

情報探索雑談におけるトピック遷移検出

宮村 祐一[†] 徳永 健伸[†]

[†] 東京工業大学 大学院情報理工学研究科

{miyamura,take}@cl.cs.titech.ac.jp

概要 従来のタスク志向型対話システムでは「目的地までの経路を知りたい」等のユーザーからの具体的な要求には対応できるが、「何か面白い情報を知りたい」といったユーザーの漠然とした要求には対応が難しい。本研究では、雑談を通してユーザーの漠然とした要求に関する情報を提供する対話システムの構築を目指しており、そのために必要となるトピック遷移のリアルタイム検出手法を提案する。提案手法では、先行研究において提案された手がかり語と対話主導権の変化による検出に加え、語彙的連鎖を併用する。雑談コーパスを用いた検出実験を行い、語彙的連鎖を併用する事で検出精度が高まることを確認した。

Real-time Topic Segmentation of information seeking chat

Yuichi Miyamura[†], Takenobu Tokunaga[†]

[†] Department of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology

{miyamura,take}@cl.cs.titech.ac.jp

Abstract Research on dialogue systems has focused mainly on ones that are task-oriented, such as planning a travel itinerary, where users have specific goals and demands within a specific domain. However, user's demands are sometimes more ambiguous and vague, e.g. simply desiring interesting information. Our ultimate goal is to build a dialogue system which provides useful information in response to the user's vague demands through chatting. To continue natural chatting, detecting topic shifts by the user in real time is crucial. In this paper, we propose a method for detecting real-time topic shifts in chat. Our method combines a method based on cue phrases and initiative change, and a lexical chain-based method. The evaluation of our method by using a chat corpus showed a decrease in error rates in detecting topic shifts.

1 はじめに

自然言語による対話システムは長年に渡って研究されているが、それら既存の対話システムは大きく以下の2種類に分類することができる。

- タスク指向型対話システム
- チャットボット

タスク志向型対話システムとは、ユーザーが明確な情報要求を持っていることを想定したシステムであり、その要求を達成するために必要な対話を行なうものである。例として目的地までの経路を検索するシステムなどが挙げられる。それに対し、チャットボットとはユーザーへの情報提供を目的とせず、単に対話を

継続することを目的としたシステムである。例として ELIZA[1] などが挙げられる。

対話システムの応用例として家庭用ロボットが考えられる。その際、ロボットに必要な機能として、「テレビをつけて」や「明日の天気を教えて」といった明確な情報要求に対応することが挙げられるが、同時に、「何か面白い情報を知りたい」といった漠然とした要求に対応できることが望ましい。しかしながら、従来のタスク志向型対話システムではユーザーが明確な情報要求を持っている事を想定しているため、これら漠然とした要求には対応することが難しい。人間同士の場合、雑談のように様々なトピックの話を行なうこと

で漠然とした情報要求を解決していると考えられ、このような雑談つまり「探索的でタスク指向性が低い対話」を [2] は情報探索雑談と定義した。そこで本研究では、情報探索雑談を通してユーザーの漠然とした要求に対応できる対話システムの構築を目的とする。この対話システムを実現するためには以下の2つの課題を解決する必要がある。

1. トピック遷移の検出
2. トピック遷移先の推定

本研究では、特に (1) のトピック遷移検出に焦点をあて、検出手法を提案する。

2 関連研究

トピック遷移検出の先行研究として、市川ら [3] が挙げられる。[3] では、明示的なトピック遷移のキーワードと会話主導権の変化を併用した検出手法を提案している。

また、トピック遷移検出に類似した研究としてトピック分割が挙げられる。トピック遷移検出とトピック分割では、処理の際に用いることができる情報量に違いがある。トピック遷移検出では対話中にリアルタイムで処理を行なう、つまりこれから話される未来の発話内容が未知の状況で検出を行なうのに対し、トピック分割では対話内容が全て既知の状況で分割を行なう。よってトピック遷移検出の方がトピック分割より難しいタスクと言える。

トピック分割の従来研究として、コサイン類似度を用いたクラスタリング手法 [4] や語彙的連鎖を用いた手法 [5][6][7]、サポートベクターマシンを用いた手法 [8] などが挙げられる。特に語彙的連鎖がトピック分割において有効であることが確認されており、[5] では会議の内容を書き起こした ICSI Meeting Corpus に対してトピック分割を行なったところ、トピック分割の誤り率を表す評価指標 Pk、WinodwDiff の値はそれぞれ 23.0 %、25.5 % であった (Pk 及び WindowDiff については 6.1 節を参照)。[6] ではトピック境界をトップレベルとサブトピックの2種類に分けて語彙的連鎖を用いた手法による実験を行なった。また、機械学習を利用した手法も盛んに行なわれており、[8] らは単語出現頻度を基にベクトルを作成し、サポートベクターマシンによってトピック境界検出を行なう手法を提案し、ICSI Meeting Corpus やニュース放送等からなる TDT Data(LDC, 2006) を用いて評価実験を行なった。

3 コーパス

2章で述べたように、従来のトピック分割研究ではコーパスとしてニュース放送や会議の書き起こしが用いられているが、本研究で使用するコーパスは人間同士の雑談を収録した Mister 0 コーパスである。これは、「びっくりしたこと」をテーマとする2話者による対面対話の記録が合計 26 対話収録されているものである。それらに対して、3人のアノテータに人手によるトピック分割をしてもらい、2者以上のアノテータがトピック境界と判断した発話を正解境界とする。ただし、収録された 26 対話のうち、1つの対話において1名のアノテータがトピックが全く切れないと判断したため、その対話はコーパスから除外し、計 25 対話で評価を行なった。25 対話のコーパスは 101 のトピック境界からなり、トピック境界間の平均距離は 40 発話であった。

4 雑談コーパスにおける語彙的連鎖の効果

トピック分割において有効である語彙的連鎖を用いた手法は、類似したタスクであるトピック遷移検出に対しても有効であると考えられる。そこでまず始めに、雑談コーパスにおける語彙的連鎖の効果を調べるため、[7] において用いられた語彙的連鎖を用いた手法 DiaSeg を本研究で使用する雑談コーパスに対して行なった。

4.1 語彙的連鎖の作成

DiaSeg とは、まず始めにテキストを一定の window 幅によって等分し、語彙的連鎖を用いて隣接する window 同士の結びつきの有無を調べる。結びつきが存在すればそれらをまとめて1つのトピックとし、結びつきのないところをトピック境界とする手法である。そのため、望月らの語彙的連鎖計算プログラム Lexical Chainers[9] を用いて語彙連鎖を作成した。連鎖を作る際の候補語となる品詞は名詞、動詞、形容詞とし、分類語彙表を用いて同概念の語をまとめて同一の語として扱った。また、除外単語リストとして以下の5種類を指定した。

1. 動詞-接尾
2. 動詞-非自立
3. 名詞-非自立
4. 名詞-代名詞
5. びっくり、笑い

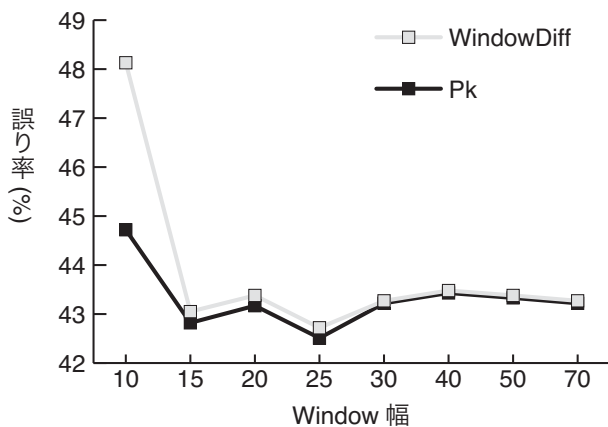


図 1: 雑談コーパスにおける DiaSeg の性能

4.2 結果

DiaSeg による雑談コーパスのトピック分割結果を図 1 に示す。window 幅を 10~70 までの間で実験したところ、window 幅が 25 のとき誤り率最小となり、Pk は 42.51 %、WindowDiff は 42.72 % となった (Pk 及び WindowDiff については 6.1 節を参照)。[7] らによる ICSI Meeting Corpus での実験結果が誤り率 18 % 程度であることと比べると、著しく悪い結果と言える。その原因として、雑談コーパスでは作成される語彙的連鎖が少ないことが挙げられる。雑談では、多くの発話は短く、また、含まれる候補語の数も少ない。そのため、まったく語彙的連鎖が作成されない区間がたびたび発生してしまい、それらをトピック境界と誤検出するようになる。本研究で目的としているトピック遷移検出では、上述のように未来発話が未知のため、語彙的連鎖作成時に使用できる情報量がトピック分割より少なく、語彙的連鎖作成時の候補語の数もより少なくなる。そのため、雑談コーパスに対しては DiaSeg のように語彙的連鎖のみによってトピック遷移をリアルタイムで検出する事は非常に困難と言える。

5 提案手法

5.1 提案手法の概要

4 章において語彙的連鎖のみでは雑談コーパスのリアルタイムトピック遷移検出は困難である事を述べた。そこで本章では、先行研究である市川らの明示的なトピック遷移のキーワード及び対話主導権による手法と語彙的連鎖を併用した手法を提案する。提案手法の流れを図 2 に示す。本手法は以下の 2 段階のステップからなる。

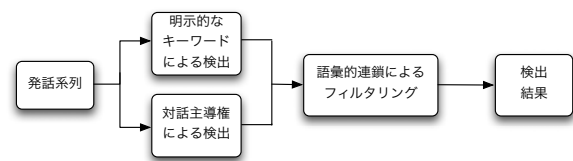


図 2: 提案手法の流れ

- ステップ 1: キーワード及び対話主導権に基づく検出
- ステップ 2: 語彙的連鎖に基づくフィルタリング

5.2 キーワード及び対話主導権に基づく検出

まず始めに、発話系列から明示的なトピック遷移のキーワードと対話主導権の交代を検出する。検出には、市川らが用いたキーワードセットに更にキーワードを追加したものを用いた。キーワードを追加した理由は、市川らの手法と違い、本手法では第 2 ステップにおいて語彙的連鎖によるフィルタリングを行なうため、第 1 ステップでより多くのトピック遷移境界候補を検出しておく必要があるためである。

明示的なトピック遷移のキーワードの検出表 1 に市川らが使用したキーワードと新たに追加したキーワードを示す。それらが表れた発話をトピック境界候補として検出する。市川らが使用したキーワードはトピック境界にのみ表れやすい語であるが、新たに追加したキーワードは、「ほかに」や「たとえば」などの短く、かつトピック境界以外でも使われる可能性が高いものを含んでいる。これによって、検出精度は悪化すると思われるが、市川らのキーワードセットでは検出できなかった境界も検出できるようになり、再現率が向上すると思われる。

対話主導権の交代検出表 1 に市川らが使用した対話主導権を表す語と新たに追加した語を示す。これらの語を発した話者が以降の対話の主導権をとると判定する。よって、これまで主導権をとっていた話者と異なる話者がこれらの語を発した時、対話主導権が交代したと判定し、その発話をトピック境界候補として検出する。

5.3 語彙的連鎖に基づくフィルタリング

4 章において語彙的連鎖が全く作成されない区間が出来てしまい、トピック境界の検出が困難である事を述べた。しかしながら、対話中には語彙的連鎖が作成できる区間もあり、その区間内では「語彙的な結びつきが存在するためトピック境界とはなりにくい」こと

表 1: キーワード例

種類	例
明示的なトピック遷移のキーワード	「ほかはないの？」 「そういえば」「ところで」
明示的なトピック遷移のキーワード (新たに追加)	「ほかに」「たとえば」「何か～こと」
対話主導権を表す語	「私」「俺」 「さっき」「この前」
対話主導権を表す語 (新たに追加)	「最近」「むかし」

は判定できる。そこで第2ステップでは、第1ステップにおいて検出されたトピック境界候補の発話と、それ以前の発話との語彙的な結びつきを語彙的連鎖の有無によって判定し、結びつきがあると判定されたトピック境界候補の除去を行なう。そして除去されずに残ったトピック境界候補を最終的な検出結果とする。使用した語彙的連鎖は4.1節において作成したものである。

6 評価実験

6.1 評価指標

評価指標としてRecall、Precision、F値、Pk、WindowDiffの5つを使用した。PkとはBeefermanら[10]が提案したトピック分割の誤り率を表す評価指標である。Pkでは、距離kだけ離れた2つの発話が同一のトピックに属しているかどうかをシステム出力結果と正解データの両方で計算する。そして、両者の一致しない割合がPkのスコアとなる。それに対し、Hearstら[11]はPkのFalse Positiveに対するペナルティが少なく計算されてしまうという問題点を指摘し、新たな評価指標WidowDiff(WD)を提案した。WindowDiffでは、距離kだけ離れた2つの発話間にいくつのトピック境界が含まれるかを計算し、その不一致数をスコアとする。これによりFalse Positiveに対するペナルティの問題は解決された。

6.2 実験結果

6.2.1 キーワード追加の効果

まず始めに、市川らのキーワードセットに新たにキーワードを追加したことの効果を調べた。用いた手法は以下の2種類である。

- 市川手法：市川らが用いたキーワードセットを用いた手法
- 提案手法1：市川手法に新たにキーワードを追加した手法

表 2: キーワード検出パターン追加による効果

	Rec.	Prec.	F 値	Pk	WD
市川手法	24.75	21.55	23.04	34.68	38.55
提案手法1	47.52	20.87	29.00	35.62	51.97

その結果を表2に示す。トピック境界以外でも使われる可能性が高い語を追加したため、Precisionが若干悪化してしまったものの、Recallは20%以上改善し、F値で約6%の改善が見られた。その一方、トピック分割での評価指標Pk及びWindowDiffでは、両者とも悪化し、特にWindowDiffでは13%の悪化が見られた。その原因として、キーワード追加によるFalse Positiveの増加が考えられる。

6.2.2 語彙的連鎖によるフィルタリングの効果

次に、語彙的連鎖によるフィルタリングの効果を調べた。その際に用いた3つの手法は以下の通りである。

- 提案手法2：提案手法1に加え、明示的なキーワードによって検出されたものみにフィルタリングを行なったもの
- 提案手法3：提案手法1に加え、対話主導権の交代によって検出されたものみにフィルタリングを行なったもの
- 提案手法4：提案手法1の全出力結果に対してフィルタリングを行なったもの

その結果を表3に示す。

表 3: 語彙的連鎖によるフィルタリングの効果

	Rec.	Prec.	F 値	Pk	WD
提案手法2	42.57	20.67	27.83	38.39	52.71
提案手法3	45.54	25.84	32.97	30.66	43.58
提案手法4	40.59	26.45	32.03	32.55	43.76

まず始めに、手法1と手法2を比較し、明示的なキーワードに対するフィルタリングの効果を調べた。

手法2ではフィルタリングによって検出結果の除去が行なわれたため、Recallは悪化し、Precisionは改善することが予想されたが、結果は両者とも悪化してしまった。Pk、WindowDiffによる評価でも手法2は手法1を下回る結果となり、明示的なキーワードに対するフィルタリングの効果は見られなかった。

次に、手法1と手法3を比較し、対話主導権に対するフィルタリングの効果を調べた。手法3の結果では手法2のときと違い、Precisionの改善がみられ、手法1よりF値で約4%の改善がなされた。Pk及びWindowDiffによる評価でもPkで5%、WindowDiffで8%の改善がみられ、対話主導権に対するフィルタリングは有効であることが確認された。

最後に、手法4の結果から全ての検出結果に対するフィルタリングの効果を調べた。明示的なキーワードに対するフィルタリングが悪影響を及ぼし、手法1と比べると改善されているものの、手法3には及ばない結果となった。

6.3 考察

6.2.2節の結果から、語彙的連鎖によるフィルタリングは対話主導権の交代による検出結果のみに対して行なった方が良いことが分かった。この結果と6.2.1節の市川手法の結果を比較すると、Recall、Precision、F値、Pkの4指標では提案手法3の方が上回る結果が得られた一方、WindowDiffでは市川手法の方が良い結果となった。その原因として、6.2.1節でも述べたように、明示的なトピック遷移のキーワードと対話主導権を表す語の追加によるFalse Positiveの増加が挙げられる。そのため、False Positiveに寛容な評価指標であるPkでは語彙的連鎖によるフィルタリングの効果によって市川らの手法を上回ることができたが、False Positiveにより重くペナルティをかけるWindowDiffでは、フィルタリングによって改善はするもののキーワード追加による悪影響の方が大きく出てしまったと考えられる。この問題への対処策として、主にFalse Positiveが正解境界周辺で多発してしまっていることから、トピック遷移検出後は一定期間(以降、skip数とする)検出は行なわないという方法が考えられる。

6.4 False Positiveの除去

skip数を設定することで、WindowDiff悪化の原因と考えられるFalse Positiveの除去を行なった。skip数を3~40までの間で実験を行い、その結果を図3と図4に示す。図3より、skip=15でPk、WindowDiffと

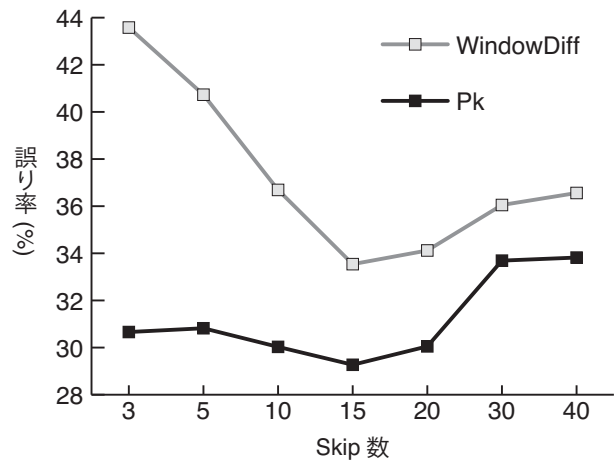


図 3: skip 数による Pk 及び WindowDiff の変化

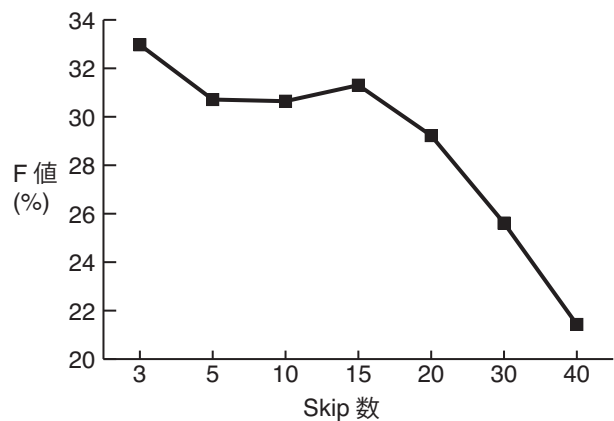


図 4: skip 数による F 値の変化

も誤り率最小となる事が分かる。また、図4よりskip数15まではF値の悪化が少なく、skip数15を境に悪化の度合いが増すことが分かる。以上のことから、トピック境界検出後15発話程度は検出を行なわない方が良い事が分かる。最も結果の良い手法3に対してskip数15を設定したとき(提案手法5とする)の全ての評価指標の結果を表4に示す。skipを設定する前と比べるとRecallが約10%悪化したものの、全ての指標で市川らの手法を上回る結果が得られた。

表 4: skip 数による False Positive の除去

	Rec.	Prec.	F 値	Pk	WD
提案手法 5	35.64	27.91	31.30	29.27	33.54

7 まとめと今後の課題

本研究では、まず始めにトピック分割において有効性が確認されている語彙的連鎖を用いた手法を雑談コーパスに対して行い、語彙的連鎖のみでは雑談のトピック遷移を検出する事は困難である事を示した。そして、市川らが提案した明示的なトピック遷移のキーワードと対話主導権の交代による検出と語彙的連鎖を併用する手法を提案した。その際、明示的なトピック遷移のキーワードと対話主導権を表す語を追加することで Recall の大幅な改善がみられたが、その反面、False Positive の増加を引き起こし、Pk 及び WindowDiff を悪化させた。また、語彙的連鎖によるフィルタリングは明示的なトピック遷移のキーワードによって検出されたものに対しては効果がなく、対話主導権の交代によって検出されたもののみに対して行なう方が良いことが示された。

キーワードの追加によって False Positive が増加し、それによって引き起こされた Pk 及び WindowDiff の悪化は、トピック遷移検出後 15 発話程度検出を行なわないようにすることで改善できる事が示された。

今後の課題として以下の項目が考えられる。

1. 本研究では、トピック遷移を全て同一視し、トピック遷移の種類による分析は行なわなかった。しかし、「全く異なるトピックに遷移する場合」と「現在のトピックの詳細化のために遷移する場合」では、遷移時の状況は異なると考えられるため、トピック遷移の種類による分析が必要となる。
2. False Positive を減少させるために一定期間検出を行なわない方法を行ない、Pk 及び WindowDiff の改善は確認できたものの、それによって Recall が 10% 低下してしまった。そのため、より高度な False Positive 除去方法が必要となる。
3. 本研究ではトピック遷移検出のみを扱ったが、最終目標である雑談を通じた情報提供システムの構築にはトピック遷移先の推定も必要となる。

謝辞

本研究では、プロジェクト「科学研究費補助金・基礎研究 (B)(1) 15320054 平成 15 年度～17 年度『アジアの文化・インターアクション・言語の相互関係に関する実証的・理論的研究』研究代表者 井出祥子 日本女子大学教授」によって作成された「Mister 0 コーパス」を利用させていただきました。

参考文献

- [1] Weizenbaum, J. Eliza - a computer program for the study of natural language communication between man and machine. Communications of the ACM, 1966
- [2] Stede, M. and Schlangen, D. Information Seeking Chat: Dialogue Management by Topic Structure. Proceeding of CATALOG'04 8th Workshop on the Semantics and Pragmatics of Dialogue. Barcelona, Spain, 117-124, 2004
- [3] 市川宙, 徳永健伸. 情報探索雑談における自然なトピック遷移の実現. 言語処理学会第 13 回年次大会, pp.151-154, 2007
- [4] Freddy Y. Y. Choi, Advances in domain independent linear text segmentation. ACM International Conference Proceeding Series, 2000
- [5] Michel Galley, Kathleen Mckeown, Eric Fosler-Lussier, Honbyan Jing. Discourse Segmentation of Multi-Party Conversation. Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1, 562-569, 2003
- [6] Pei-Yun Hsueh, Johanna D. Moore. Combining Multiple Knowledge Sources for Dialogue Segmentation in Multimedia Archives. Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, 1016-1023, 2007
- [7] Margot Mieskes and Michael Strube. Parameters for Boundary Detection in Multi Party Dialogues. LREC2008, 2008
- [8] Maria Georgescu, Alexander Clark, Susan Armstrong. Word Distributions for Thematic Segmentation in a Support Vector Machine Approach. LREC2008, 2008
- [9] 望月源. 語彙的連鎖計算プログラム Lexical Chainers. <http://www.jaist.ac.jp/~motizuki/software/chainers/>
- [10] Doug Beeferman, Adam Berger and Jhon Lafferty. Statistical Models for Text Segmentation. Machine Learning 34, 117-210, 1999
- [11] Lev Pevzner, Marti A. Hearst. A Critique and Improvement of an Evaluation Metric for Text Segmentation. Computational Linguistics, 2002