題内容は、“砂漠での遭難”、“月での遭難”、“ジャングルでの遭難”などの非日常的な状況において、いくらかあるアイテムの中で生き残る為に重要だと思われるものを話し合いながら決定していくというものである。本研究では、二人一組で8組に話し合ってもらい、話し合うテーマは組によって異なる。

2.2 必要・不要発話タグの付与

対話の脱線を検出する実験を行う為に、2.1節で得られた対話コーパスに対して、必要・不要のタグを付与する必要がある。その為に、人手で書き起こした対話データに対して、アノテーター三名によるタグ付与を行った。本研究では対話の脱線を“目的とは関係のない不要な発話と人間が感じた状態”とする。以下の三点を基準に各発話に対する不要タグ付与を行い、タグが付与されなかった発話を必要発話とした。

- 目的を達成する上で建設的ではない発話
- 新しい情報が含まれていない、冗長な発話
- 削除しても対話に支障がでない発話

アノテーター三名によるタグ付与の平均一致率は、71%であった。意見が割れた3割程度に関しては、アイテムの使い方に対して言及する発話が多かった。例えば「宇宙服から食べ物ってどうやって食べるんだろう?」という発話を現状確認という面から必要発話とするアノテーターと現状では確認不能な事柄の為に不要発話とするアノテーターが存在し、意見が割れた。また、相手の意見に対して何度も同意を示す発話や「うーん」などの独り言のような発話で必要か不要かの意見が割れやすい傾向が見られた。

3 不要発話の検出

本研究では不要発話を検出する為に統計的な枠組みを採用する。分類器として、条件付き確率場 (Conditional Random Fields; CRF) [3] を使用する。CRFは入力系列 \mathbf{x} から出力系列 \mathbf{y} を推測する分類器として広く利用されている。本研究で扱う不要発話の検出も、対話の連続性という観点から系列ラベリング問題の一つと考えられ、CRFによる高い分類性能が期待される。本研究で利用するCRFでは、 \mathbf{x} は発話文の系列、 \mathbf{y} は必要か不要かを表すラベルの系列を指す。分類器の素性として言語的特徴量と話題遷移を利用した特徴量を用いる。3.1節、3.2節にそれぞれの特徴量の抽出方法を示す。

3.1 言語的特徴量

人間が不要発話を検出する際、特定の表現が手掛かりになることが多い。例えば、「関係ないけど」という

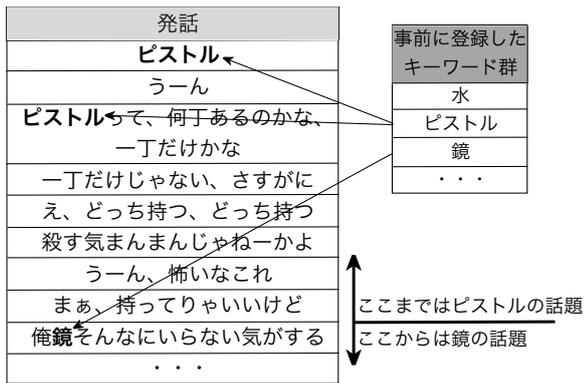


図 1: キーワードを用いた話題遷移推定の例

表現が入った発話は、不要発話である可能性が高い。逆に「話し戻そう」という発話後には、対話が本題に戻る可能性が高いと予想される。以上の理由より、言語的特徴量として、単語頻度ベクトル (1-gram) を素性に用いる。書き起こされた発話文に対して形態素解析を行い、各形態素の頻度を計算し、単語頻度ベクトルを求める。また、発話の長さが不要発話に関与する可能性がある為、発話内の形態素数も素性として加える。

3.2 話題遷移特徴量

対話の脱線を検出するもうひとつの手段として、話題の遷移を利用するという方法も提案する。対話の脱線を検出する為に、発話列を個々の話題へと分割し、各話題が必要か不要かを判断することが効果的であると考えられる。話題遷移推定の手法として、キーワードを利用した手法 [4] や語彙的連鎖性を用いた手法 [5, 6] が知られている。キーワードを用いた話題遷移とは、登録されたキーワードが出現した位置を利用して話題推定を行う手法である。語彙的連鎖性を用いた話題遷移とは、同じトピック中に類似した名詞や形容詞が出やすいという知見を利用した手法であり、類似した名詞や形容詞が含まれる連続した発話を一つのトピックとして推定する。先行研究では、トピック推定においては語彙的連鎖性が特に有効であることが確認されている。以上の2手法をそれぞれ素性とする事で、対話の脱線を検出する。また、さらに2手法を組み合わせることで、対話の脱線の検出率の向上を狙う。つまり、語彙的連鎖で推定したトピック内にキーワードが含まれている場合は対話の目的に必要な可能性が高く、逆にキーワードが含まれていない場合には対話の目的には不要である可能性が高いと考えられる。

3.2.1 キーワードを用いた話題遷移推定

キーワードを用いた話題遷移推定の例を図1に示す。文献 [4] では、対話中に初めて出現する内容語や「ほかはないの?」などの登録された表現を使用して話題遷移の推定を行っている。事前に登録したキーワードが現れた場合に、その発話から次のキーワードが表れるまでの発話をそのキーワードの話題と推定する。キーワードを用いて話題遷移を推定する事前準備として、

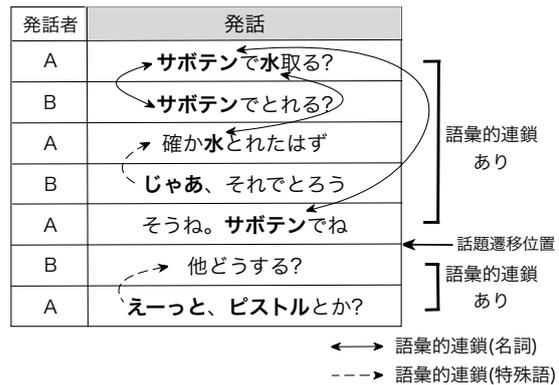


図 2: 語彙的連鎖を用いた話題遷移推定の例

対話目的で重要なキーワードとなる名詞句を抽出する必要がある。文献 [4] のように、対話目的に合わせて人手でキーワードを登録する方法も考えられるが、本研究では対話目的を説明する為に用意された資料に対して形態素解析を行い、抽出された名詞句をキーワードとする。

3.2.2 語彙的連鎖性を用いた話題遷移推定

語彙的連鎖性を用いて発話同士の結びつきを調べ、結びつきが存在する発話をまとめて一つの話題とし、結びつきがない発話を話題が遷移したと推定する手法が語彙的連鎖性を用いた話題遷移推定である。語彙的連鎖性は、連続する発話に出現する意味的に関連した語の繋がりのことである。例えば、文献 [5, 6] では、名詞、動詞、形容詞などの繋がりを使用して発話同士の語彙的連鎖を検出する。

本研究で用いる語彙的連鎖性を用いた対話遷移の検出アルゴリズムを以下に示す。ここでは便宜上、語彙的連鎖が確認された発話群を連鎖発話と呼び、事前に人手で登録した直前の発話と結びつきやすい単語を特殊語と呼ぶ。

- 現発話文に対して形態素解析を行い、名詞および特殊語を抽出
- 直前の連鎖発話に抽出された名詞が含まれている場合に、現発話を連鎖発話に追加
- 特殊語が現発話で検出された場合に、現発話を連鎖発話に追加
- 自発話と相手発話が連鎖発話に続けて追加されなかった場合に、対話遷移であると推定

図2に語彙的連鎖を利用した対話の例を示す。

3.2.3 話題遷移素性

最後にキーワードと語彙的連鎖性を用いた話題遷移を素性に必要がある。まず、キーワードを用いた話題遷移の場合には、“キーワードの出現位置”か“それ以外”という状態を定義し、これを素性とする。語彙的連鎖性を用いた話題遷移の場合には、“トピックの開始位置”、“自発話に同じ名詞が存在”、“相手の発

表 1: タグを一意に定めた際の平均評価値

使用手法	A	R	P	F
MMR	0.62	0.54	0.53	0.53
AllUnnecessary	0.40	1.00	0.40	0.57
1-gram	0.71	0.58	0.65	0.61
Chain	0.64	0.21	0.71	0.32
Key	0.65	0.40	0.58	0.46
Chain+Key	0.66	0.42	0.60	0.49
1-gram+Chain	0.71	0.57	0.65	0.60
1-gram+Key	0.73	0.59	0.68	0.63
1-gram+Chain+Key	0.73	0.60	0.66	0.63

表 2: 意見が割れた発話を除いた際の平均評価値

使用手法	A	R	P	F
MMR	0.65	0.56	0.55	0.54
AllUnnecessary	0.39	1.00	0.39	0.55
1-gram	0.76	0.66	0.70	0.67
Chain	0.68	0.28	0.78	0.40
Key	0.69	0.45	0.62	0.51
Chain+Key	0.70	0.47	0.64	0.54
1-gram+Chain	0.76	0.65	0.71	0.66
1-gram+Key	0.78	0.65	0.73	0.68
1-gram+Chain+Key	0.78	0.66	0.72	0.68

話にのみ同じ名詞が存在”, “特殊語のみ存在” という状態を定義し, これを素性とする.

4 不要発話の分類実験

4.1 実験条件

実験を行う為に 2.2 節で付与したタグデータを使用する. なお, 本稿ではアノテーター間で意見が割れた発話は多数決で一意にタグを定めた. 実験に用いたデータセットの総数は必要発話が 1743, 不要発話が 1429, 総発話数が 3172 となった. 3 節で説明した素性データの抽出の為に, 形態素解析エンジンである MeCab[7] を使用した. 分類器には CRFsuite[8] を使用した. 不要発話の分類率の評価には, 7 対話を学習用のデータとし, 残り 1 対話をテスト用のデータにする 8 対話分割交差検定を用いた. 評価尺度として, 精度 (Accuracy; A), 再現率 (Recall; R), 適合率 (Precision; P), F 値 (F-measure; F) を採用した. 再現率が高いほど人間が不要と考えた発話をシステムが再現できており, 適合率が高いほどシステムが不要と判断した発話が人間の考えに適合していることを表す. F 値は, 再現率と適合率の調和平均として定義される.

本研究で取り扱う問題において, 提案法の有効性を調べる為に, 幅広い自動要約研究で代表的な手法として知られている Maximal Marginal Relevance(MMR)[9] との比較を行った. MMR は重要文の抽出数を事前に決定する必要があるが, 本研究では抽出する対話以外の 7 対話で必要とされた発話の平均を抽出数とした. また, もう一つの比較対象として, AllUnnecessary を用いた. AllUnnecessary は, 推定される発話を全て不要としたものであり, システムが必要発話を全く検出できていない場合の評価指標である. 本実験で使用する素性としては言語的特徴量 (1-gram), キーワード (Key), 語彙的連鎖性 (Chain) を用いた特徴量とし, これらの素性の組み合わせによる評価値を調査した.

4.2 実験結果と考察

表 1 に, 多数決によって必要・不要発話タグを一意に定めた際の平均評価値を載せ, 表 2 に, 評価値を計算する際に意見が割れた発話を除いた値を載せる.

発話者	発話	Human	1-gram	Chain+Key	1-gram+Chain+Key	MMR
A	照明が九で, 八がヒーター	+	+	+	+	+
B	うん	+	+	+	+	-
A	で, 七位がロープで, えー	+	+	+	+	+
B	うん	+	+	+	+	-
A	六位, 六位FMで五位が星座とかにする? 逆の方がいい?	+	+	+	+	+
B	そーね	+	+	+	+	-
A	えー, 星座六にしようか, じゃあ	+	+	+	+	+
B	おー	+	+	+	+	-

+: 必要発話

-: 不要発話

図 3: 提案手法が有効であった対話例

フィッシャーの正確確率検定の 5%水準で本研究の全提案手法と従来手法である MMR との精度に有意差が確認された. また, 適合率は全提案手法が MMR よりも優れており, 1-gram が含まれる素性の組み合わせの場合に, MMR よりも全ての評価値が優れていた. このことから, 対話のプロセスも重要視した分類問題では, 提案手法が MMR よりも優れていることが分かる. また, AllUnnecessary との比較結果より, MMR も含めた全手法で明らかに精度の面で優れていた. このことから, システムが不要発話を偶然に選んでいないことが示された.

提案素性の中で検定を実施した結果, 1-gram と語彙的連鎖, キーワード, 語彙的連鎖とキーワードを組み合わせた場合に精度間で有意差が確認されたが, 1-gram と 1-gram と別の素性を組み合わせた場合の精度間では有意差が確認されなかった. 精度が最も良かった 1-gram, 語彙的連鎖性, キーワードの組み合わせとこの組み合わせに対して有意差が確認されなかった手法の精度を太字で表す. 以上の結果より, 精度の面で最も貢献する素性は 1-gram であることを示した. また, 表 1, 2 の結果より, 語彙的連鎖の素性は再現率に効果的であり, 1-gram の素性が適合率に効果的であることがわかった.

図 3 に提案手法が有効であった対話例を示す. Human は人間によるアノテーション結果を表す. 発話者 B の「うん」や「そーね」などの発話は, 発話者 A

5 おわりに

本稿では、目的を持った対話において脱線を検出することを目的とし、言語的特徴量と対話遷移特徴量を使用することで発話の必要・不要の自動分類を行った。結果として、対話のプロセスも重要視した場合に、従来手法のMMRよりも本研究が優れていることを示した。また、1-gramを素性として使用する場合には、不要発話の分類精度が高く、再現率や適合率の向上にも役立つことがわかった。また、対話遷移特徴量により、複数発話を話題としてひとまとめにして、不要発話検出が行えることも示した。一方で、同じ話題で必要発話と不要発話が混在するような対話では、不要発話を分類することが難しいという課題が残った。

今後は、本稿とは別の対話目的を設定した場合と対話人数を増やした場合の対話収録を行い、本提案手法の有効性の調査を行っていく。また、本研究の最終的な目標は人間の会議にシステムを参加させ、システムによって対話の脱線を回避させることである為、音声認識や音声合成も含めた総合的な対話システムの開発を目指していく。

参考文献

- [1] Korbinian Riedhammer, Benoit Favre, and Dilek Hakkani-Tür. Long story short—global unsupervised models for keyphrase based meeting summarization. *Speech Communication*, Vol. 52, No. 10, pp. 801–815, 2010.
- [2] JC Lafferty, PM Eady, and J Elmers. The desert survival problem. *Experimental Learning Methods*, 1974.
- [3] John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *ICML '01 Proceedings of the Eighteenth International Conference on Machine Learning*, pp. 282–289, 2001.
- [4] 宮村祐一, 徳永健伸. 情報探索雑談におけるトピック遷移検出. 情報処理学会研究報告 自然言語処理研究会報告, Vol. 187, pp. 71–76, 2008.
- [5] Michel Galley, Kathleen McKeown, Eric Fosler-Lussier, and Hongyan Jing. Discourse segmentation of multi-party conversation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, pp. 562–569. Association for Computational Linguistics, 2003.
- [6] Pei-Yun Hsueh and Johanna D Moore. Combining multiple knowledge sources for dialogue segmentation in multimedia archives. In *annual meeting-association for computational linguistics*, Vol. 45, p. 1016, 2007.
- [7] 工藤拓, 山本薫, 松本裕治. Conditional random fields を用いた日本語形態素解析. 情報処理学会自然言語処理研究会 SIGNL-161, Vol. 47, pp. 89–96, 2004.
- [8] Naoaki Okazaki. Crfsuite: a fast implementation of conditional random fields (crf), 2007.
- [9] Jaime Carbonell and Jade Goldstein. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 335–336. ACM, 1998.

発話者	発話	Human	1-gram	Chain+Key	1-gram+Chain+Key	MMR
A	素人がピストルぶっぱなした ところであたへんて	+	+	+	+	+
B	ジャングルのガイド兼船長やで	+	+	+	+	+
A	船長カスやねん	-	-	+	-	+
B	そっか、船長運転誤ってるもんね	-	+	+	+	-
A	そうやで、切り株に激突してん	-	+	+	+	+
B	カスやん船長	-	-	+	+	+
B	や、でも自分がカスなんやから それぐらいせなあかんやろ	-	-	+	+	+
A	カスやから最後までカスかも知らん	-	-	+	-	-
B	最後までカス	-	-	+	-	-
A	カスやきつと	-	-	+	-	-
B	どっちか微妙やな船長にかけろか	-	-	+	-	+
A	あん	-	-	+	-	-
B	船長カスやと見限る?	-	+	+	-	+
A	いや、いまそ論点ちゃうからな、 絶対に、絶対論点ちゃうから	-	+	+	-	+
A	あの、これ終わらんやつやで	-	-	+	-	-
B	微妙やな	-	-	+	-	-
A	今んとこ一位が水の缶で決まってて、 それ運ぶポートいると思うねんけど	+	+	+	+	+

+: 必要発話
-: 不要発話

図 4: 対話遷移特徴量が有効ではなかった対話例

の意見に対する同意を表しており、対話を円滑に進めていく為に重要な意味を持つ。図 3 の発話を見ると、MMR では情報の新規性に基づいた冗長性の低い発話を抽出する為、同意を表す発話を不要発話と判断した。一方で、本手法では対話のプロセスも重要視する為、同意を表す発話も必要発話と判断した。システムを会議に参加させることを考えた場合に、同意発話の度に会議を止めるシステムの導入は、対話の円滑性の面で難しいと考えられる。このことから本研究のプロセスも重要視するという考え方がシステムを会議に参加させる際に重要である事が分かる。

図 4 に対話遷移特徴量が有効ではなかった対話例を示す。2 発話目の“船長”と 3 発話目の“船長”が語彙的連鎖で結びつき、3 発話目の“カス”と 6 発話目の“カス”が語彙的連鎖で結びつき、その後も“船長”と“カス”で語彙的連鎖が続いた。“船長”や“ジャングル”、“ガイド”はキーワードとして登録されており、ひとまとめにされた語彙的連鎖の発話群が結果的に必要発話と判断された。話題としてひとまとめにする考え方では、図 4 のような、対話の中に必要と不要が混在する場合に対処することが難しいと考えられる。語彙的連鎖とキーワードを用いた手法の精度が当初考えていたよりも低かった理由も、同じ話題で必要と不要が混在する対話に対処できなかった為と考える。