博 士 論 文

ユーザの発話理解の精緻化によって
円滑な雑談を実現する自由対話システム

指導教官　白井 清昭 准教授

北陸先端科学技術大学院大学
情報科学研究科情報システム学専攻

福岡 知隆

2017 年 3 月 24 日

Abstract

In recent years, the study of non-task-oriented dialog system that can freely talk with users attracts more research interests, since it can be used for a chat robot or user friendly interface. To appropriately respond to user's utterance, it is important to precisely recognize user's intention. This thesis proposes fundamental techniques to realize smooth conversation between the user and dialog system.

First, a novel method to identify a dialog act, which represents the user's intension, for a given utterance is proposed. In previous studies based on supervised machine learning, a unique set of features is used for classification of all dialog acts. However, not all features may be effective to identify all dialog acts. Some features may be effective for classification of a particular dialog act only, and such features may cause errors for classification of other dialog acts. In the proposed method, an optimized set of the features is determined for each dialog act. Then, binary classifiers are trained with the optimized features for individual dialog acts. Finally, one dialog act is chosen from the results of these classifiers by the several sophisticated methods. The experimental results showed that proposed method significantly improved the F-measure by 0.6% over a baseline that was trained with the unique feature set.

Second it is important to consider a timing of changing the topic in order to continue chat with the user. If the user shows the sympathy for the current topic, the system should continue the conversation with the same topic. On the other hand, if the user does not display the sympathy, the system should provide other topics. This thesis proposes a method to identify if a speaker displays sympathy in his/her utterance. The method is based on supervised machine learning. New features are proposed to train a classifier for identifying the sympathy in user's utterance. A problem for supervised learning of sympathy identification is that a number of positive samples is much fewer than that of negative samples in general, i.e. the sympathetic utterance does not frequently appear in free conversation. To tackle this problem, a filtering process to remove the redundant negative samples is introduced to correct imbalance of the training data. The results of the experiments showed that the proposed features improved the F-measure by 3-4% over a baseline and the filtering of negative samples was effective to improve performance.

Third, an initiative of the conversation is often altered among the user and dialogue system in natural free conversation. To take the initiative, the system is required to produce a sequence of consistent utterance of the same topic. In this study, an anecdote of a person is regarded as a sequence of consistent utterance that can be provided by the dialog system, because the anecdote might be a good topic that attracts user's interest and has suitable length. Therefore, this thesis proposes a method to retrieve the anecdotes of the given person from Web. First, passages that are related to the given person are retrieved by searching relevant web pages with the query "[person] & anecdote" and segmenting the obtained pages based on the analysis of Document Object Model (DOM) trees of them. Then, each passage is judged whether it is the anecdote by several rules based on linguistic features of the anecdote. The experimental results showed that the precision of the proposed method was increased by 11% comparing to a baseline with a little loss of the recall. However, it was also found that deep understanding of the text would be required to precisely filter out non-anecdote passages.

Keywords: Non-task-oriented dialogue system, Supervised machine learning, Dialog act, Sympathy, Initiative of dialog, Anecdote

要 旨

近年は人間とコンピュータが自然言語を用いて対話を行う対話システムの研究が盛んに行われている．特に，タスクや対話の内容を限定しない自由対話システムは，人間と雑談できるロボットやユーザフレンドリーなインターフェースを担う機能として注目を集めている．自由対話システムがユーザの発話に対して適切に応答するためには，ユーザの発話意図の正確な認識が特に重要である．本稿では，人間と対話システムとの円滑な雑談を実現するための要素技術の研究に取り組む．

まず，ユーザの発話に対し，その発話意図 (対話行為) を自動推定する新しい手法を提案する．従来の教師あり機械学習に基づく手法では，対話行為を推定するための素性のセットを 1 つだけ設定する．しかし，全ての素性が全ての対話行為の分類に有効というわけではなく，特定の対話行為の分類のみに貢献する素性は，他の対話行為の分類の精度を低下させる要因となりうる．本研究では，対話行為毎に最適な素性のセットを推定し，その素性を基に個々の対話行為毎に分類器を学習する．さらに，個々の対話行為毎の分類器の出力の結果から，最適な対話行為をひとつ選択するいくつかの手法を提案する．評価実験の結果，提案手法の対話行為分類の F 値は 0.825 となり，素性のセットを 1 つだけ用いる従来手法と比べて有意に 0.6 ポイント高いことを確認した．

雑談を長続きさせるためには，ユーザが現在の話題に共感を示しているときはその話題を継続し，共感を示していないときは話題を変える戦略が有効である．本研究では，ユーザの発話が共感を示しているか否かを推測する手法を提案する．実際に対話例を分析し，ユーザの共感を示唆する素性タイプをいくつか提案し，それを用いて教師あり機械学習により共感の有無を判定する分類器を学習する．また，実際の発話では共感を示す発話がそうでない発話と比べて数が少ない，すなわち機械学習の訓練データにおいて正例の数が負例の数よりも少ないという問題に対し，冗長な負例をフィルタリングすることで正例と負例の偏りを是正する手法を提案する．評価実験の結果，本研究で提案する素性を用いた推定結果の F 値は 0.18 となった．この値はベースラインよりも 3 ポイント高く，提案手法が共感推定システムの性能向上に貢献することを確認した．また，負例のフィルタリング手法が有効であることも確認した．

自然な雑談を実現するためには，ユーザとシステムが交互に対話を主導する必要があり，システムが対話を主導するときは同じ話題に関する一連の発話を順次生成することが求められる．本研究では，そのような一連の発話として人物の逸話に注目する．逸話はユーザの興味を引く内容を含むことが多く，かつある程度の長さを持つことから，システム主導の対話におけるトピックとして適していると考えられるためである．提案手法では，人物名と「逸話」というキーワードを含むウェブページの DOM ツリーを解析し，それをパッセージに分割した後，逸話ではないパッセージを除外するルールをいくつか提案し，そのルールを用いたフィルタリング処理によって逸話を獲得する．評価実験により，ベースラインの逸話抽出の精度が 23.2%であったのに対し，提案手法の精度は 34.8%となり，およそ 11 ポイント向上したことを確認した一方，逸話とそうでないパッセージを正確に識別するためにはより深い意味解析が必要であることも明らかにした．

目 次

第 1 章

序論

対話は人間同士がコミュニケーションをとるための代表的な手段である．一方的に情報を発信，受信する場合と異なり，対話では複数の参加者がお互いに情報を送受信するために，相手の意図や心的状態を正確に理解することができ，結果として円滑なコミュニケーションが実現できることが多い．例えば，学習における教師と学習者の言語を通じた相互作用は，対話による意志疎通が最も効果的に働く例といえる．このように対話は人間の知的・創造的活動には欠かせないものである．

近年は人間とコンピュータが自然言語を用いて対話を行う対話システムの研究が盛んに行われている．対話システムには大きく分けて 2 つの種類がある．特定の場面で使用することを想定し，対話の内容をあらかじめ限定したタスク指向型対話システムと，対話内容を限定しない非タスク指向型対話システムである．

タスク指向型の対話システムとして，観光の案内 [21] やスマートフォンの利用方法の説明[1]など，様々な対話システムが開発され，我々は実際に利用することができる．これらのタスク指向型のシステムでは，対話の目的があらかじめ決まっており，ユーザからの入力やそれに対する応答の内容や表現は，対話の目的を達成するために必要なもののみに限定されている．そのため，システムが応答文を生成する手法は，あらかじめ入力文と出力文の組を用意して，人間の入力文に対するパターンマッチによって応答文を生成するルールベース型 [49] が一般的である．

一方，非タスク指向型対話システムでは対話内容はあらかじめ設定されていない．すなわち，対話システムは様々な内容の対話を実現する．例えば，雑談，質問，相談，説明などが例として挙げられる．また，前述のタスク指向型対話システムが取り扱うような目的が明確に定められている対話も含まれる．非タスク指向型の対話システムもまたすでにいくつも開発されており，我々は気軽に利用することができる．例えば,「りんな」はマイクロソフト社が開発した対話 AI(Artificial Intelligence) であり，ウェブ上で公開されている[2]．非タスク指向型対話システムに

[1] http://www.apple.com/jp/ios/siri/
[2] http://rinna.jp/

おいても，ルールベースの手法は自然な応答文を生成することができるが，膨大な入力文と応答文の組をルールとして用意しなければならない．ルールベースの手法においては，応答文は入力文が対話システムが保持しているルールに合致した場合にのみ正しく生成される．想定していない入力に対してはうまく応答文を生成できないことがルールベースの手法の欠点である．一方，非タスク指向型の対話におけるユーザからの入力は，様々なトピックが含まれるため，想定していない入力文が存在しないようにするためには膨大なルールを用意する必要がある．そのため，ルールベース型の非タスク指向型対話システムを構築するのは非常にコストがかかる．また，新しいトピックは日々生まれているため，すべてのトピックに対するルールをあらかじめ用意することは困難である．そのため，非タスク指向型の対話システムでは統計的手法による応答文生成が一般的である [9].

統計的手法による応答文生成は，発話とそれに対する応答の組となる実際の対話データや新聞記事などのテキストデータを大量に用意し，入力となる発話に対する適切な応答文をテキストデータから検索する手法である．この際，テキストデータに存在する応答文の候補に対するスコアは様々な統計的手法で計算され，そのスコアが最大の応答文が選択される．統計的手法ならば，入力に対する応答文を人間が手作業で設定する必要はないので，様々なトピックを含む大規模コーパスを用意すれば，非タスク指向対話におけるユーザからの入力にも対応可能である．近年ではインターネットの普及により，ウェブから大規模なコーパスを容易に取得することが可能であり，また Twitter[3]などのマイクロブログを用いたコミュニケーションの流行もあり，新聞記事などよりも砕けた表現のテキストも容易に得ることが出来る．一方で，統計的手法を用いて生成された応答文は，ルールベースで生成された応答文と比べて質が悪い場合が多い．これまでの統計的手法では文脈や発話文の内容が十分に考慮できていないため，的外れな応答や意味不明な応答を生成する場合が，ルールベースによる応答文生成よりも多い．

従来の対話システムの研究はタスク指向型対話システムが中心であったが，現在は非タスク指向型の対話システムの研究も盛んに行われている．非タスク指向型対話システムは，自由対話システムとも呼ばれている．本論文では自由対話システムを研究の対象とする．

自由対話システムが実現する対話とは主に雑談である．対話システムが雑談を実現できることの主な利点は以下の通りである．

ユーザとの親密度の向上

人間同士が円滑に関係を築く上で，雑談は大きな役割を果たしている [1]．これは人間とコンピュータの対話においても同様であると考えられる．これまでの対話システムの多くはユーザとの一度きりの対話を行うだけであったが，近年はユーザの個人情報を利用した対話システムも研究されている [27][14]．特定のユーザにパーソナライズされた対話システムとの対話は，一度きりで

[3]https://twitter.com/

はなく何度も対話を行うことが想定され，これによりユーザとの間に親密な関係を築くことも可能である．

目的を持った対話の中での雑談

対話システムが雑談を行うことは，タスク指向の対話システムにおいても重要である．なぜなら，ユーザがタスク指向型対話システムと対話をする場合でも，タスクとは関係のない内容の発話を発っすることがあり，ときにはタスクを達成することを目的としない対話が行われる可能性があるためである．こういった雑談はタスク達成のために回避するべきものではなく，むしろ円滑にタスクを達成するためには必要なものである [48]．対話システムがこういった雑談をこなすことができなければ，タスク指向の対話であったとしても継続することが難しくなる．

本研究が想定する自由対話システムの構成を図 1.1 に示す．対話システムは，入力処理部，出力処理部，対話制御部の 3 つのサブシステムから構成されている．

- 入力処理部は，ユーザの発話を入力として受けとり，それを解析するサブシステムである．まず，ユーザの発話に対して，音声認識，形態素解析，構文解析，述語項構造解析を行い，その意味内容を理解すると同時に，発話に含まれる言語的特徴を抽出する．次に，その結果を基に，対話行為推定，話題推定，共感推定，ユーザ状態推定などの処理を行う．対話行為推定とは，発話の対話行為 (質問，応答，あいづちなどの発話のタイプ) を推定する処理である．話題推定とは，発話が入力された時点における最も主要な話題は何かを推定する処理である．共感推定は，ユーザがシステムに対して共感を示しているかを推定する処理である．この処理ではユーザ発話の対話行為を手がかりとするため，対話行為推定結果を用いる．ユーザ状態推定とは，ユーザの喜び，悲しみ，怒りなどの感情や，共感推定の結果を手がかりとし，現在の対話の話題に興味を示しているか，または退屈しているかを推定する処理である．

- 出力処理部は，ユーザへの応答文を生成するサブシステムである．応答文を生成するいくつかのモジュールをあらかじめ用意しておく．QA モジュールは，ユーザの発話が質問であったとき，それに対する回答を対話システムが持つデータベースやウェブなどの外部知識から検索し，応答文として生成する．詳述モジュールは，現在の話題に関する新しい事実を説明する発話を生成する．話題提供モジュールは，新しい話題をユーザに提供するような発話を生成する．あいづちモジュールは，適当なあいづちを生成し，ユーザの発話を促す．自己紹介モジュールは，対話の冒頭に対話システムの自己紹介をする働きをする．対話システムがどのような人物を模しているかはあらかじめ設定しておく．逸話紹介モジュールは，現在の話題が特定の人物のとき，その人物に関する逸話を紹介するモジュールである．このモジュールでは，

対話システムが複数の発話を続けて生成することで，対話を主導することを狙う．

- 対話制御部は，入力処理部の解析結果を受けとり，出力処理部におけるどのモジュールを用いて応答文を生成するかを決定するサブシステムである．対話制御部は話題を転換するか否かを決定する機能も有する．ユーザが現在の話題に共感を示しているときは現在の話題を継続するための発話を，逆に飽きているときは新しい話題を提供する発話を生成する．

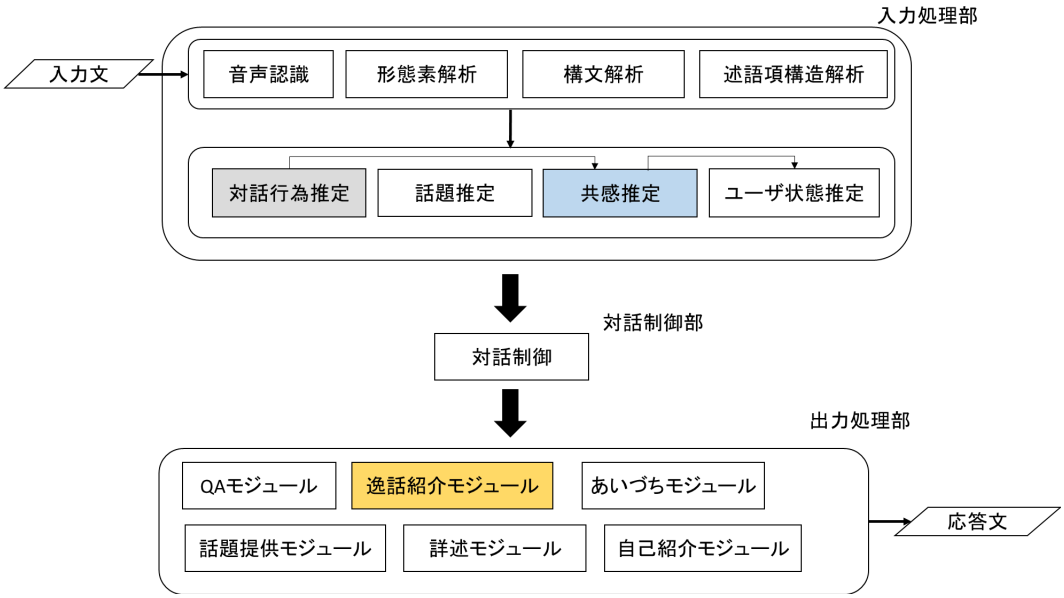

図 1.1: 対話システムの構成

図 1.1 に示した自由対話システムを実現するためには多くの解決するべき課題がある．本研究では，この中で特に重要と考えられる 3 つの研究課題に取り組む．

1. 対話行為の自動推定
   入力処理部における発話の対話行為の推定は，適切な出力モジュールを選択し，自然な応答を生成するために必要不可欠な技術である．本論文では対話行為推定の性能を向上させる課題に取り組む．詳細は 1.1 節で述べる．

2. 共感の自動推定
   入力処理部における共感推定は，適切な話題転換を行うために重要である．本論文では発話の共感を推定するために有効な特徴は何かを探究する課題に取り組む．詳細は 1.2 節で述べる．

3. 逸話の自動獲得
   逸話紹介モジュールでは，人物とその逸話のデータベースをあらかじめ用意

しておくことを想定している．本論文ではウェブから人物の逸話を自動獲得するという新しい課題に取り組む．詳細は 1.3 節で述べる．

1.1 対話行為の自動推定

対話行為とは，発話をその目的や意図などに基づいて分類した発話のタイプである．自由対話システムでは対話行為の情報が利用されることが多い [10]．なぜならば，対話行為はユーザの目的や意図の種類とも言い換えることができ，対話行為を推定することはユーザの発話意図を推察することに他ならないからである．対話システムはユーザの発話の対話行為を推定し，ユーザがその発言によって何を意図しているか考慮した上で応答文を生成すれば，対話を続けやすい．逆にユーザの意図を無視した応答を生成すれば，例えば，ユーザが何か質問をしているときに新しい話題を始めたり，ユーザが単にあいづちをしているのに対話システムもあいづちを返すなど，不適切な応答をすれば，ユーザは不快感を抱き，対話が破綻することが予想される．

ユーザの意図を正しく汲むため，対話システムがユーザ発話から対話行為を正確に推定することは重要である．例えば，対話システムが対話行為「質問」である入力を受け付けたならば，対話システムは質問に対する「応答」となる応答文を生成するべきである．人間同士の対話では質問に対して応答を返すということは自然な流れであるが，このような入力に対する応答のパターンは，対話システムが適切な応答文を生成するための重要な知識である．自由対話システムは様々な入力に対応するため，様々な応答文を生成できなければならない．対話行為に応じて入力に対する応答生成のモジュールをあらかじめ用意することで，様々な発話に対して適切な応答を生成できる．対話行為に応じて適切な応答文モジュールを選択するアプローチでは，ユーザの発話の対話行為を推定することが前提となっている．そのため，対話行為を自動的に推定する技術は重要である．

対話行為を自動的に推定するために，教師あり機械学習がよく用いられる [33, 17, 42, 24, 31]．また，対話行為を推定するための素性も様々なものが提案されている．しかし，機械学習に用いる素性を設定する際，個々の対話行為の特徴が十分に考慮されていないという問題点がある．既存研究の多くは，対話行為の自動推定を多値分類問題と捉え，対話行為の分類に有効と思われる素性のセットを 1 つ設定する．しかし，素性の中には，ある特定の対話行為の分類にしか有効に働かないものもある．例えば，ユーザの発話の対話行為が (質問に対する)「応答」であるかを判定するためには，発話者が交替したかという素性は重要だが，対話行為が「質問」であるかを判定するためには，相手の発話の後に質問することもあれば自身の発話に続けて質問することもあるので，話者交替は重要な素性とは考え難い．ある素性が特定の対話行為の推定に有効に働かないだけでなく，その対話行為の推定精度を低下させることもある．このため，推定精度が高い対話行為と低い対話行為が存在する．推定精度が高い対話行為の発話が入力されたときと

そうでないときでは，後者の方が適切な応答を生成でない可能性が高い．すなわち，ユーザの発話の対話行為によって応答の質に差が生じてしまう．これは自由対話システムとして望ましくない．

本研究では，この問題に対し，機械学習による対話行為推定を 1 つの素性セットを用いて多値分類問題として解くのではなく，それぞれの対話行為の特徴を考慮した素性セットを用いて，対話行為毎の二値分類問題として解き，その結果を基に最適な対話行為を選択する手法を提案する．機械学習に用いる素性セットをそれぞれの対話行為の分類に最適化することで，個々の対話行為の推定精度を向上させ，ひいては全体的な推定精度の向上を図る．

1.2 共感の自動推定

共感は人間とコンピュータが自由対話を実現する上で考慮しなければならない重要な要素である [29]．対話システムの実装において，共感は対話行為の一つとして扱われることも多いが，本研究では独立した要素として扱う．なぜならば，共感は人が他人と円滑に人間関係を築くための基本的かつ重要な能力であると考えられるからである [7]．例えば，自らの意見や思想を相手に伝えた際，相手が共感しているか否かを察する必要がある．相手が共感していないならば，理解しやすいように話し方や言いまわしを変える．あるいは，この相手から到底共感を得られないと思えば，相手に自分の考えを理解してもらうのをあきらめる．また，相手の意見や思想に対して共感を示すことで，相手の話を促したり，より親密な関係を築くことも可能である．

共感は，人と人の関係構築の助けとなるだけでなく，自由対話システムが人との雑談を継続するためにも必要である．本研究では，発話者の共感を検出することは，自由対話中における話題の転換のタイミングを計る上で，重要な手かがりになるものと考える．自由対話では話題は固定されておらず，任意のタイミングで変更することが可能である．しかし，話題を変更するタイミングはいつでも良いわけではない．相手がまだその話題について話をしたい時に話題を変更したり，またはその話題についてはこれ以上対話を続けたくないにも関わらず対話を続けることは，相手に不快感を与え，場合によっては対話を打ち切られる可能性がある．この話題転換のタイミングを計る要素の 1 つが話者の共感である．話し相手の共感が得られているならば現在の話題を継続し，逆に共感が得られていないならば異なる話題を提供することが，自由対話を継続させるために重要である．

人間の発話から共感を検出する試みはこれまでもなされてきた．しかし，前述のように，多くの場合，共感は対話行為の一種として扱われている．つまり，共感は対話行為を分類する際のクラスの一つとして扱われている．したがって，共感の検出に関する研究は十分になされているとはいえない．人間が対話中に相手の共感を得られているかを判断する手がかりとしては，相手の表情，しぐさ，発話の内容などが考えられる．これらのうち，本研究では発話の内容，具体的には発

話テキストに着目する．すなわち，相手の表情や音韻情報は手がかりとせず，テキストとして入力された発話が共感を示しているかを判定するモデルを機械学習する手法を提案する．特に，テキストから得られる情報のうち，何が共感推定に有効な素性となるかを明らかにする．テキストのみから共感の有無を判定する技術は，テキストベースのチャットシステムを実装するときに特に有効である．

1.3 逸話の自動獲得

ユーザを退屈させない応答文とは，ユーザの興味を引く応答文と考えられる．対話システムが，ユーザが興味を持っているトピックに関連した話題を応答文として生成すれば，ユーザとの対話を継続しやすい．対話システムがユーザの興味を自動的に知る手法は，あらかじめ取得したユーザのプロファイルや検索エンジンの履歴などのパーソナルデータから推測する手法 [46][13] と，ユーザとの対話中にユーザの発話から推測する手法 [22][9] の 2 つに大別される．前者の手法によりユーザの興味をあらかじめ把握しておけば，対話システムが対話の開始時からユーザの興味を引くことも可能である．また，後者の手法により，対話システムは対話中のトピック変化に柔軟に対応することが可能である．

また，自由対話は常に一人の発話者が対話を主導するわけではない．対話の話題は刻々と変化し，話題によって対話を主導する人物も変化する [23]．よって，自由対話システムの実現のためには，ユーザからの入力に答えるだけでなく，システムからも様々な話題を提供し，時には対話を主導することも重要である．

自由対話では様々な話題に対する応答が求められるため，自由対話システムの応答文生成手法では，あらかじめ入出力のパターンを決めておくルールベースの手法よりも統計的手法が用いられることが多い．応答文生成のための言語資源として，ウェブ上のニュース記事や Twitter などを利用する方法がすでにいくつか提案されている [38, 43, 10]．これらの手法では，システムは様々な話題の発話を様々な言い回しで生成することが可能である．ところが，先行研究の多くは 1 つの応答文が生成されるが，システムが対話を主導するためには，1 つではなく一貫性のある複数の発話を連続的に発することが求められる．一方，システムが対話を主導するにあたり，対話システム自身が展開する発話の構成を自動的に構築することは困難である．

この問題の解決方法として，あるイベントに対する詳述内容をあらかじめ用意し，それを連続した応答文として生成する手法が考えられる．話題として提供する内容に対して，システムが予めコーパスから応答文のまとまりを取得しておき，それらを順次生成する．例えば，ある年のサッカーワールドカップについて，どの国が優勝し，どの選手が優れていたかなどをシステムが出力し，一方ユーザがその話題に興味を持ち，システムの発話の合間にあいづちをいれれば，システム主導の対話が実現できる．

本研究では，自由対話システムが提供する話題として，人物の逸話に着目する．人物の逸話は，ある程度の文の集まりからなる．対話システムがユーザに対して逸話を紹介することで，内容に一貫性を持ったまま，複数の発話を連続して発することが可能であると考えられる．また，逸話のテキストはウェブ上にも存在するため，逸話データベースの構築を自動化することも可能である．本論文では，逸話データベースを構築するために，ウェブから人物の逸話を自動獲得する手法を提案する．ウェブから様々な知識を獲得する先行研究が存在するが，人物の逸話の獲得はこれまでにない新しい試みである．

1.4 論文の構成

本論文の構成は以下の通りである．第 2 章では本研究の関連研究について述べる．第 3 章では対話行為推定手法について述べる．第 4 章ではユーザの発話の共感の有無を推定する手法について述べる．第 5 章ではウェブから逸話を抽出する手法について述べる．最後に第 6 章で結論を述べる．

第 2 章

関連研究

本章では，本論文の関連研究について論じる．2.1 節では，対話システムの過去の研究事例を概観する．次に，1 章で述べた本論文の 3 つの研究課題に関する過去の研究について述べる．2.2 節では，対話行為の自動分類に関する先行研究について論じる．2.3 節では，対話における共感を対象とした研究について述べる．最後に，2.4 節では，自由対話システムが生成する応答文をウェブなどの外部リソースから自動収集する研究を紹介する．

2.1　対話システム

これまで数多くの対話システムが作成されており，音声や画像によってユーザとコミュニケーションを取るシステム [26][36][35][51] も存在する．本研究で研究対象とする対話はトピックを設定しない非タスク指向型のものであるため，以降ではタスク指向型の対話システムは対象とせず，非タスク指向型の対話システムのみについて論じる．

一般的な対話システムの処理の流れは以下の通りである．ユーザから発話が入力されると，形態素解析，構文解析，意味解析によってその内容を理解する．また，多くの対話システムでは発話の対話行為を推定する．次に，対話マネージャーと呼ばれるモジュールで，ユーザの発話理解の結果，対話行為の推定結果，過去の対話の記録などを基にシステムが生成する発話の内容を決める．さらに，その発話の内容を表わす文を生成し，これをシステムの応答文として出力する．ここでは，応答文の生成方法によって過去に研究された対話システムを分類する．

ELIZA[49] は初期の対話システムとして有名である．応答文の生成のための入力と出力のペアをあらかじめ用意するルールベース型の対話システムである．ユーザの入力から応答文生成のルールに合致するキーワードがある場合，あらかじめ決められた応答文を出力する．ELIZA は診療療法士によるセラピーを模してユーザと対話することがあり，このときには対話の内容には制限が設けられない．そのためユーザから多様な発話が入力されるが，ルールから外れた入力にはうまく対

応できない．また，ルールベース型ではルールを改善することで大幅な性能の向上を図ることは難しく [11]，未知のトピックへ対応するためにはコストがかかる．

あらかじめ入力文と応答文の組を用意するのではなく，コーパスを用意し，そこから統計的に応答文を出力する手法もある．吉野らは，質問への回答にウェブ検索の結果を利用する質問応答対話システムを作成した [53]．この対話システムでは，応答文を生成するためのコーパスとしてウェブを利用している．

樋口らが作成した対話システム [12] では，ユーザの入力文から重要と思われるキーワードを取り出し，その連想概念のテキストをウェブから抽出し，応答文としている．

大西らが作成した対話システム [40] では，対話行為の推定や，対話の話題を認識することで，ユーザの発話の意図や文脈を理解した上での応答文を生成している．

東中らが作成した対話システム [9] では，応答文の生成手法を一つに限らず，複数の応答文生成モジュールを実装している．それらの中にはルールベース型や Twitter から応答文を抽出する統計的手法などの様々な応答文生成手法を採用している．

近年の対話システムでは抽象的な情報を扱う処理が研究されている．例えば，発話テキストを解析する際に，構文的，意味的な情報だけではなく，ユーザの状態も解析の対象とする場合がある．水上らは，ユーザの快適度を推定し，その推定結果に応じて適切な用例を選択して応答文を生成する対話システムを提案している [37]．ここでの快適度とは，Yang らの研究 [54] に従い,「システムの応答をどの程度快適であると感じたか」を 1 から 6 までの整数で表わした指標である．作成した対話システムでは，発話や応答文における単語 n-gram や単語共起の情報を手がかりにユーザの快適度を推定し，快適度が高くなるように応答文を生成する．例えば，ユーザ発話「小腹がすいたなー」に対するシステムの応答文として,「何か食べる？」は快適度が高い発話である．大竹らは，高齢者との対話に適した応答を生成するため，発話意図を考慮し，相手の発話意欲を促す対話システムを作成した [41]．対話システムの作成を前提としているわけではないが，Hasegawa らは，Twitter から収集した大規模な対話コーパスを用いて，発話から発話者の感情を推定する手法と，発話者の感情を誘導する発話の生成手法を提案した [8]．

2.2 対話行為推定

対話を形成する上で，話者の対話行為は対話の展開に強い影響を与えるだけでなく，対人印象や対人関係にも影響を及ぼしている [39]．自由対話においては，対話を継続するか否かの判断はユーザに委ねられており，対話の内容のみならず，対話システムの不自然な応答や不快な発話は対話の破綻に繋がる．自由対話を継続するには，ユーザと対話システムが良好な関係を築く必要があり，そのためには，話者の対話行為を正確に推測し，それに応じて適切な応答を返さなければならない．したがって，ユーザの発話の対話行為の推定は自由対話システムにおける重要な要素技術である．

自由対話システムでは，対話行為は，ユーザ意図の理解，システムの応答文生成における条件づけ，システムの対話制御などに利用されている．

東中らが作成した対話システム [9] では，対話行為推定は意図理解のモジュールの一つとして実装されている．また，対話システムが応答文を生成する際には，システムが生成するべき発話の対話行為を決定し，その対話行為を持つ発話を生成している．

乾らの作成した対話システム [16] では，例えば，質問の対話行為の入力文には応答の対話行為の応答文が自然であるといったパターンを応答文生成時の条件としている．

前田らの研究 [30] では，強化学習により対話制御を学習する手法を提案しており，対話行為はその対話制御の出力結果として用いられている．

また，ユーザの発話を分析し，その対話行為を推定することは，ユーザの意図理解の 1 つとみなせる．ユーザが挨拶をしているか，何かを質問しているのかなどをシステムが理解することで，その後の対話の展開を決定する．南らは行動予測確率に基づく報酬を設定する部分観測マルコフ決定過程 (POMDP) を用いた対話制御手法において，対話行為列の tri-gram による行動予測確率を導入した手法を提案し，その有効性を確認した [34].

また，発話からの情報抽出のためのフィルタリング条件としても対話行為の情報は用いられる．平野らは，ユーザの発話からユーザ情報を抽出する手法を提案し，その手法では発話がユーザ情報を含むか否かを対話行為に基づき判断している [14].

対話行為の自動推定に関する先行研究においては，教師あり機械学習に基づく手法が主流である．この際，基本的な素性として単語 n-gram が利用されることが多い．これに加えて独自の素性も提案されている．

単語 uni-gram は語順を考慮していないため，Milajevs らは単語 bi-gram を素性として用い，単語 uni-gram のみよりも bi-gram を併用したときの方が高い精度が得られることを示した [33]．また，対話の流れを考慮するために前の発話の対話行為を素性として利用し，その効果を評価した．磯村らは，頻度 2 以上の単語 uni-gram と単語 bi-gram，及び 1 つ前の発話の対話行為を素性として，Conditinal Random Field(CRF) を用いて対話行為を推定し，75.77%の推定精度を得たと報告している [17]．機械学習アルゴリズムとして Support Vector Machine(SVM) と Naive Bayes を用いた実験も行ったが，これらでは 1 つ前の発話の対話行為を素性として利用しておらず，推定精度はそれぞれ 66.95%と 60.14%となり，CRF より劣る．関野らは，素性として発話文字数，内容語数，発話順番を提案し，磯村の手法 [17] の素性にこれらを 1 つ以上追加したモデルを評価した [42]．全ての組み合わせにおいてその有効性が確認され，内容語数と発話順番を追加した場合が最も高い精度となった．Kim らは，bag-of-words に加え，対話中の話者の役割などの構造的な特徴と，直前の発話や同一話者によるこれまでの対話行為などといった対話の依存関係を機械学習の素性として提案した [24]．ドメインが限られた対話を評価の対

象としているが，96.86%という高い推定精度が得られている．

目黒らは，多種多様な話題や語彙を含み，また非文法的な文が多いマイクロブログ中の発話に対する対話行為自動付与のため，シソーラスを用いて抽象化した単語 n-gram と文字 n-gram を素性とする手法を提案した [31]．評価実験の結果，Bag-of-Ngrams 素性を用いたベースライン手法よりも高い精度を得た．

これらの先行研究では，機械学習のために用いる素性のセットは 1 つであり，それで全ての対話行為を推定している．しかし，どの素性がどの対話行為の推定に有効に働くかなど，素性と対話行為の関係については議論されていない．本研究では，発話がある対話行為を持つか否かを推定する機械学習において，対話行為それぞれに対して有効な素性を自動的に選択する点に特長がある．

2.3 共感の自動推定

一般に，対話の自動タグ付けには機械学習がよく使われる．一方，共感の有無も対話におけるタグの一種と考えられる．発話の共感を推定する先行研究においても，機械学習を用いた手法が主流となっている．

Bo らは単語の n-gram を素性とし，言語モデル学習ツール SRILM [44] を用いて共感発話の推定を行った [50]．彼らは，bi-gram を素性とする結果が最も精度が高く，正答率が 6 割程度であることを示した．

南らによって提案された，重み付け有限状態トランスデューサーを用いて対話行為を自動認識する手法 [34] では，共感は対話行為の一つのカテゴリとして扱われた．彼らが定義した 6 カテゴリ 29 対話行為の中には,「共感」のカテゴリが定義されており，発話が共感を示す対象を条件として，6 種類のサブカテゴリに細分化されている．彼らが実施した評価では，個々の対話行為の正解率は示されていないため，彼らが提案した手法が共感の推定に有効か否かは不明である．

関野らが提案した，CRF を用いて行った対話行為タグの自動付与 [42] でも共感の推定がなされている．彼らは対話行為タグの定義に SWBD-DAMSL タグセット [19] を用いており，これは sympathy(共感) というタグを含む．しかし，この対話行為の定義は非常に細分化されており，実験に用いた対話コーパスにおいて共感のタグが付与された発話の数は僅かであった．また，彼らが実施した評価では，個々の対話行為の正解率は示されていないため，彼らが提案した素性が共感の推定に有効か否かは不明である．

目黒らが提案したシソーラスを用いて抽象化した単語 n-gram と文字 n-gram を素性とする手法 [31] では，共感は 34 個の対話行為のうちの 1 つとして定義されている．目黒らは，評価実験により，提案手法の対話行為推定の正解率が従来手法に比べて向上したと報告している．しかし，対話行為毎の正解率は示していないため，提案手法が共感推定の正解率向上に有効であるかは不明である．

共感の自動推定に関する先行研究では，様々な素性が提案されている．例えば，日本語では，特定の文末表現が共感の存在を示唆することが報告されており [15][18],

文末表現は共感推定に有効な素性であるといえる．しかしながら，共感は対話行為のひとつとして扱われることが多いこともあり，どのような素性が共感の推定に有効であるかははっきりしない．本論文では，先行研究の知見を踏まえ，与えられた発話が共感を示しているか否かを判定する手法を提案する．提案手法は教師あり機械学習に基づくが，学習のための素性は，対話データの分析によって独自に提案するものを含む．さらに，共感の推定に有効な素性を実験的に調査する．

2.4 ウェブからの応答文の獲得

自由対話システムでは多様なトピックについて対話を行う必要があるため，適切な応答文を生成するのは難しい．人手であらかじめ応答のパターンを用意するアプローチは，応答できる文の数が限られるため，自由対話システムには不向きである．一方，ウェブには大量のテキストが存在することから，適切な応答文をウェブから検索するアプローチが近年注目を集めている．この節では，ウェブ上のテキストを利用した応答文生成手法の先行研究について述べる．

水野らは，ユーザを飽きさせない雑談対話システムの実現のため，ウェブニュースを応答文生成に用いる手法の有用性を評価した [38]．単語の重複，記事内の文の位置，文長を尺度として，システムが応答するのにふさわしいニュース記事をウェブから抽出する．対話システムを実装し，それとユーザとの対話の感性評価を行った．その結果，ニュース記事を利用した発話でも対話をある程度継続させることが可能であることを確認した．

柴田らは，ユーザ発話と表層的結束性および意味的整合性を満足する文をウェブ上のテキストから検索し，対話システムの応答文とする手法を提案した [43]．トピックを映画と限定した会話実験を行い，約 66%の発話について，ユーザが許容できる程度に流れが自然で意味がある応答を返した．

吉野らは，質問への回答にウェブ検索の結果を利用する質問応答対話システムを作成した [53]．質問クエリの述語項構造を取得し，それと部分的に一致するウェブ上の文を検索することで，質問の答えに完全に合致する検索結果が得られないときでも，関連した内容の応答を生成することができる．ドメインを野球に限定した評価実験において，この手法の応答文生成の F 値は Bag-of-Words を用いたベースラインより 17 ポイント高かった．

杉山らは，ユーザ発話内の頻出単語と関連語を用いたテンプレート型の応答文生成手法を提案した [45]．この手法では，関連語は Twitter から収集したコーパス用いて獲得した単語間の依存構造から抽出される．提案手法をチャットボットとして実装し，その評価実験を行った．従来の質問応答対話システムと比べて，テンプレート型で生成した発話はユーザから高い評価を得た．

東中らが実装した自由対話システム [9] では，応答文生成モジュールの一つに Twitter のテキストを利用している [10]．この手法では，Twitter に固有の表現，キー

ワード，統語構造によるフィルタリングと対話のトピックワードとの関連語を条件として，Twitter に投稿された文から対話システムの応答文を選択する．

ウェブ上のテキストを用いた応答文生成手法の多くは，ユーザ主導の対話における応答文をウェブから取得する．ユーザからの 1 つの入力文に応じるための文を生成しており，その文を生成した後にシステムが対話を主導するわけではない．しかし，自由対話では，ときにはシステムが対話を主導することが望ましい．対話システムが対話を主導するためには，一貫性のある発話を順次生成する必要があると考えられる．本研究では，人物の逸話は複数の文からなる場合が多いことに着目し，対話システムが対話を主導する際にユーザに提供する内容に一貫性のあるテキストとして，人物の逸話をウェブから自動的に獲得する手法を提案する．ウェブから人物の逸話の獲得は，これまでに研究されていない新しい研究課題である．

第 3 章

対話行為推定

対話行為は，ユーザの発話意図を推察する重要な手がかりである．対話システムがユーザ発話の対話行為を推定し，ユーザが何を意図しているか考慮した上で応答文を生成すれば，対話が破綻することなく自然な雑談が実現できる．

本章では，自由対話における発話を入力とし，その対話行為を推定する手法を提案する．対話行為の分類クラスをあらかじめ定義し，その中から適切な対話行為のクラスを 1 つ選択する．従来手法の多くは教師あり機械学習に基づくが，学習のための素性タイプのセットはあらかじめ一律に定められている．しかし，全ての素性が全ての対話行為の分類に必要というわけではなく，ある素性が特定の対話行為の分類に貢献しないことがある．そのような素性は正解率を低下させる要因となりうる．この問題を解決するために，提案手法では，対話行為の分類クラス毎に異なる素性タイプのセットを設定する．

提案手法の処理の流れを図 3.1 に示す．対話行為毎に，入力発話がその対話行為に該当するか否かを判定する二値分類器を学習する．その際，対話行為毎に最適な素性タイプのセットを実験的に決める．また，分類と同時に判定の信頼度も算出する．次に，二値分類器による判定の結果，ならびに判定の信頼度を基に，入力発話の対話行為をひとつ選択する．本章では，対話行為を選択するアルゴリズムとして，3.5 節で述べる 4 つの手法を提案する．本手法は複数の分類器を学習し，それぞれの分類結果を統合するという点ではアンサンブル学習におけるバギング法 [2] に類似しているが，それぞれの分類器は特定の対話行為に対して最適化された二値分類器となっており，対話行為に固有の特徴を分類モデルに反映させることが可能である．

本研究では，各対話行為の二値分類器を L2 正則化ロジスティック回帰によって学習し，学習ツールとして LIBLINEAR[6] を用いた．LIBLINEAR の学習パラメタはデフォルト値を用いた．判定の信頼度は LIBLINEAR が出力する確率を用いた．

機械学習アルゴリズムとして L2 正則化ロジスティック回帰を選択した理由は以下の通りである．

- 学習に要する時間が他の機械学習アルゴリズムに比べて短い．後述するよう

に，提案手法では対話行為毎に最適な素性のセットを実験的に決定するため，様々な素性を用いて分類器の学習とテストを繰り返し行う．そのため，SVM のような学習に時間を要するアルゴリズムでは，現実的な時間で対話行為推定システムを構築することができない．

- 提案手法では対話行為毎の二値分類の際，判定の信頼度を算出する必要がある．LIBLENEAR を使えば，その確率を判定の信頼度として利用できる．
- 予備実験では，機械学習アルゴリズムとして線形カーネルを用いた SVM を試したが，その正解率は L2 正則化ロジスティック回帰と大差はなかった．

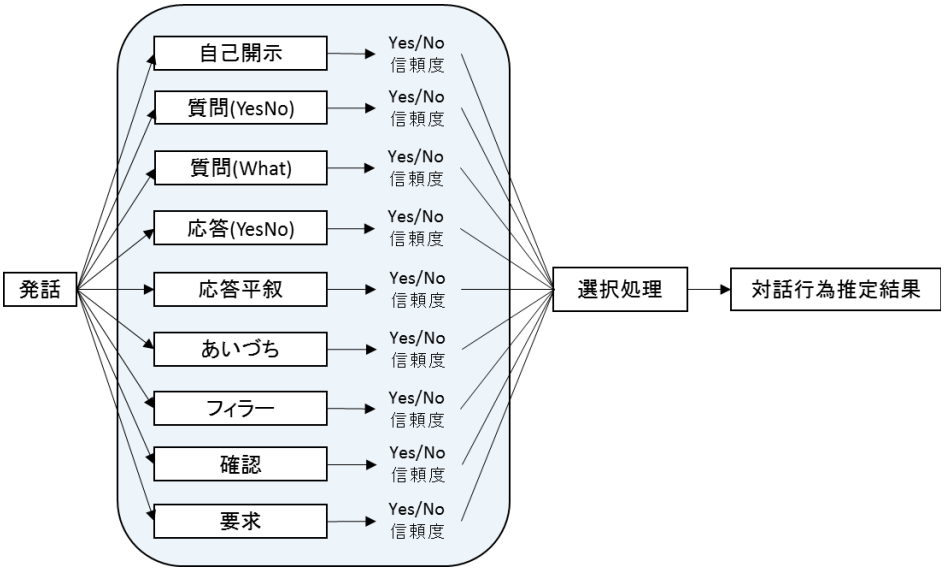

図 3.1: 提案手法の流れ

3.1 対話行為の定義

対話行為の定義としては SWBD-DAMSL[20] が著名だが，かなり詳細な対話行為が定義されており，また自由対話を対象としたものではない．自由対話を想定した対話行為のセット [32] も提案されてはいるが，本研究では，今後構築を目指す自由対話システムの仕様を考慮して独自に定義した 9 つの対話行為のセットを用いる．対話行為の一覧を表 3.1 に示す．また，それぞれの対話行為の発話の例を表 3.2 に示す．

9 つの対話行為のうち,「あいづち」と「フィラー」は類似しており，人間でも区別が難しいと考えられる．また，表 3.2 の例に示したように，対話行為が「確認」の発話は表層的には Yes か No を問う疑問文と似ているため,「質問 (YesNo)」

と「確認」も区別が難しい．しかし，対話システムが適切な応答文を生成するためには，それぞれ別々の対話行為として処理する必要があると考えられる.「あいづち」と「フィラー」の場合，対話システムがあることを詳述しているとき，ユーザがあいづちの発話を発しているのならば内容を理解していると考えられ，そのまま詳述を続けても良いが，フィラーの場合は理解しているとは限らず，あいづちと同様に詳述を継続してはユーザの理解が追いつかない場合が考えられる.「質問 (YesNo)」と「確認」の場合,「確認」は相手が伝聞したことまたは理解したことを確認する発話であるため，対話システムがユーザの理解度を推し量るための指標になると考えられる．例えば，ユーザからの「確認」が頻発するような対話となっているならば，ユーザが対話の内容を理解し切れていない可能性があるため，対話システムは内容を理解しやすいように言い換える必要がある．また，対話システムがあることを詳述しているとき,「質問 (YesNo)」が頻発するならば，対話システムはユーザが詳述内容に興味を持っているものと判断することができる.

表 3.1: 対話行為の定義

d_1 自己開示	発話者の考えや事実を述べる発話
d_2 質問 (YesNo)	相手に対して「はい」「いいね」などの返答を求める質問
d_3 質問 (What)	具体的な内容を問う質問
d_4 応答 (YesNo)	「はい」「いいえ」に相当する短い応答
d_5 応答 (平叙)	質問に対して具体的な内容を返す応答
d_6 あいづち	あいづちを表わす発話
d_7 フィラー	意味を持たないが間をつなぐための発話
d_8 確認	相手が伝聞・理解したことを確認する発話
d_9 要求	相手に対する何らかの要求を表わす発話

表 3.2: 対話行為の発話例

d_1 自己開示	それは楽しみ
d_2 質問 (YesNo)	こっちでやってみる？
d_3 質問 (What)	これ何だろう。
d_4 応答 (YesNo)	うん。
d_5 応答 (平叙)	まあ、それはそのとき。
d_6 あいづち	ふーん。
d_7 フィラー	だからあの。
d_8 確認	平気だよねえ。
d_9 要求	うん、見てみてくださる？

表 3.3: 対話行為推定のための素性タイプ

f_1:単語 n-gram	f_{11}:質問キーワード	f_{21}:話者交代
f_2:前発話の単語 n-gram	f_{12}:質問 (What) キーワード	f_{22}:自立語の有無
f_3:自立語	f_{13}:応答 (YesNo) キーワード	f_{23}:自立語の繰返しの有無
f_4:前発話の自立語	f_{14}:あいづちキーワード	f_{24}:自立語繰返し 1
f_5:文末の単語 n-gram	f_{15}:フィラーキーワード	f_{25}:自立語繰返し 2
f_6:前発話の文末単語 n-gram	f_{16}:文末要求表現	f_{26}:一単語発話 (自立語)
f_7:文末付属語列	f_{17}:文末あいづち表現	f_{27}:一単語発話 (非自立語)
f_8:前発話の文末付属語列	f_{18}:相手の過去の発話の対話行為	f_{28}:発話内単語繰返し
f_9:文末 n-gram ペア	f_{19}:話者の過去の発話の対話行為	
f_{10}:文末付属語列ペア	f_{20}:発話長	

3.2 素性タイプ

本稿では，機械学習において対話行為の識別のために用いる情報を「素性」，素性の種類 (タイプ) のことを「素性タイプ」と呼ぶ[1]．人間同士の自由対話を人手で分析し，対話行為の言語的特徴を考慮して，対話行為の推定に有効と思われる 28 個の素性タイプを設定した．その一覧を表 3.3 に示す．これらは大きく 4 つのグループに分けられる．

グループ 1　f_1〜f_{10} は，発話の内容を表わし，全ての対話行為の分類に有効と考えられる素性タイプである．

f_1：単語 n-gram
　解析対象の発話に含まれる単語の n-gram である．n は 1,2,3 のいずれかとする．この素性タイプを導入することにより，発話の内容と対話行為の関連性が学習できると考えられる．

f_2：前発話の単語 n-gram
　解析対象の前の相手の発話に含まれる単語の n-gram である．n は 1,2,3 のいずれかとする．この素性タイプを導入することにより，相手の発話の内容と対話行為の関連性が学習できると考えられる．

f_3：自立語
　解析対象の発話に含まれる自立語である．f_1 における単語 1-gram と似ているが，自立語に限定することにより，発話の内容を表わす単語を素性として利用できる一方，付属語のような発話の内容を表わさない単語を除外できる．

f_4：前発話の自立語
　解析対象の前の相手の発話に含まれる自立語である．f_2 における単語 1-gram と似ているが，自立語に限定することにより，発話の内容を表わす単語を素

[1]例えば,「単語 3-gram」は素性タイプ,「思い+ます+か」はその素性タイプに属する素性である．

性として利用できる一方，付属語のような発話の内容を表わさない単語を除外できる．

f_5：文末の単語 n-gram
解析対象の発話に含まれる文末の単語の n-gram である．n は 1,2,3 のいずれかとする．文末に出現する表現は対話行為を決める有力な手がかりになると考えられる．

f_6：前発話の文末単語 n-gram
解析対象の前の相手の発話に含まれる文末の単語の n-gram である．n は 1,2,3 のいずれかとする．相手の発話の文末に出現する表現は対話行為を決める有力な手がかりになると考えられる．

f_7：文末付属語列
解析対象の発話に含まれる文末付属語列である．f_5 とは異なり，長さの制限のない付属語の列を素性とすることで，3 つより多くの付属語列から成る文末表現と対話行為の関連性が学習できると考えられる．

f_8：前発話の文末付属語列
解析対象の前の相手の発話に含まれる文末付属語列である．f_6 とは異なり，長さの制限のない付属語の列を素性とすることで，3 つより多くの付属語列から成る相手の発話の文末表現と対話行為の関連性が学習できると考えられる．

f_9：文末 n-gram ペア
解析対象の発話に含まれる文末の単語の n-gram と解析対象の前の相手の発話に含まれる文末の単語の n-gram の組である．n は 1,2,3 のいずれかとする．相手と自身の発話の文末表現の組もまた対話行為を識別するための手がかりとなりうる．

f_{10}：文末付属語列ペア
解析対象の発話に含まれる文末付属語列と解析対象の前の相手の発話に含まれる文末付属語列の組である．この素性タイプを導入した理由は f_9 と同じだが，付属語列のみを素性とする点が異なる．

グループ 2　f_{11}〜f_{17} は，発話の内容を表わし，特定の対話行為の推定に有効に働くと考えられる素性タイプである．

f_{11}：質問キーワード
解析対象の発話が「？」を含むか否かを判定する素性タイプである．

f_{12}：質問 (What) キーワード
解析対象の発話が「だれ」,「誰」,「どこ」,「何で」,「どう」,「どうして」,「いつ」,「何時」,「どちら」を含むか否かを判定する素性タイプである．

f_{13}：応答 (YesNo) キーワード
解析対象の発話が「はい」,「いいえ」,「そう」,「うん」を含むか否かを判定する素性タイプである.

f_{14}：あいづちキーワード
解析対象の発話が「はい」,「うん」を含むか否かを判定する素性タイプである.

f_{15}：フィラーキーワード
解析対象の発話が「はい」,「そう」,「うん」,「うーん」,「あー」,「えー」を含むか否かを判定する素性タイプである.

f_{16}：文末要求表現
要求の発話の文末によく見られる表現である.解析対象の発話が文末が命令形の動詞，動詞基本形+「な」の否定の命令形，動詞連用形+「て」，動詞連用形+「や」，これらの表現+「よ」or「ね」，のいずれかを含むか否かを判定する素性タイプである.

f_{17}：文末あいづち表現
解析対象の発話が文末表現「ね」を含むか否かを判定する素性タイプである.

f_{11}〜f_{17} で使われているキーワードは訓練データを参照して人手で選定した.
グループ 3　f_{18}〜f_{21} は，発話の内容以外の情報を表わし，全ての対話行為の分類に有効と考えられる素性タイプである.

f_{18}：相手の過去の発話の対話行為
人間同士の対話において，質問の発話に対して応答の発話を返すように，前の相手の発話の対話行為は重要であると考えられる．なお，相手の前の発話の対話行為の長さは実験的に決める．詳細は 3.3.2 項で述べる.

f_{19}：話者の過去の発話の対話行為
人間同士の対話において，質問の発話を発してすぐに応答の発話を続けることが起こりにくいように，話者自身がこれまでにどのような対話行為の発話を発してきたかは重要であると考えられる．なお，話者自身の前の発話の対話行為の長さは実験的に決める．詳細は 3.3.2 項で述べる.

f_{20}：発話長
解析対象の発話文中の文字数に基づく発話の長さである．発話長を機械学習の素性として用いる場合，長さを適当な間隔（1 ∼ 5，6 ∼ 10，11 以上，など）に切って発話長を分類するのが一般的であるが，その適切な間隔を決めるのは難しい. 本論文では,「発話長が $l \pm 2$ である」$(3 \leq l \leq 19)$,「発話長が 20 以上である」といった素性で発話長を表現する．例えば，発話長が 10 の発話に対しては，$l = 8, 9, 10, 11, 12$ の素性の重みを 1 とする.

f_{21}：話者交代
現在と直前の発話の話者が同じかどうかを表わす．実験に用いた自由対話コーパスでは，同じ話者が 2 つ以上の発話を連続して発言することがあるため，この素性タイプを導入した．

グループ 4　f_{22}〜f_{28} は，発話の内容以外の情報を表わし，特定の対話行為の推定に有効に働くと考えられる素性タイプである．

f_{22}：自立語の有無
解析対象の発話が自立語を含むか否かを判定する素性タイプである. 自立語を含まなくても生成できる「応答（YesNo）」,「あいづち」,「フィラー」とその他の対話行為の区別に有効であると考えられる．

f_{23}：自立語の繰返しの有無
相手の前の発話の自立語が現在の発話で繰り返し用いられるかを表わす．単語を繰り返して聞き返す「確認」や，反復による「あいづち」の特徴を捉えられる．

f_{24}：自立語繰返し 1
より厳密に「確認」,「あいづち」を示唆する自立語の繰り返しを区別するため，繰り返される自立語が相手の前発話の文末に出現するか否かを表わす素性タイプである．

f_{25}：自立語繰返し 2
より厳密に「確認」,「あいづち」を示唆する自立語の繰り返しを区別するため，発話で繰り返される自立語が現在の発話における唯一の自立語であるかを表わす素性タイプである．

f_{26}：一単語発話 (自立語)
解析対象の発話が自立語 1 語であるか否かを判定する素性タイプである. 自立語 1 語で表現されることの多い対話行為「応答（平叙）」,「あいづち」の識別に有効であると考えられる．

f_{27}：一単語発話 (非自立語)
解析対象の発話が非自立語 1 語であるか否かを判定する素性タイプである. 非自立語 1 語で表現されることの多い対話行為「応答（YesNo）」の識別に有効であると考えられる．

f_{28}：発話内単語繰返し
解析対象の発話が同じ単語を二つ以上含むか否かを判定する素性タイプである. 応答表現やあいづちによく見られる繰り返しによる強調表現の有無を表わす．

対話行為を推定する二値分類器を学習する際には，発話を素性ベクトルで表現する．素性ベクトルの重みは，その素性が発話に出現していれば 1，それ以外は 0 とする．

3.3 素性タイプの最適化

本節では，個々の対話行為毎に，対話行為推定のための素性タイプを最適化する手法について述べる．

3.3.1 最適な素性タイプセットの決定

個々の対話行為に対し，表 3.3 に示した素性タイプの中から，その対話行為の分類に有効でないものを削除することで，対話行為毎に最適な素性タイプのセットを決める．そのアルゴリズムを図 3.2 に示す．E は全素性タイプの集合，E' は最適化された素性タイプの集合である．$F_{dev}(X)$ は，X を素性タイプとして学習した分類器の開発データ[2]における F 値[3]である．素性タイプ f_i を除いたときの F 値 $F_{dev}(E \setminus \{f_i\})$ が全素性タイプを用いたときの F 値 $F_{dev}(E)$ よりも低ければ，f_i を有効な特徴とみなして E' に入れ，そうでなければ削除する．これを全ての素性タイプについて行い，1 つ以上の素性タイプが削除されたら，残された素性タイプを新たに全素性タイプの集合とみなして同様の操作を行う．ただし，個別に評価したときに有効でない素性タイプは E' に残されていないにも関わらず，7 行目の段階で複数の素性タイプが削除された E' を用いたときの F 値がもとの E と比べて低くなることがある．そのときは，素性タイプを削除することによって最も F 値が向上する（最も悪影響を与える）ものを 1 つ選択し，それのみを削除した素性タイプの集合を新たな E とする（10 行目）．これを素性タイプが削除されなくなるまで繰り返す.

学習に用いる素性を対話行為毎に最適化する方法として，提案手法のように素性タイプを取捨選択するのでなく，個々の素性を取捨選択する手法も考えられる．同じ素性タイプの素性でも，特定の対話行為の分類に有効な素性とそうでない素性が存在することが予想される．したがって，素性タイプではなく素性の集合を最適化した方が，対話行為推定の F 値は向上すると考えられる．しかし，素性の数は素性タイプの数 (本研究では 28) よりもはるかに多いため，素性を 1 つ除いて開発データでの F 値の測定を繰り返すのは非常に時間を要する．そのため，本研究では，素性タイプを単位として学習素性のセットの最適化を試みる．

[2]この実験に用いた開発データの詳細は 3.6.1 項で述べる．
[3]発話がある対話行為に該当するか否かを判定する二値分類の F 値.

Input: $E = \{f_1, f_2, \cdots, f_n\}$
Output: E'

while true **do**
  $E' \leftarrow \emptyset$
  for all $f_i \in E$ **do**
    if $F_{dev}(E) \geq F_{dev}(E \setminus \{f_i\})$ **then** $E' \leftarrow E' \cup \{f_i\}$
  end for
  if $E = E'$ **then return** E'
  if $F_{dev}(E') \geq F_{dev}(E)$ **then**
    $E \leftarrow E'$
  else
    $f_x = \arg\max_{f_i} F_{dev}(E \setminus \{f_i\}) - F_{dev}(E)$; $E \leftarrow E \setminus \{f_x\}$
  end if
end while

図 3.2: 素性タイプの選択アルゴリズム

3.3.2 対話行為列の長さの最適化

素性タイプ f_{18} と f_{19} は,「質問 (YesNo)」の次には「応答 (YesNo)」の発話が出現しやすいといったように，対話行為の並びを考慮するために導入した．しかし，直前だけでなく，2 つ以前の発話からの対話の流れが対話行為の推定に有効である場合も考えられる．このとき，どれくらい前の発話を辿ればよいか，つまり過去の発話の対話行為列の長さをいくつに設定すればよいかは，分類対象とする対話行為によって異なると考えられる．

本手法では，素性タイプ f_{18} と f_{19} をそれぞれ相手もしくは話者自身の過去の $N_h(= 1, 2, 3, 4, 5)$ 個の発話の対話行為の列とし，N_h の値を対話行為に応じて最適化する．すなわち，対話行為毎に，開発データでの F 値が最大となる N_h を選択する．また，N_h の値が大きいときには素性数が増えるため，素性選択を行う．具体的には，素性と対話行為の相関の強さを χ^2 値 [55] で測り，それが閾値 T_h よりも小さい素性を削除する．χ^2 値は式 (3.1) で計算される．

$$\chi^2 = \frac{N(o_{11}o_{22} - o_{12}o_{21})^2}{(o_{11}+o_{12})(o_{11}+o_{21})(o_{12}+o_{22})(o_{21}+o_{22})} \tag{3.1}$$

o_{11} は対話行為が d_i でかつ素性 f_i が出現する発話数，o_{12} は対話行為が d_i でかつ素性 f_i が出現しない発話数，o_{21} は対話行為が d_i ではなくかつ素性 f_i が出現する発話数，o_{22} は対話行為が d_i ではなくかつ素性 f_i が出現しない発話数である．表 3.4 は o_{ij} の定義を表としてまとめたものである．T_h は 0,1,5,10 のいずれかとし，N_h と同様に開発データでの F 値が最大となる値を選択することで最適化する．

表 3.4: o_{ij} の対応表

	f_i あり	f_i なし
対話行為 d_i の発話	o_{11}	o_{12}
対話行為 d_i 以外の発話	o_{21}	o_{22}

N_h と T_h の最適化は，3.3.1 で述べた最適な素性タイプセットを決定する前に行う．このとき，素性タイプは f_1(単語 n-gram) と f_{18} もしくは f_{19} のみを使用する．

3.4 組み合わせ素性

本手法で使用する LIBLINEAR では素性間の相関関係は考慮されていない．しかし，素性の組み合わせが対話行為の分類に特に有効に働く可能性がある．そのため，表 3.3 の素性タイプの素性に加え，2 つの素性を組み合わせた素性も使用する．以下，これを「組み合わせ素性」と呼ぶ．ただし，全ての素性タイプの素性を組み合わせると素性数が増大するため，図 3.2 のアルゴリズムにより得られたそれぞれの対話行為に最適な素性タイプセットの F 値と，その素性タイプセットから 1 つの素性タイプを除いた場合の F 値の差が最も大きい素性タイプを「最も有効な素性タイプ」と定義し，最も有効な素性タイプの素性とそれ以外の素性タイプの素性の組のみを組み合わせ素性として導入する．

3.5 対話行為の選択

本節では，個々の対話行為の二値分類器の出力結果から，最も適切な対話行為を 1 つ選択する手法について述べる．

3.5.1 判定の信頼度による選択

対話行為の二値分類器が出力する信頼度を比較し，それが最も高い対話行為を選択する．具体的には，式 (3.2) にしたがって最終的に選択する対話行為 $\hat{d}$ を決定する．$r(d_i)$ は対話行為 d_i の判定の信頼度を表わす．

$$\hat{d} = \arg\max_{d_i} r(d_i) \tag{3.2}$$

3.5.2 信頼度を素性とする機械学習による手法

9 つの対話行為の二値分類器の出力結果を素性とし，対話行為を選択するモデルを機械学習する．当然だが，3.5.1 項で述べた手法において，信頼度 1 位の対話行

為が常に正解となるわけではない．ここでの狙いは,「対話行為 d_a と d_b について，d_a の信頼度が 1 位であるが，d_a と d_b の信頼度の差がそれほど大きくないときは，d_b が正解である可能性が高い」といった傾向を自動的に学習することにある．この手法では以下の学習素性を用いる．

- 対話行為 d_i の判定の信頼度．
- 信頼度の順位が n 位の対話行為の判定の信頼度．$(n = 1, 2, 3)$

これらの素性の重みは信頼度の値とする．後者の素性は，テキスト分類において，他クラスの信頼度を考慮する有効性が高橋らにより報告されている [47] ことから設定した．機械学習アルゴリズムとして L2 正規化ロジスティック回帰 (LIBLINEAR) を用いた．

3.5.3 信頼度に対する重み付けに基づく手法

予備実験の結果,「自己開示」以外の対話行為を持つ発話に対して「自己開示」が誤って選択される事例が多いことがわかった.「自己開示」の信頼度は他の対話行為に比べて平均的に高く,「自己開示」が最終的に選ばれやすいためであった．これは，訓練データにおける「自己開示」の出現頻度が高いためと考えられる．このような信頼度の不均衡を是正するため，式 (3.3) にしたがって対話行為を選択する．

$$\hat{d} = \begin{cases} \arg\max_{d_i} w_i \cdot r(d_i) & \text{if rank(1)=自己開示} \\ \arg\max_{d_i} r(d_i) & \text{if それ以外} \end{cases} \tag{3.3}$$

rank(1) は信頼度の順位が 1 位の対話行為を表わす．w_i は対話行為 d_i の信頼度に与える重みであり,「自己開示」以外の対話行為の信頼度を大きくする働きをする．また,「自己開示」に対する重みは 1 と設定する．

信頼度の重みを反復推定するアルゴリズムを図 3.3 に示す．変数 j は反復のステップを表わす変数で，7 〜13 行目の処理を繰り返す．開発データ D_{dev} における発話 u_k に対し，その正解の対話行為が自己開示ではなく，誤って自動推定された対話行為が自己開示であり，$uncertainty(u_k)$ が閾値 TU_i より大きいとき (9 行目)，正解の対話行為 d_i に対する重み $w_i^{(j)}$ を 10 行目の式にしたがって更新する．$uncertainty(u_k)$ は発話 u_k に対する対話行為推定の不確かさを表わす指標であり，9 つの対話行為に対する判定の信頼度 $r(d_i)$ を得たとき，その 1 位の信頼度と 2 位の信頼度の比と定義する[4]．TU_i は対話行為 d_i に対する重みを更新するか否かを決める $uncertainty(u_k)$ の閾値である．基本的には，不正解となった「自己開示」の信頼度と正解の対話行為 d_i の信頼度の差が大きいときほど $w_i^{(j)}$ により大きい値を加える．$w_i^{(j)}$ の値を増やすことにより，正解の対話行為 d_i の信頼度が高くなり，選ばれる可能性が増す．δ は重みの 1 回当たりの変動量を調整するパラメタである．本

[4] 1 位と 2 位の信頼度が近ければ近いほど，1 位の対話行為が正しくない可能性が高い．

1: $gold(u_k) \stackrel{def}{=}$ 発話 u_k の正解の対話行為
2: $predict_j(u_k) \stackrel{def}{=}$ j 回目の反復が終わった時点で自動推定された u_k の対話行為
3: $w_i^{(j)} \stackrel{def}{=}$ j 回目の反復における対話行為 d_i の重み
4: $r'_j(d_i) \stackrel{def}{=} w_i^{(j)} \cdot r(d_i)$ # 重み付けによって調整された対話行為 d_i の信頼度
5: $\forall i \;\; w_i^{(0)} \leftarrow 1$ # 初期化
6: **for** $j = 1$ **to** 500 **do**
7: $\quad \forall i \;\; w_i^{(j)} \leftarrow w_i^{(j-1)}$
8: $\quad$ **for all** $u_k \in D_{dev}$ **do**
9: $\quad\quad$ **if** $gold(u_k) = d_i$ **and** $d_i \neq$ 自己開示 **and** $predict_{j-1}(u_k)$ = 自己開示 **and** $uncertainty(u_k) > TU_i$ **then**
10: $\quad\quad\quad w_i^{(j)} \leftarrow w_i^{(j)} + \delta \times \left(\frac{r'_{j-1}(\text{自己開示}) - r'_{j-1}(d_i)}{r'_{j-1}(\text{自己開示})} \right)$
11: $\quad\quad$ **end if**
12: $\quad$ **end for**
13: $\quad update(predict_j)$
14: **end for**
15: $\forall i \;\; w_i \leftarrow w_i^{(j)}$ where $j = \arg\max_j \; eval_j(d_i)$
16: **return** $\{w_i\}$

図 3.3: 信頼度に対する重みを決定するアルゴリズム

研究では $\delta = 0.001$ とした．開発データの全ての発話について重みの調整が終わったら，新しい重みを用いて，システムによる自動推定の結果を更新する (13 行目).

一般に $w_i^{(j)}$ は収束するが，本研究では収束後の重みではなく，1 回の反復毎に開発データにおける対話行為推定の改善度 $eval_j(d_i)$ を測り，これが最も高い時点での重みを選択する (15 行目)．$eval_j(d_i)$ の定義は式 (3.4) であり，対話行為が d_i である発話のうち重み付けによって新たに正解となった発話数 ($|B|$) と，対話行為が「自己開示」である発話のうち重み付けによって新たに不正解となった発話数 ($|W|$) の差である[5].

$$
\begin{aligned}
&eval_j(d_i) = |B| - |W| \\
&\quad B = \{u_k \mid gold(u_k) = d_i \wedge predict_0(u_k) \neq gold(u_k) \wedge predict_j(u_k) = gold(u_k)\} \\
&\quad W = \\
&\{u_k \mid gold(u_k) = \text{自己開示} \wedge predict_0(u_k) = gold(u_k) \wedge predict_j(u_k) \neq gold(u_k)\}
\end{aligned}
\tag{3.4}
$$

本手法では，$uncertainty(u_k)$ が低いときは重みの更新を行わない．これは個々の対話行為の二値分類器の結果が十分に信頼できるとみなしているためである．閾値 TU_i は重みの更新を行うか行わないかをコントロールする働きをする．TU_i は

[5] $predict_0(u_k)$ は重み付けしない手法で選択された発話 u_k の対話行為を表わす.

表 3.5: 信頼度 1 位が不正解，2 位が正解となる対話行為の組と発話数

	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
d_1		42	72	1	231	161	111	118	54
d_2			36	0	4	1	0	102	3
d_3				0	5	8	4	1	1
d_4					6	6	8	0	0
d_5						2	4	6	7
d_6							151	9	0
d_7								0	0
d_8									0

d_1:自己開示，d_2:質問 (YesNo)，d_3:質問 (What)，d_4:応答 (YesNo)，
d_5:応答 (平叙)，d_6:あいづち，d_7:フィラー，d_8:確認，d_9:要求

重み w_i の推定に用いたものとは別の開発データを用いて最適化する．TU_i を変動させ，学習した重みを用いたシステムの eval の値が最大となる閾値を選択する．

3.5.4 特定の対話行為の組に対して機械学習で識別する手法

対話行為の中には互いに識別が難しい組み合わせがある．表 3.5 は，対話行為のそれぞれの組に対し，一方の対話行為の信頼度の順位が 1 位でかつ不正解，もう一方の対話行為の信頼度の順位が 2 位でかつ正解となる発話の開発データにおける数を示している．この表において発話数 (誤り数) の多い対話行為の組は，特に判定が難しいと考えられる．ここでは，このような対話行為の組に対し，適切な対話行為を選択する分類器を機械学習することを試みる．ただし,「自己開示」(d_1) については，3.5.3 項 で述べた信頼度の重み付けによる手法で対応することとし，ここでは d_1 を含まない組の中で表 3.5 における誤り発話数が多い組に着目する．具体的には，他と比べて誤り発話数の多い (あいづち, フィラー) と (質問 (YesNo), 確認) の 2 つの組について，機械学習により適切な対話行為を選択する．以上をまとめると，本手法は式 (3.5) にしたがって $\hat{d}$ を決定する．

$$\hat{d} = \begin{cases} \arg\max_{d_i} w_i \cdot r(d_i) & \text{if rank(1)=}d_1\text{(自己開示)} \\ \text{classify}(\text{rank}(1), \text{rank}(2)) & \text{if \{rank(1),rank(2)\} = } \{d_6, d_7\} \text{ or } \{d_2, d_8\} \\ \arg\max_{d_i} r(d_i) & \text{if それ以外} \end{cases} \tag{3.5}$$

rank(1)，rank(2) は判定の信頼度が 1 位，2 位の対話行為を表わし，classify(x, y) は 2 つの対話行為 x, y の中から一方を選択する分類器である．classify(x, y) の学習に使う素性は，組み合わせ素性も含めて対話行為 x と y の分類に用いる素性タイプの和集合とし，学習には LIBLINEAR を用いる．

3.6 評価実験

3.6.1 データ

対話コーパスとして，人間同士の自由対話を書き起こした名大対話コーパス [28] を用いた．実験では，対話コーパスの中から参加者が二名の対話のみを選択して用いた．対話数は 97，発話数は 91,906 である．コーパスを対話を単位としておよそ 80%，10%，10%に分割し，それぞれ訓練データ，開発データ，テストデータとした．開発データは最適な素性タイプの選択やパラメタの最適化に用いた．それぞれのデータセットの対話数と発話数を表 3.6 に示す.

対話コーパスにおける各発話に対し，対話行為タグを人手で付与した．タグの付与は言語学の知識を持つ作業者 1 名が実施した．対話行為タグの付与の一致率を調べるため，3 つの対話に対してのみ 2 名の作業者が対話行為タグを付与した．対話行為の一致率は 77.3%，κ 係数は 0.636 であった．コーパス全体における対話行為の頻度分布を表 3.7 に示し，訓練，開発，テストデータならびにコーパス全体における対話行為の相対的な頻度分布を表 3.8 に示す.「自己開示」の発話が全体の 6 割弱を占めていることがわかる．一方,「要求」の発話は全体の 1%程度と少ない.

表 3.6: 対話コーパス

データ	対話数	発話数
訓練	77	74228
開発	10	8984
テスト	10	8694

表 3.7: 実験データにおける対話行為の出現頻度の分布

d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
53625 (58.3%)	6423 (7.0%)	3943 (4.3%)	2123 (2.3%)	7492 (8.2%)	9217 (10.0%)	4404 (4.8%)	3930 (4.3%)	749 (0.8%)

表 3.8: 対話コーパスにおける対話行為の相対出現頻度 (%)

データ	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
全て	58.3	7.0	4.3	2.3	8.2	10.0	4.8	4.3	0.8
訓練	58.6	7.0	4.2	2.3	8.1	10.0	4.7	4.1	0.7
開発	56.0	7.1	4.1	2.4	9.0	10.7	4.3	5.2	1.1
テスト	58.1	6.4	5.0	2.0	7.4	9.4	5.9	4.6	1.3

3.6.2 パラメータ最適化

素性タイプ f_{18}（相手の過去の発話の対話行為列），f_{19}（話者の過去の発話の対話行為列）について，3.3.2 項で述べたように，過去の対話行為列の長さ N_h ならびに素性選択の閾値 T_h の最適化を行った．

本実験では，テストデータにおける発話の対話行為を推定する際，f_{18} と f_{19} の素性は正解の対話行為を用いる．実際には，過去の発話の対話行為は自動的に推定するべきである．しかし，このような実験設定では，対話行為推定の誤りが次の発話の対話行為の推定に影響し，対話行為の誤推定が前の発話の対話行為の誤りによるものか，それとも提案手法の不備など他の要因によるものなのかを区別できない．今回の実験では，提案手法の有効性を確認することに重点を置き，過去の発話の対話行為の分類に誤りはないという理想的な条件下で実験を行った．開発データにおける F 値が最大となった N_h と T_h の値を表 3.9 に示す．この表ではパラメタの最適化を行わないとき（$N_h = 1$，$T_h = 0$）の F 値も示した．'-' は $N_h = 1$，$T_h = 0$ のときに F 値が最大になった場合，すなわちパラメタの最適化によって F 値が向上しなかった場合を表わす．

この結果から，対話行為毎に話者自身の過去の発話の対話行為列，相手の過去の発話の対話行為列の最適な長さが異なることが確認できる．特に,「フィラー」については F 値が 11 もしくは 14 ポイント向上しており，パラメータ最適化の影響が大きい．これは「フィラー」の発話を認識するためにはそれまでの対話の流れが重要な情報であることを示唆する．一方で,「質問（YesNo）」,「応答（YesNo）」については自身の過去の発話，相手の過去の発話ともに $N_h = 1$，$T_h = 0$ のときが最良となっている.「質問（YesNo）」については，前の発話の対話行為の影響が小さいため，対話行為列の長さを変化させても影響がなかったと考えられる.「応答（YesNo）」については，前の相手の最後の発話が「質問」であることが多いため，f_{18} についてはひとつ前の相手の発話の対話行為だけを素性とすれば十分と考えられる．一方，f_{19} については，F 値が他の対話行為と比べて極端に低い．N_h，T_h の最適化の際には f_1(単語 n-gram) のみを素性としていることが原因と考えられる．

3.6.3 素性タイプの最適化

個々の対話行為に対して選択された素性タイプを表 3.10 に示す．また，図 3.10 のアルゴリズムでは素性タイプの集合 E が更新されるが，その更新の過程，すなわち素性タイプが選ばれる過程を付録 A.1 に示す．表 3.10 の結果から，対話行為毎に有効な素性タイプが大きく異なることが確認された．f_1(単語 n-gram) は全ての対話行為に共通して有効な素性タイプである．一方で，f_2(前発話の単語 n-gram) や f_8(前発話の文末付属語列) は全ての対話行為で不要であり，前の相手の発話の内容は有効な素性タイプではないと考えられる．表 3.11 は 3.4 節で定義した最も有効な素性タイプの一覧である．これらも対話行為毎に異なるが，f_1，f_{18}，f_{19} の

表 3.9: 過去の発話の対話行為の特徴のパラメタ

対話行為	f_{18}(相手) N_h	T_h	F 値	f_{19}(話者) N_h	T_h	F 値
自己開示	1	0	0.856	1	0	0.851
	–	–	–	5	10	0.851
質問 (YesNo)	1	0	0.722	1	0	0.737
	–	–	–	–	–	–
質問 (What)	1	0	0.714	1	0	0.707
	–	–	–	2	0	0.708
応答 (YesNo)	1	0	0.771	1	0	0.033
	–	–	–	–	–	–
応答 (平叙)	1	0	0.467	1	0	0.205
	2	0	0.483	–	–	–

対話行為	f_{18}(相手) N_h	T_h	F 値	f_{19}(話者) N_h	T_h	F 値
あいづち	1	0	0.637	1	0	0.588
	2	10	0.651	3	1	0.604
フィラー	1	0	0.278	1	0	0.355
	2	5	0.390	4	5	0.496
確認	1	0	0.333	1	0	0.357
	5	0	0.342	–	–	–
要求	1	0	0.405	1	0	0.395
	2	5	0.411	3	0	0.410

いずれかが選ばれており，これらが特に重要な素性タイプであることがわかる．表 3.12 は選択された素性タイプの数と素性数の一覧である．素性数は，組み合わせ素性を含めた場合と含めない場合の両方を示す．組み合わせ素性ありの場合に対話行為によって素性数が大きく異なるのは，対話行為毎に最も有効な素性タイプが異なるためである．最も有効な素性が f_1(単語 n-gram) のとき，単語 n-gram の素性数は多いので，組み合わせ素性の数も多くなる．

3.6.4 信頼度の重みの推定

3.5.3 項で述べた手法において，対話行為毎の信頼度の重み w_i は開発データを用いて推定した．一方，閾値 TU_i は，開発データとは別のデータで最適化する必要がある．本実験では，訓練データの 8 分割交差検定により TU_i を最適化した．交差検定の際には，機械学習の素性タイプや重み w_i は開発データで決定したものを用いるが，分類器の学習は分割されたデータ毎にやり直した．TU_i を 0 から 0.9 まで 0.1 刻みで変動させ，式 (3.4) の eval の値が一番大きい閾値を選択した.「自己開示」以外の対話行為に対する w_i と TU_i の一覧を表 3.13 に示す．d_4(応答 (YesNo)), d_7(あいづち)，d_8(確認) の 3 つの対話行為については，信頼度に対する重み付けを行っても対話行為推定結果は向上しなかったため，重みを 1 に設定している．すなわち，これらの対話行為の信頼度に対しては重み付けを行わない.

表 3.14 は，例として d_2，d_5，d_7 の 3 つの対話行為について，8 分割交差検定において分割された個々のデータ ($TR_1 \sim TR_8$) に対する eval の値を示している[6]．eval の値 は対話データによってばらつきが見られ，負の値になる (重み付けによって悪化する) こともある．この結果から，信頼度に対する重み付けに基づく手法は，対話によって効果的に働く場合とそうでない場合があることがわかった．信頼度に対する重み付けは,「自己開示」の判定の信頼度が他の対話行為に比べて高いこと

[6] TU_i は表 3.14 における「合計」が最も大きい値を選んで最適化している.

表 3.10: 選択された素性タイプ

	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{18}	f_{19}	f_{20}	f_{21}	f_{22}	f_{23}	f_{24}	f_{25}	f_{26}	f_{27}	f_{28}
d_1	✓				✓		✓		✓				✓	✓				✓	✓	✓	✓		✓	✓	✓		✓	✓
d_2	✓				✓		✓				✓	✓		✓	✓		✓		✓		✓	✓				✓	✓	✓
d_3	✓		✓		✓						✓		✓	✓	✓	✓	✓			✓	✓	✓		✓				✓
d_4	✓		✓				✓						✓			✓	✓	✓			✓							✓
d_5	✓					✓			✓	✓	✓	✓				✓		✓	✓	✓		✓	✓	✓				✓
d_6	✓		✓		✓					✓	✓	✓			✓			✓	✓	✓	✓		✓		✓	✓		
d_7	✓			✓	✓	✓					✓		✓	✓		✓	✓	✓	✓	✓		✓	✓			✓		✓
d_8	✓		✓		✓	✓				✓					✓					✓				✓		✓		
d_9	✓			✓	✓	✓	✓		✓		✓	✓	✓	✓	✓	✓			✓		✓	✓		✓	✓	✓	✓	

d_1:自己開示，d_2:質問 (YesNo)，d_3:質問 (What)，d_4:応答 (YesNo)，d_5:応答 (平叙)，d_6:あいづち，d_7:フィラー，d_8:確認，d_9:要求

表 3.11: 対話行為の分類に最も有効な素性タイプ

自己開示	f_{18}:相手の過去の発話の対話行為
質問 (YesNo)	f_1:単語 n-gram
質問 (What)	f_1:単語 n-gram
応答 (YesNo)	f_{18}:相手の過去の発話の対話行為
応答 (平叙)	f_{19}:話者の過去の発話の対話行為
あいづち	f_{18}:相手の過去の発話の対話行為
フィラー	f_{19}:話者の過去の発話の対話行為
確認	f_1:単語 n-gram
要求	f_1:単語 n-gram

表 3.12: 選択された素性タイプと素性数

対話行為	素性タイプ数	素性数	
		組み合わせ素性なし	組み合わせ素性あり
自己開示	15	206,330	632,662
質問 (YesNo)	14	146,262	4,125,975
質問 (What)	14	156,414	8,763,099
応答 (YesNo)	9	117,395	446,720
応答 (平叙)	14	226,650	668,491
あいづち	14	192,279	691,132
フィラー	16	182,670	586,103
確認	9	221,229	12,493,568
要求	20	248,706	14,127,841

表 3.13: 対話行為ごとの w_i と TU_i

	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
w_i	1.123	1.437	1.000	1.207	1.000	1.630	1.000	1.537
TU_i	0.4	0.5	-	0.5	-	0.5	-	0.5

を是正するための手法であるが，自己開示の発話の出現のしやすさは対話の内容に強く依存しており，自己開示の発話が多く出現する対話に対しては,「自己開示」以外の対話行為を選択しやすくする本手法が有効に働かなかったと推察できる．

表 3.14: 8 分割交差検定における分割データ毎の eval 値

	TR_1	TR_2	TR_3	TR_4	TR_5	TR_6	TR_7	TR_8	合計
d_2: $(TU_2 = 0.4)$	1	-1	-2	1	1	2	1	2	5
d_5: $(TU_5 = 0.5)$	5	1	10	8	6	9	0	4	43
d_7: $(TU_7 = 0.5)$	8	12	15	8	4	4	4	-6	49

3.6.5 対話行為推定の評価

対話行為を推定する提案手法の性能を評価する．評価基準は，各対話行為の推定の精度，再現率，F 値，ならびにこれら 3 つの全対話行為についてのマクロ平均とマイクロ平均である．なお，精度および再現率のマイクロ平均は正解率 (システムが選択した対話行為と正解の対話行為が一致する割合) に等しい．提案手法を 2 つのベースラインと比較する．一つは，全ての素性タイプを用いて，9 つの対話行為のいずれかを選択する分類器を LIBLINEAR で学習する手法 (BL_a) である．もう一つは，3.3 節で説明した方法で素性タイプを選択する手法 (BL_s) である．提案手法が個々の対話行為毎に最適な素性タイプを選択するのに対し，BL_s では 9 つのクラスの分類に有効な素性タイプのセットを 1 つだけ選択し，それを用いて全ての対話行為を分類する．BL_s で選ばれた素性タイプは，f_1，f_5，f_6，f_8，f_{14}，f_{15}，f_{18}，f_{19}，f_{22}，f_{24} であった．一方，提案手法として，3.5.1 項で述べた信頼度を比較する手法 (Pro_p)，3.5.2 項で述べた信頼度を素性とした機械学習を用いる手法 (Pro_m)，「自己開示」以外の対話行為の信頼度に対して高い重みを与える手法 (Pro_w)，判定の難しい対話行為の組に対して機械学習で適切な対話行為を選択する手法 (Pro_b) の 4 つを評価する．

まず，発話がある対話行為を持つか否かを判定するタスク (以下,「個別対話行為判定タスク」と呼ぶ) についてベースラインと提案手法を比較する．言い換えれば，個別対話行為判定タスクでは，図 5.3 の第 1 段階における対話行為毎に学習し

表 3.15: 個別対話行為判定タスクの結果

(a) 開発データ

	BL_s			Pro_p		
	P	R	F	P	R	F
自己開示	.851	.907	.878	.855	.920	.886
質問 (YesNo)	.699	.734	.716	.732	.751	.742
質問 (What)	.820	.590	.687	.809	.651	.721
応答 (YesNo)	.902	.847	.874	.951	.875	.911
応答 (平叙)	.737	.687	.711	.760	.748	.754
あいづち	.713	.591	.647	.763	.609	.678
フィラー	.652	.359	.463	.699	.440	.540
確認	.566	.236	.333	.644	.279	.389
要求	.750	.207	.324	.850	.293	.436
マクロ平均	.743	.573	.626	.785	.618	.673

(b) テストデータ

	BL_s			Pro_p		
	P	R	F	P	R	F
自己開示	.848	.919	.881	.856	.925	.889
質問 (YesNo)	.630	.838	.719	.763	.680	.719
質問 (What)	.327	.919	.483	.787	.672	.725
応答 (YesNo)	.827	.871	.848	.872	.885	.879
応答 (平叙)	.804	.741	.771	.804	.798	.801
あいづち	.776	.731	.753	.763	.717	.739
フィラー	.619	.311	.414	.612	.356	.450
確認	.191	.819	.310	.680	.254	.370
要求	.618	.347	.444	.714	.204	.317
マクロ平均	.627	.722	.625	.761	.610	.654

た分類器の性能を評価する．表 3.15 は同タスクにおける BL_s と Pro_p の精度 (P), 再現率 (R)，F 値 (F) を示している．表 3.15(a) は開発データの結果であり，対話行為毎に素性タイプを最適化することによって，全ての対話行為について評価値が同等もしくは向上していることが確認できる．一方，表 3.15(b) はテストデータの結果であり，Pro_p は BL_s に比べて F 値のマクロ平均が 2.9 ポイント向上した．しかしながら,「あいづち」と「要求」については F 値が低下している．これは開発データとテストデータとで対話の内容が異なるため，両データにおいて最適な素性タイプが一致していないためと考えられる．この結果は，自由対話では様々なトピックが話題に挙がるため，対話行為分類のための最適な素性タイプを実験的に決定することが難しいことを示唆する．

表 3.16 は，発話に対して 9 つの対話行為の中から該当するものを推定するタスク (以下,「対話行為推定タスク」と呼ぶ) における各手法の評価値を示している．2 つのベースラインを比較すると，BL_s はマクロ平均では BL_a を上回るが，マイクロ平均は等しい．対話行為を区別せずに単純に素性タイプを最適化しても，正解率は向上しないことがわかる．一方，4 つの提案手法の F 値のマイクロ平均はいずれもベースラインよりも高い．最も結果が良かったのは手法 Pro_b であった．BL_s と Pro_b の結果をマクネマー検定で検定したところ，5%の有意水準で有意差があった．また，正解の対話行為と Pro_b が選択した対話行為の対応表を付録 A.2 に示す.

対話行為毎に結果を比較すると,「応答 (YesNo)」「あいづち」「確認」「要求」については，Pro_b は BL_s に比べて F 値は改善しなかったが,「自己開示」「質問 (YesNo)」「質問 (What)」「応答 (平叙)」「フィラー」については F 値が 0.3〜9.7 ポイント改善した.

Pro_p と Pro_m を比較すると，Pro_m は「あいづち」「確認」「要求」以外の対話行為でより高い F 値が得られており，信頼度を素性とした機械学習の手法が有効であることを示している.「自己開示」について Pro_p と Pro_w を比較すると，再現率は Pro_p の方が高いが，精度ならびに F 値では Pro_w が上回る．信頼度に重み付けを行う Pro_w は，判定の信頼度が全般に高い「自己開示」が過度に選ばれることを抑制するための手法であるが，この手法により「自己開示」の false positive の誤りが減少したことが確認された．また，表 3.13 で重みを 1 より大きく設定した全ての対話行為で F 値が向上した．Pro_b は Pro_w と比べて，誤りが多かったために改めて機械学習で分類し直した「質問 (YesNo)」「フィラー」「確認」の結果が改善されていることが確認できた．ただし,「あいづち」については，精度，再現率に変化はあったが，F 値は変化しなかった.

本論文では，ベースラインで精度や再現率が低い対話行為に対して推定の性能を向上させることを目指したが，一部の対話行為についてはその目標が達成されていない．具体的には，ベースラインで性能の低い「フィラー」の評価値は向上しているが,「確認」や「要求」については逆にベースラインよりも低くなっている.「確認」については，表 3.15 より，個別対話行為分類タスクでは提案手法はベースラインを上回っているため，図 5.3 における二段階の処理のうち，第 1 段階で「確

表 3.16: 対話行為推定タスクの結果

	BL_a			BL_s		
	P	R	F	P	R	F
自己開示 (d_1)	.855	.949	.899	.851	.951	.898
質問 (YesNo)(d_2)	.743	.742	.742	.762	.745	.753
質問 (What)(d_3)	.739	.667	.701	.787	.672	.725
応答 (YesNo)(d_4)	.877	.885	.881	.874	.900	.887
応答 (平叙)(d_5)	.820	.804	.812	.819	.772	.795
あいづち (d_6)	.751	.751	.751	.768	.730	.748
フィラー (d_7)	.660	.378	.481	.608	.412	.491
確認 (d_8)	.658	.289	.402	.634	.318	.424
要求 (d_9)	.808	.214	.339	.724	.214	.331
Ma	.768	.631	.667	.759	.635	.672
Mi	.819	.819	.819	.819	.819	.819

	Pro_p			Pro_m			Pro_w			Pro_b		
	P	R	F	P	R	F	P	R	F	P	R	F
d_1	.852	.953	.900	.858	.951	.902	.859	.949	.901	.859	.949	.901
d_2	.754	.751	.752	.755	.761	.758	.754	.753	.753	.760	.753	.756
d_3	.807	.689	.743	.799	.700	.746	.797	.706	.749	.797	.706	.749
d_4	.876	.880	.878	.889	.885	.887	.876	.880	.878	.876	.880	.878
d_5	.818	.812	.815	.805	.846	.825	.811	.839	.824	.811	.839	.824
d_6	.758	.724	.741	.763	.716	.738	.758	.724	.741	.790	.699	.741
d_7	.607	.399	.482	.593	.423	.494	.598	.423	.495	.627	.553	.588
d_8	.678	.265	.381	.709	.258	.379	.678	.265	.381	.687	.276	.394
d_9	.773	.173	.283	.680	.173	.276	.643	.184	.286	.643	.184	.286
Ma	.769	.628	.664	.761	.635	.667	.753	.636	.668	.761	.649	.680
Mi	.819	.821	.821	.823	.823	.823	.824	.824	.824	.825	.825	.825

表 3.17: 組み合わせ素性の評価

	BL_a	BL_s	Pro_p	Pro_m	Pro_w	Pro_b
組み合わせ素性なし	.808	.815	.816	.816	.819	.823
組み合わせ素性あり	.819	.819	.821	.823	.822	.825

認」に該当するかを判定する時点では性能の向上が見られるものの，第 2 段階の対話行為を推定する段階で誤りを多く生じていることがわかる．一方,「要求」については表 3.15 でも表 3.16 でも提案手法はベースラインより劣る．この原因として，表 3.8 に示すように，コーパスにおいて「要求」の対話行為を持つ発話の数が他の対話行為と比べて極端に少ないことが考えられる．

組み合わせ素性の有効性を評価するために，組み合わせ素性を使用したモデルと使用しないモデルの F 値のマイクロ平均 (正解率) を比較した．結果を表 3.17 に示す．いずれの手法も組み合わせ素性を用いることで F 値が向上していることから，組み合わせ素性の有効性が確認できた．

3.6.6 考察

その他の機械学習手法との比較

前項までの実験では機械学習アルゴリズムとして L2 正則化ロジスティック回帰を用いたが，本項ではこれと他の機械学習アルゴリズムを比較する．また，対話行為毎に適切な素性タイプセットを設定するという提案手法の基本的な考え方が他の機械学習アルゴリズムでも有効であるかを検証する．そのため，BL_s(対話行為を区別しないで素性タイプを選択したベースライン) と提案手法のうち最も基本的な Pro_p を比較する実験を行う．

比較する機械学習アルゴリズムは SVM とする．カーネル関数として，線形カーネル，多項式カーネル，Radial Basic Function(RBF) カーネル，シグモイドカーネルの 4 つを用いる．多項式カーネルの次数は 3 とした．Pro_p では，対話行為毎に素性タイプを選択するために，素性タイプセットを変えて学習とテストを繰り返す必要があるが，SVM の学習は非常に時間がかかるため，現実的な時間では素性タイプの選択が終了しない．例えば，対話行為「自己開示」に該当するかを判定する二値分類器の学習に，LIBLINEAR では 16 秒を要するのに対し，SVM を LIBSVM[5][7] で学習するにはおよそ 168 倍の 2697 秒を要する．そこで，高速な LIBLINEAR を用いて選択された素性タイプのセット (表 3.10) を用い，対話行為毎の二値分類器を学習するときのみ SVM を用いる．同様に，BL_s も LIBLINEAR を用いて選択された素性タイプのセットを用いる．また，L2 正則化ロジスティック回帰とは異なり多

[7] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

表 3.18: 機械学習アルゴリズムの比較

		SVM							
ロジスティック回帰		線形		多項式		RBF		シグモイド	
BL_s	Pro_p	BL_s	Pro_p	BL_s	Pro_p	BL_s	Pro_p	BL_s	Pro_p
.815	.816	.806	.801*	.560	.560*	.764	.773*†	.763	.770*

*: $p < 0.05$ (vs. ロジスティック回帰の Pro_p), †: $p < 0.05$ (vs. BL_s)

項式カーネルの SVM では素性の組み合わせも学習時に考慮されるため，ここでの実験では組み合わせ素性は用いない．SVM の学習には LIBSVM を用いる．Pro_p で用いる個々の対話行為判定の信頼度は，LIBSVM が出力する確率とする．

SVM による対話行為推定の F 値のマイクロ平均 (正解率) を表 3.18 に示す．比較のため，L2 正則化ロジスティック回帰を用いたときの結果 (表 3.17 の組み合わせ素性なしの結果) も再掲する．*と†はマクネマー検定で Pro_p と他の手法の差を検定した結果を表わす．*はロジスティック回帰の Pro_p との間に有意水準 5%で有意差が，†は同じ学習アルゴリズムの BL_s との間に有意差があることを示している．多項式カーネルの SVM では，BL_s，Pro_p ともに全ての発話に対して「自己開示」が選択された．これは過学習のためと考えられる．

異なる機械学習アルゴリズムの Pro_p の結果を比較すると，L2 正則化ロジスティック回帰は全ての SVM よりも正解率が有意に高い．ただし，今回の実験では，素性タイプ選択の際に用いた機械学習アルゴリズム (ロジスティック回帰) と対話行為推定の分類器の学習に用いたアルゴリズム (SVM) が異なる．素性タイプ選択も SVM で行えば，SVM での分類に適した素性タイプセットが選ばれて，正解率が向上する可能性がある．とはいえ，LIBLINEAR 以外では素性タイプ選択に非常に時間がかかるという問題がある．今回の実験結果からは，LIBLINEAR を用いる手法が対話行為推定の F 値ならびに計算時間の両方の観点から最も優れているといえる．

機械学習アルゴリズム毎に BL_s と Pro_p を比較すると，ロジスティック回帰，RBF カーネルの SVM，シグモイド関数の SVM において，差はそれほど顕著ではないものの，Pro_p は BL_s よりも正解率が高かった．有意義な分類器が学習できなかった多項式カーネルの SVM を除けば，4 つのうち 3 つの機械学習アルゴリズムについて，対話行為毎に最適な素性タイプを選択するというアプローチは有効と言える．但し，提案手法のアプローチの妥当性をより正確に検証するためには，異なる対話行為のセットを用いた実験や，異なる対話コーパスを用いた実験などを行う必要があるだろう．

次に，文献 [17] で比較的良好な結果を収めていることが報告されていることから，CRF を用いて対話行為を推定する追加実験を行った．CRF の学習には CRFsuite[8] を用いた．CRF では分類対象とする系列全体を入力としなければならない．つまり，対話が終了した後に，その対話の全ての発話の列を入力する必要がある．しかし，

[8] http://www.chokkan.org/software/crfsuite/

表 3.19: CRF による対話行為推定の結果

CRF_{all}		CRF_{seq}	
BL_a	BL_s	BL_a	BL_s
.825	.828	.810	.809

対話システムへの組み込みを想定する場合には，解析対象となる発話より後の発話は未知であるため，対話内の全ての発話を利用することができない．そのため，対話システムに CRF による対話行為推定モジュールを組み込む場合は，解析対象となる発話とそれ以前の発話の列が入力となる．入力となる発話の列は一つの発話を解析するたびに逐次更新され，対話が進むほど利用できる発話列が増える．よって，解析対象の発話が対話の先頭に近いほど利用できる情報が少なく，対話内の発話全てを用いた CRF よりも分類性能が低下することが予想される．この追加実験では，以下の 2 種類の手法を評価する．実際の対話システムの中に組み込むことはできないが，対話内の全ての発話列を用いて CRF で対話行為を推定する手法（CRF_{all}）と，実際の対話システムへの実装を想定し，解析対象となる発話までの発話列を用いて CRF で対話行為を推定する手法（CRF_{seq}）である．なお，CRF_{all} は，文献 [17] で行われているように，コーパスに対話行為のタグ付けをする際には利用できる．学習に用いる素性タイプのセットは，BL_a と同じ素性タイプセット (28 個の全ての特徴) を用いた場合と，BL_s と同じ素性タイプセットを用いた場合の二通りとする．表 3.19 に CRF による対話行為推定の F 値のマイクロ平均を示す.

CRF_{all} において，BL_s と同じ素性タイプセットを用いたときは 0.828 であり，提案手法の最高の成績 (Pro_b の 0.825) をわずかに上回る．しかしながら，前述のように，この手法は対話システムへは組み込めない．CRF_{seq} の F 値は BL_a の素性タイプセットを用いた場合の方が高いが，Pro_b よりは低い．マクネマー検定で Pro_b と CRF の結果を比較したところ，Pro_b と CRF_{seq} の間には 5%の有意水準で有意差があったが，Pro_b と CRF_{all} の間には有意差はなかった．以上から，対話システムに組み込むことを想定した場合，本論文の提案手法は CRF による手法を上回ると言える．

最後に，ランダムフォレスト [3] と提案手法を比較する実験を行った．ランダムフォレストの学習には scikit-learn[9] を用いた．学習に用いる素性タイプのセットは，CRF と同様に，BL_a，BL_s と同じとする．表 3.20 にランダムフォレストによる対話行為推定の F 値のマイクロ平均を示す．提案手法 Pro_b の推定精度は 0.825 で，この結果よりも高く，またマクネマー検定で 5%の有意水準で有意差があることを確認した．よって，本論文の提案手法はランダムフォレストによる手法を上回ると言える．

[9] http://scikit-learn.org/

表 3.20: ランダムフォレストの結果

BL_a	BL_s
.763	.774

素性タイプ選択と対話行為列の長さの最適化の順序に関する考察

提案手法では，素性タイプ f_{18}(相手の過去の話者の対話行為)，f_{19}(話者の過去の発話の対話行為) における対話行為列の長さを最適化してから個々の対話行為の素性タイプのセットの最適化を実施している．しかし，先に個々の対話行為毎に最適な素性タイプのセットを決めてから，その中に f_{18} や f_{19} が選ばれたとき，対話行為列の長さを最適化する方法も考えられる．ここでは両者を実験的に比較する．

提案手法と比較するために，以下の手続きによって対話行為推定システムを構築する．まず，対話行為列の長さのパラメタを $N_h = 1$，対話行為列の素性選択の閾値を $T_h = 0$ と設定する．次に，各対話行為における素性タイプセットを 3.3.1 項の手順に従い最適化する．その次に，3.3.2 項で述べたように N_h と T_h を最適化する．この手法は，対話行為毎の素性タイプの選択と，対話行為列の素性のパラメタの最適化を行う順序が，提案手法と逆になっている．

個々の対話行為の素性タイプを最適化してから対話行為列の長さを最適化する手法を StoL，対話行為列の長さを最適化してから個々の対話行為の素性タイプの最適化を行う手法を $LtoS$ と記す．LtoS が提案手法である．それぞれの手法の開発データ上での評価結果を表 3.21 に示す．また，StoL について，3.6.2 項と同様に，開発データにおける F 値が最大となった N_h と T_h の値を表 3.22 に示す．ただし,「質問 (YesNo)」「フィラー」「確認」「要求」の対話行為については f_{18} が,「質問 (What)」「応答 (平叙)」の対話行為については f_{19} が，有効な素性タイプとして選ばれなかった．その場合，表 3.22 では全て ‘-’ を埋めた．

表 3.21 に示されているように，手法 $LtoS$ の F 値のマイクロ平均は手法 $StoL$ の F 値のマイクロ平均より 2 ポイント高い．また，表 3.22 の結果は，手法 $LtoS$ の結果である表 3.9 と比較すると，長さの最適化による F 値の向上が小さい．つまり，素性タイプ最適化後では，素性タイプ最適化前と比べて，対話行為推定の F 値の向上に対する対話行為列の長さの最適化の貢献度が低くなる．

これらの結果から，処理の順序として，対話行為列の長さを最適化した後，素性タイプセットを選択する手法が妥当であるといえる．

Pro_w に関する考察

3.5.3 項で説明した手法では，信頼度に対して重みを与えているのは，信頼度 1 位の対話行為が「自己開示」の場合のみである．一方，信頼度が 1 位の対話行為が何かによらず，常に信頼度に重みをかけて，その値が最大の対話行為を選択す

表 3.21: 最適化の順番による結果の比較

	$StoL$			$LtoS$		
	P	R	F	P	R	F
自己開示	.849	.948	.896	.846	.951	.895
質問（YesNo）	.763	.750	.756	.756	.754	.755
質問（What）	.796	.700	.745	.806	.697	.748
応答（YesNo）	.892	.866	.879	.880	.880	.880
応答（平叙）	.791	.784	.787	.801	.755	.782
あいづち	.760	.723	.741	.753	.734	.743
フィラー	.552	.327	.411	.601	.356	.447
確認	.645	.300	.410	.673	.300	.415
要求	.677	.214	.326	.714	.204	.317
マクロ平均	.747	.624	.661	.760	.626	.665
マイクロ平均	.814	.814	.814	.816	.816	.816

る手法も考えられる．この手法は式 (3.6) のように定式化される．

$$\hat{d} = \arg\max_{d_i} w_i \cdot r(d_i) \tag{3.6}$$

式 (3.3) に示した Pro_w とは異なり，常に重み w_i をかけて対話行為の信頼度を比較し，最大のものを選ぶ．ここでは,，式 (3.6) による選択手法を採用せず,「自己開示」が 1 位のときのみ重みを与えた理由を説明する．

本研究の初期の段階では式 (3.6) に基づいて対話行為を選択する手法を検討し，実装した．この際，対話行為の重み w_i は図 3.4 のアルゴリズムで推定した．変数 j は反復のステップを表わす変数で，7 〜16 行目の処理を繰り返す．開発データ D_{dev} における発話 u_k に対し，誤った対話行為が自動推定されたとき (9 行目)，正解の対話行為 d_i に対する重み $w_i^{(j)}$ と誤った対話行為 d_p の重み $w_p^{(i)}$ を 12 行目，13 行目の式にしたがって更新する．基本的には，不正解となった対話行為 d_p の信頼度と正解の対話行為 d_i の信頼度の差が大きいときほど $w_i^{(j)}$ により小さい値を加える[10]．$w_i^{(j)}$ の値を増やすことにより，正解の対話行為 d_i の信頼度が高くなり，選ばれる可能性が増す．一方で，$w_p^{(i)}$ の値を減らすことで，誤った対話行為 d_p の信頼度が低くなり，選ばれる可能性が減る．δ は重みの 1 回当たりの変動量を調整するパラメタである．この実験では $\delta = 0.005$ とした．開発データの全ての発話について重みの調整が終わったら，新しい重みを用いて，システムによる自動推定の結果を更新する (16 行目).

[10]予備実験では，信頼度の差が大きいほど $w_i^{(j)}$ に大きい値を加えるように重みの更新式を定義した手法も試したが，対話行為推定の F 値は図 3.4 の手法と比べて低かった．

表 3.22: StoL における過去の発話の対話行為の特徴のパラメタ

対話行為	f_{18}(相手) N_h	T_h	F 値	f_{19}(話者) N_h	T_h	F 値
自己開示	1	0	0.882	1	0	0.882
	3	0	0.883	5	10	0.885
質問 (YesNo)	–	–	–	1	0	0.740
	–	–	–	–	–	–
質問 (What)	1	0	0.719	–	–	–
	4	5	0.723	–	–	–
応答 (YesNo)	1	0	0.899	1	0	0.899
	2	1	0.903	–	–	–
応答 (平叙)	1	0	0.750	–	–	–
	2	10	0.757	–	–	–

対話行為	f_{18}(相手) N_h	T_h	F 値	f_{19}(話者) N_h	T_h	F 値
あいづち	1	0	0.686	1	0	0.686
	–	–	–	3	1	0.690
フィラー	–	–	–	1	0	0.518
	–	–	–	–	–	–
確認	–	–	–	1	0	0.363
	–	–	–	–	–	–
要求	–	–	–	1	0	0.456
	–	–	–	2	5	0.465

式 (3.6) にしたがって対話行為を選択する手法を $Pro_{w'}$ とし，これを重み付けをしないで対話行為を選択する手法 Pro_p(式 (3.2)) と比較した．表 3.23 は，Pro_p では不正解だった発話が $Pro_{w'}$ では正解した発話，すなわち重み付けによって不正解が正解に転じた発話の数を示す．行は Pro_p によって推定された対話行為，列は $Pro_{w'}$ によって推定された対話行為であり，行が d_i で列が d_j のセルは，Pro_p では対話行為 d_i を選んで不正解であったが，$Pro_{w'}$ では対話行為 d_j を選んで正解となった発話の数を示している．一方，表 3.24 は，Pro_p では正解だった発話が $Pro_{w'}$ では不正解となった発話，すなわち重み付けによって正解が不正解に転じた発話の数を示す．行が d_i で列が d_j のセルは，Pro_p では対話行為 d_i を選んで正解であったが，$Pro_{w'}$ では対話行為 d_j を選んで不正解となった発話の数を示している．

表 3.23 の発話数の和は表 3.24 の発話数の和よりも多いことから，重み付けによって対話行為推定の正解率が向上することが確認できる．また，表 3.23 で不正解から正解に変わった発話のほとんどが，あるいは表 3.24 で正解から不正解に変わった発話のほとんどが，Pro_p で d_1(自己開示) が選ばれたとき (信頼度が 1 位の対話行為が「自己開示」のとき) であることがわかる．言い換えれば，Pro_p で「自己開示」以外の対話行為が選ばれたときは，信頼度に対する重み付けを行っても結果は大きく変化しない．重み付けに基づく手法は,「自己開示」の対話行為の信頼度が他の対話行為の信頼度と比べて大きいために過度に選ばれやすいという問題に対応するためのものであった．開発データでは信頼度が 1 位の対話行為が「自己開示」のときの誤りが多く，したがって「自己開示」に対する重みが低くなるように学習される．このことが，重み付けによって正解・不正解が変化するのは「自己開示」の信頼度が 1 位の場合がほとんどであった理由と考えられる．上記の考察を踏まえ，本論文では，信頼度 1 位の対話行為が「自己開示」のときのみに重み付けを行う手法を提案した．

表 3.23: 重み付けによって不正解から正解となった発話数

	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
d_1	0	0	11	0	38	20	27	8	14
d_2	0	0	6	0	0	0	0	7	1
d_3	0	0	0	0	0	0	0	0	0
d_4	0	0	0	0	1	2	3	0	4
d_5	0	0	0	0	0	0	0	0	0
d_6	0	0	0	0	0	0	1	0	0
d_7	0	0	0	0	0	0	0	0	0
d_8	0	0	0	0	0	0	0	0	0
d_9	0	0	0	0	0	0	0	0	0

表 3.24: 重み付けによって正解が不正解となった発話数

	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
d_1	0	0	13	0	18	18	9	10	6
d_2	0	0	5	0	0	0	0	5	0
d_3	0	0	0	0	0	0	0	0	0
d_4	0	0	0	0	0	2	0	0	0
d_5	0	0	0	0	0	1	0	0	0
d_6	0	0	0	0	0	0	3	0	0
d_7	0	0	0	0	0	0	0	0	0
d_8	0	0	0	0	0	0	0	0	0
d_9	0	0	0	0	0	0	0	0	0

d_1:自己開示，d_2:質問 (YesNo)，d_3:質問 (What)，d_4:応答 (YesNo)，
d_5:応答 (平叙)，d_6:あいづち，d_7:フィラー，d_8:確認，d_9:要求

$gold(u_k) \stackrel{def}{=}$ 発話 u_k の正解の対話行為
$predict_j(u_k) \stackrel{def}{=}$ j 回目の反復が終わった時点で自動推定された u_k の対話行為
$w_i^{(j)} \stackrel{def}{=}$ j 回目の反復における対話行為 d_i の重み
$r'_j(d_i) \stackrel{def}{=} w_i^{(j)} \cdot r(d_i)$ # 重み付けによって調整された対話行為 d_i の信頼度
$\forall i \;\; w_i^{(0)} \leftarrow 1$ # 初期化
for $j = 1$ **to** 1000 **do**
    $\forall i \;\; w_i^{(j)} \leftarrow w_i^{(j-1)}$
    for all $u_k \in D_{dev}$ **do**
        if $gold(u_k) = d_i$ **and** $predict_{j-1}(u_k) \neq d_i$ **then**
            $d_i \leftarrow gold(u_k)$
            $d_p \leftarrow predict_{j-1}(u_k)$
            $w_i^{(j)} \leftarrow w_i^{(j)} + \delta \times \left(1 - \frac{r'_{j-1}(d_p) - r'_{j-1}(d_i)}{r'_{j-1}(d_p)}\right)$
            $w_p^{(j)} \leftarrow w_p^{(j)} - \delta \times \left(1 - \frac{r'_{j-1}(d_p) - r'_{j-1}(d_i)}{r'_{j-1}(d_p)}\right)$
        end if
    end for
    $update(predict_j)$
end for
return $\{w_i\}$

図 3.4: $Pro_{w'}$ に用いた重みの最適化アルゴリズム

Pro_b に対する考察

3.5.4 項で述べた手法 (Pro_b) では，信頼度の順位が 1 位と 2 位になるときに誤りが多く発生する対話行為の組について，それらを判別する分類器を別途学習し，それを用いて改めて対話行為の選択を行う．この際，分類器を学習する対話行為の組 (Pro_b の場合は $\{d_6, d_7\}$ と $\{d_2, d_8\}$) は人手で選定している．しかし，分類器を別途学習する対話行為の組は自動的に選択される方が望ましい．最も簡単な方法は，表 3.5 における誤り発話数に対する閾値を設定し，その閾値以上の誤りが発生する対話行為の組を選択する手法や，誤り発話数の順位に対して閾値を設定し，その順位以上の対話行為の組を選択する手法が考えられる．ただし，閾値をどのように設定するかという問題がある．また，分類器を学習する対話行為の組を変え，それぞれについて開発データ上の F 値を測定し，F 値が最も高くなる対話行為の組のセットを選ぶ方法も考えられる．これらの手法の実験による評価は今後の課題とする．

3.7 まとめ

本論文では，自由対話における発話の対話行為を自動推定する新しい手法を提案した．提案手法は，個々の対話行為毎に発話がその対話行為に該当するかを判定する第 1 段階と，第 1 段階の結果から最終的に最も適切な対話行為を選択する第 2 段階から構成される．第 1 段階において，教師あり機械学習のために有効な素性タイプは対話行為毎に異なるという仮定の下，対話行為毎に最適な素性タイプのセットを自動的に決定する点に特長がある．評価実験の結果，対話行為を区別せずに素性タイプの選択を行う手法と比べて，提案手法の対話行為推定の F 値は 0.6 ポイント高かった．F 値の差はそれほど大きくはないものの，統計的に有意な差があることを確認した．

本章の貢献は以下の通りである．表 3.10 に示すように，有効な素性タイプのセットは対話行為によって異なることを実験的に確認し，対話行為毎に素性タイプの最適化を行う提案手法のアプローチが有望であることを示した．また，過去の対話行為を素性タイプとするとき，その最適な長さは対話行為毎に異なることを確認した．さらに，提案手法の第 2 段階において，分類の信頼度を単純に比較して対話行為を 1 つ選択すると，分類の信頼度が対話行為によって大きく差があるために特定の対話行為 (具体的には「自己開示」) が選ばれやすいという問題に対し，適切な対話行為を選択する 3 つの手法を提案し，それらが F 値の向上に貢献することを確認した．一方，対話行為によっては，素性タイプの最適化により，開発データでは F 値が向上するもののテストデータは低下することがわかった．自由対話システムでは様々なトピックが話題になることから，対話によって有効な素性タイプや素性が異なる可能性があり，素性タイプの最適化を実験的に行う提案手法のアプローチの問題点も明らかにした．表 3.14 に示したように，信頼度の重み付けに基づく手法が対話によって有効に働く場合とそうでない場合があることがわかったが，これも自由対話システムにおけるトピックの多様性に起因すると考えられる．

今後の課題としては，F 値が依然として低い「フィラー」「確認」「要求」に対して対話行為推定の性能を向上させることが挙げられる．これらの対話行為の推定に有効な新たな素性タイプを発見したり，提案手法の第 2 段階における対話行為選択手法を洗練する必要がある．また，上記の考察を踏まえ，領域適応の技術を応用し，対話の内容が訓練データ・開発データとテストデータとで異なる場合でも F 値を低下させない方法を探究することも重要な課題である．

第 4章

共感の推定

発話者の共感は，自由対話中における話題の転換のタイミングを計る際に重要な役割を果たす．自由対話では話題は固定されておらず，任意のタイミングで変更することが可能である．しかし，話題を変更するタイミングはいつでも良いわけではない．相手がまだその話題について話をしたい時に話題を変更したり，またはその話題についてはこれ以上対話を続けたくないにも関わらず対話を続けることは，相手に不快感を与え，場合によっては対話を打ち切られる可能性がある．話者が共感を示しているかを推察することができれば，適切なタイミングで話題を転換することが可能である．

本章では，自由対話における発話に対し，それが共感発話であるか否かを推定する手法を提案する．ここで，共感発話とは，相手の発話を受けて，その相手に対する共感や賛意を示している発話と定義する．単に同意を示す発話は含まれない．

本論文では，共感発話を推定するために教師あり機械学習を用いる．具体的には，Suport Vector Machine(SVM) を用いて，与えられた発話が共感を示しているかを判定する二値分類器を学習する．また，共感発話の推定に有効と考えられる学習素性 (素性) を提案し，その有効性を実験的に評価する．

4.1 素性ベクトル

4.1.1 共感推定のための素性タイプ

発話が共感を示しているかを判定するために，発話を素性のベクトルで表現する．本論文では，共感の有無を判定する手がかりになると考えられる 9 つの素性タイプを提案する．また，ベクトルの重みは二値とする．すなわち，ベクトルの次元に対応する素性が発話に出現していれば重みを 1，出現していなければ 0 と設定する．以下，提案する 9 つの素性タイプを順に説明する．

F_{ng}: 単語 n-gram

発話に出現する単語 n-gram(n=1,2,3) を素性とする．この素性は，発話の内容

を反映し，既存研究においても広く用いられている基本的な素性である．また，現在の発話内容だけでなく，前の発話内容も重要であると考えられるため，現在の発話と前の発話の両方の単語 n-gram を素性とする．

F_{len}: 発話長

共感を示す発話は短い傾向にあるため，発話長（文字数）を素性として用いる．発話長を二値の素性ベクトルの次元とする場合，長さを適当な間隔に分けて，例えば「1〜5」「6〜10」「11 以上」などに分けて，発話長を分類するのが一般的である．しかしながら，その適切な間隔を決めるのは難しい．本研究では，発話の長さを表わす素性を式 (4.1)，(4.2) のように定義する．

$$f_{len}^{(i)} = \begin{cases} 1 & \text{if } l_u \text{ is in } [i-2, i+2] \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

$$f_{len}^{(long)} = \begin{cases} 1 & \text{if } l_u \geq 20 \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

式 (4.1) は，発話長が $i \pm 2$ の範囲にあることを表わす素性である．ただし，$3 \leq i \leq 19$ とする．一方，式 (4.2) は，発話長が長い (20 以上である) ことを表わす素性である．発話長は，$F_{len}^{(i)}$ と $F_{len}^{(long)}$ の合計 18 個の素性で表わす．上記の手法は，発話長の範囲をあらかじめ固定する手法に比べて，発話の長さをより柔軟に表現できると考えられる．

F_{tu}: 話者交代

本研究で使用した対話コーパスでは一人の発話者が連続して発話を行う場合がある．共感は相手の発話を受けて発話されるため，つまり話者交替があったときに共感を示すことが多いため，発話者に変更があったか否かを素性タイプとする．

F_{rw1}: 自立語繰り返し (1)

発話者は他の発話者の以前の発話の単語を繰り返すことで共感を示す場合がある．以下の例では，発話者 A が「あれは傑作だった」と言ったのに対して話者 B が「傑作だね」と応じ，相手の発話中の語を繰り返して共感を示している．

A: あの/映画/は/傑作/だ[1]
B: 傑作/だ/ね

この素性タイプは現在の発話と前の発話で同じ自立語が存在するかを示す．

F_{rw2}: 自立語繰り返し (2)

自立語の繰り返しは常に共感を示すわけではない．以下に例を挙げる．

[1] '/' は単語の境界を示す．

A: 海草/類/嫌い/な/の/?
B: そう/で/も/ない/よ/、/海草

話者 B は「海草」という単語を繰り返しているが，その発話は共感を示してはいない．この素性タイプは F_{rw1} と同じ考えに基づくが，共感を示す自立語の繰り返しをより厳密に定義する．具体的には，F_{rw2} は，以下の 2 つの条件のいずれかを満たすことを表わす．

- 相手の直近の発話の最後の用言が発話中に存在する．
- 発話に出現する自立語が 1 つのみであり，かつそれが相手の直近の発話に出現する．

F_{rc1}: 意味カテゴリの繰り返し (1)

話者は，同じ単語を繰り返さないが，似た意味を持つ単語を繰り返すことにより共感を示すことがある．以下の例では，二つの発話に自立語の重複はないが，話者 B は話者 A が発っした「面白い」と似ている「楽しい」という単語を繰り返すことで共感を示していると考えられる．

A: あの/映画/は/面白かっ/た
B: 楽しかっ/た/ね

この素性タイプは，発話の中に，前の発話に含まれる単語と同じ意味カテゴリを持つ単語が存在するかを表わす．単語の意味カテゴリは，ここでは分類語彙表 [56] における意味カテゴリを用いる．上記の例では,「面白い」と「楽しい」は分類語彙表ではともに 33011(快・喜び) という意味カテゴリを持つので，F_{rc1} の重みを 1 とする．

分類語彙表では 1 つの単語が複数の意味カテゴリを持つことがある．このとき，複数の意味カテゴリのうちどれかひとつでも一致していれば，F_{rc1} の重みを 1 とする．具体的には，現在の発話に含まれる全ての自立語に対する全ての意味カテゴリのリストと，前の発話に含まれる全ての自立語に対する全ての意味カテゴリのリストを作成し，両者に重複があるかをチェックする．

F_{rc2}: 意味カテゴリの繰り返し (2)

F_{rw2} と同様に，意味カテゴリの繰り返しをより厳密にチェックする．具体的には，以下の二つの条件のいずれかを満たすか否かを表わす．

- 相手の直近の発話の最後の用言の意味カテゴリと同じものを持つ用言が発話中に存在する．
- 発話に出現する自立語が 1 つのみであり，かつその単語の意味カテゴリと同じものを持つ単語が相手の直近の発話に出現する．

F_{da}: 対話行為

対話行為は共感の有無を判定する有力な手がかりであると考えられる．例えば，共感は相手の主張や考えなどに対して示されることが多いが，質問に対して共感することは少ない．そのため，発話の対話行為を素性タイプとして用いる．対話行為のセットは表 3.1 に示したものを使用する．

4.3 節の実験では，対話行為はコーパスに人手で付与されたものを用いる．本来，対話行為は 3 章で提案した手法で自動的に推定するべきである．しかしながら，共感推定の誤りが発生したとき，それが共感の推定手法に起因するのか，それとも対話行為の自動分類の誤りに起因するのかを判断するのが難しい．実験では共感推定モデルの評価を目的としているため，対話行為の分類に誤りはないと仮定し，正解の対話行為を素性として用いる．

F_{end}: 文末付属語列

発話者は発話の文末表現で共感を示すことがある．例えば,「〜だよね」「〜だね」といった文末表現は共感を示しているといえる．この考えに基づき，文末の表現を素性タイプとする．具体的には，文末に出現する付属語の列を素性として用いる．

4.1.2 組み合わせ素性

SVM では様々なカーネル関数が提案されている．予備実験として，線形カーネル，多項式カーネル，放射基底関数 (Radial Basis Function; RBF) の推定精度を比較した．その結果，多項式カーネルと RBF カーネルの推定精度は非常に低かった．そのため，本研究では SVM のカーネルとして線形カーネルを用いる．

線形カーネルの問題点は，学習に用いる素性が独立に評価され，素性間の相関関係もしくは依存関係が考慮されない点である．そのため，複数の素性を組み合わせた素性を導入することで，素性間の相関関係を学習に反映させることを試みる．いま，素性の集合を $F = \{\ldots f_i \ldots\}$ とおく．ここで f_i は 4.1.1 項で提案した素性タイプの素性である．そして，素性の全ての組み合わせ $[f_i, f_j](i \neq j)$ を新たに素性として導入する．以下，これを組み合わせ素性と呼ぶ．

組み合わせ素性の数は非常に多いため，過学習を引き起こす可能性がある．そのため，組み合わせ素性の中から有効なもののみを選別する．すなわち，いわゆる素性選択を行う．素性選択の手法は次項で述べる．

4.1.3 素性選択

単語 n-gram(F_{ng}) の素性と組み合わせ素性は非常に数が多い．一般に，訓練データが十分にないときに，過剰な素性を用いると過学習を起こしやすい．そこで，これら 2 つの素性に対しては，有効な素性を選別する素性選択の手法を導入する．

素性選択の手法として，3.3.2 項で述べた χ^2 値に基づく手法を採用する．まず，各素性と対話行為 d_i の相関の強さを χ^2 値で推定する．χ^2 値は式 (3.1) で計算される．以下に再掲する．

$$\chi^2 = \frac{N(o_{11}o_{22} - o_{12}o_{21})^2}{(o_{11}+o_{12})(o_{11}+o_{21})(o_{12}+o_{22})(o_{21}+o_{22})} \tag{4.3}$$

共感推定タスクの場合，o_{11} は共感を示しかつ素性 f_i が出現する発話数，o_{12} は共感を示しかつ素性 f_i が出現しない発話数，o_{21} は共感を示さずかつ素性 f_i が出現する発話数，o_{22} は共感を示さずかつ素性 f_i が出現しない発話数となる．表 4.1 は o_{ij} の定義を表としてまとめたものである．

表 4.1: o_{ij} の対応表

	f_i あり	f_i なし
共感発話	o_{11}	o_{12}
非共感発話	o_{21}	o_{22}

次に，χ^2 値が閾値よりも小さい素性を削除する．以下，単語 n-gram の素性を素性選択する際の閾値を T_{ng}，組み合わせ素性を素性選択する際の閾値を T_{comb} とおく．T_{ng} と T_{comb} は開発データで最適化する．

4.2 負例のフィルタリング

一般に，機械学習によって分類器を学習し，その分類器を用いて未知のデータを分類したとき，訓練データにおいて頻出する分類クラスが選ばれやすい傾向がある．訓練データにおける分類クラスの分布に著しい偏りがある場合，最頻出の分類クラスを常に選択するような分類器が学習されることがある．例えば，二値分類器を学習する際，訓練データにおける正例の数が負例の数と比べて極端に少ない場合，常に負と判定する二値分類器が学習されることがある．ひとつの分類クラスを常に選択する分類器は明らかに望ましくない．

しかし，一般的に自由対話において，共感を示す発話の出現比率は低い．実際，後に示す表 4.2 によれば，本手法の評価に用いたコーパスにおける共感発話の割合は 1.1% とかなり低い．このような対話コーパスをそのまま訓練データとして利用すると，その正例と負例の比率は著しく偏ったものとなる．

訓練データにおけるデータの不均衡を是正する手法として，正例を増やすオーバーサンプリングと，負例を削除するアンダーサンプリング手法が一般的である．本研究ではアンダーサンプリングの手法を採用する．アンダーサンプリングではランダムに負例を削除する手法が一般的だが [4, 52]，機械学習に有効な負例も削除してしまう可能性がある．

本論文では「冗長な負例」を削除することで正例と負例の不均衡を是正する．冗長な負例とは，ここでは，共感を示さず，かつ表現や内容が他の負例の発話と似ている発話と定義する．冗長な負例は，それから得られる特徴は他の負例からも得られるため，削除しても共感推定の精度はそれほど低下しないと考えられる．一方，冗長な負例を削除することで正例が訓練データに占める割合は増えるため，ほぼ常に負と判定する無意味な二値分類器が学習されることを妨げることができる．

冗長な負例を検出するため，訓練データにおける発話間の類似度を計算する．発話間の類似度は，発話を単語 n-gram を素性とするベクトルで表現し，それらのベクトル間のコサイン類似度で測る．もし 2 つの負例の発話の類似度が十分に高いとき，一方の発話を冗長な負例と判定する．しかしながら，訓練データの全ての発話間の類似度を計算することは多大な時間を要する．そのため，まず訓練データの発話をクラスタリングし，クラスタ内の発話間のみ類似度を計算する．クラスタリングでは，クラスタ数を 1000 として Repeated Bisection 法により負例の集合からクラスタを生成する．クラスタの作成にはクラスタリングツール Cluto[2]を用いる．次に，それぞれのクラスタにおいて，冗長な負例を検出し，それを削除する．このアルゴリズムを図 4.1 に示す．クラスタ U に属する発話集合に対して，それを訓練データから削除しない発話集合 U_k と削除する発話集合（冗長な負例の集合）U_d に分割する．U 中の発話はあらかじめ順番に並べられているが，その順序は任意である．つまり，U 中の発話の順序はランダムに決める．発話 u_i に対し，それと順位が下位の発話 u_j との類似度を計算し，類似度が閾値 S_{fil} よりも大きいとき，u_i を U_k に追加し (訓練データに残し)，u_j を U_d に追加する (訓練データから削除する)．複数の類似した発話が存在するとき，最初に現れた発話のみが訓練データに残ることに注意していただきたい．閾値 S_{fil} は削除する負例数を調整する働きをする．その最適値は開発データを用いて実験的に決定する．

4.3 評価実験

本節では提案手法の評価実験の手続きとその結果を報告する．共感タグが付与されたコーパスを訓練データとし，提案手法によって共感の有無を判定する二値分類器を学習する．また，開発データを用いてパラメタ (T_{ng}，T_{comb}，S_{fil}) を最適化する．最後に，最適化したパラメタを用いて，共感を判定する分類器を再度学習し，それを用いてテストデータの発話の共感の有無を推定する．

提案手法の評価基準は，共感推定の精度，再現率，F 値とする．それぞれの定義を式 (4.4),(4.5),(4.6) に示す．

$$\text{精度} = \frac{\text{推定された正しい共感発話の数}}{\text{推定された共感発話の総数}} \tag{4.4}$$

[2]http://glaros.dtc.umn.edu/gkhome/views/cluto

Input: $U = \{u_1, u_2, \cdots, u_n\}$
Output: U_k, U_d

$U_k \leftarrow \emptyset$, $U_d \leftarrow \emptyset$
for $j = 1$ **to** n **do**
    if $u_i \in U_d$ **then**
        next
    end if
    for $j \leftarrow i + 1$ **to** n **do**
        $sim \leftarrow cos(u_i, u_j)$
        if $sim \geq S_{fil}$ **then**
            $U_d \leftarrow U_d \cup \{u_j\}$
        end if
    end for
    $U_k \leftarrow U_k \cup \{u_i\}$
end for

図 4.1: 負例の削除アルゴリズム

$$\text{再現率} = \frac{\text{推定された正しい共感発話の数}}{\text{正解の共感発話の総数}} \tag{4.5}$$

$$\text{F 値} = \frac{2 \times \text{精度} \times \text{再現率}}{\text{精度} + \text{再現率}} \tag{4.6}$$

4.3.1 データ

評価実験のためのデータは，3.6 節における対話行為推定の実験に用いたデータと同じものを用いる．すなわち，名大対話コーパス [28] を用いる．名大対話コーパスは，2〜4 名の参加者の日本語による雑談対話を書き起こしたテキストである．この対話コーパスから対話参加人数が二人の対話を選び，対話内の個々の発話に対して共感を示しているかのタグを人手で付与した．共感タグの付与は一人のアノテーターによって行った．ただし，3 対話 (3294 発話) については，別のアノーテーターが独立してタグ付けを行い，2 名によるタグ付けがどれだけ一致しているかを調べた．κ 係数は 0.27，F 値は 0.28 となり，十分に高いとは言えなかった．これは共感発話を推定することが難しいことを示唆している．今回のタグ付けでは発話の書き起こしテキストだけで共感の有無を判定しており，音声やパラ言語情報を用いていないことも一致率が低い原因のひとつと考えられる．共感発話のより厳密な定義を定め，それを基に客観的な共感タグ付けのためのガイドラインを整備することは今後の課題である．評価実験では対話コーパスの発話をランダムに訓練データ (80%)，開発データ (10%)，テストデータ (10%) の三つの対話セット

表 4.2: コーパスの構成

	対話数	共感発話数	非共感発話数
訓練データ	77	861	73378
開発データ	10	103	8882
テストデータ	10	99	8598

に分けて利用する．この分割は 3.6 節におけるデータ分割と同一である．表 4.2 にそれぞれのデータの対話数，共感発話数，非共感発話数を示す.

コーパスにおける共感発話と非共感発話の比率は 1:86 と非常に偏っている．4.2 節で議論したように，このようなデータ上では分類器の精度や再現率が低くなりがちである．一方，本実験では，4.1.1 項で提案したそれぞれの素性タイプの有効性を評価することも目的としている．すなわち，ある素性タイプを学習に用いたときと用いないときの分類器を比較することで，共感の有無を判定するために有効な素性タイプを明らかにする．このような実験の際は，ベースとなる精度や再現率が低すぎると，素性タイプの有効性を正確に評価できない可能性がある．したがって，共感発話と非共感発話を同数含むデータを作成し，このデータも用いて提案手法を評価する．以下，表 4.2 に示したデータを「オリジナルデータ」，同数の共感発話と非共感発話から構成されるデータを「平衡データ」と呼ぶ.「平衡データ」は，テストデータ，開発データ，訓練データのそれぞれについて，非共感発話の中から共感発話と同数の発話，すなわち 99, 103, 861 個の非共感発話をランダムに選択することで作成する．また，ランダム試行による結果の揺れを考慮し，平衡データは 5 回作成し，それぞれについて精度，再現率，F 値を測る．以降で示す平衡データの実験結果は，5 回の試行の平均である.

4.3.2 共感推定の評価

ここでは提案手法による共感推定の性能を評価する.

まず，単語 n-gram 素性の素性選択の閾値 T_{ng} の最適値を求めた．図 4.2 は T_{ng} を変化させたときの開発データにおける精度 (P)，再現率 (R)，F 値 (F) の変動を示す．この結果から，$T_{ng} = 0.9$ のときに F 値が最大となった．よって T_{ng} の最適値を 0.9 と設定する．このとき，単語 n-gram の素性の総数は 4378 であった．同様に，組み合わせ素性の閾値 T_{comb} も最適化した．詳細は 4.3.4 項で報告する.

表 4.3 にオリジナルデータのテストデータにおけるベースラインと提案手法を比較した結果を示す．ベースラインは単語 n-gram のみを素性として学習した分類器である．一方，平衡データのテストデータにおける結果を表 4.4 に示す．なお，これらの表は負例のフィルタリングを適用していない手法の評価結果である．提案手法は全体的にベースラインを上回った．しかしながら，オリジナルのデータに

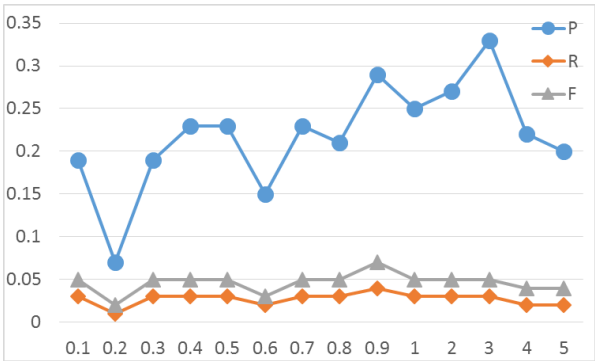

図 4.2: T_{ng} の最適化

おいては F 値は高くない．これは対話コーパスにおける共感発話の頻度が低いことが原因であると考えられる．

表 4.3: オリジナルのテストデータの結果

	P	R	F
ベースライン (F_{ng})	0.23	0.11	0.15
提案手法	0.28	0.13	0.18

表 4.4: 平衡データのテストデータの結果

	P	R	F
ベースライン (F_{ng})	0.80	0.73	0.76
提案手法	0.81	0.76	0.80

4.3.3 素性タイプの有効性の評価

素性タイプの有効性を評価するため，いくつかの素性タイプセットを比較する．単語 n-gram とその他の一つの素性タイプ ($F_{ng} + F_*$ と記す) を学習素性として用いた分類器を学習し，この分類器の評価指標を比較する．また，ベースライン (F_{ng}) の結果，全ての素性タイプを用いた分類器 (F_{all}) の結果とも比較する．表 4.5 はオリジナルの対話データを用いた結果，表 4.6 は平衡データを用いた結果を示す．なお，個々の素性タイプの有効性を比較するため，この実験では組み合わせ素性は用いていない．

オリジナルの対話データを使った結果では F_{len}，F_{rc2}，F_{da}，F_{end} を追加した場合は F 値が低下している．また，全ての素性を用いた場合も F_{ng} のみを用いた場合と比べて F 値は向上していない．しかし，平衡データではほぼ全ての素性タイプが F 値の向上に寄与している．さらに，F_{all} はベースラインよりも精度，再現率，F 値全てが向上している．話者交代 (F_{tu})，自立語の繰り返し (F_{rw1}，F_{rw2}) の素性

表 4.5: オリジナルデータにおける素性タイプの有効性の評価

素性セット	P	R	F
F_{ng}	0.23	0.11	0.15
$F_{ng} + F_{len}$	0.18	0.08	0.11
$F_{ng} + F_{tu}$	0.25	0.12	0.16
$F_{ng} + F_{rw1}$	0.25	0.11	0.15
$F_{ng} + F_{rw2}$	0.26	0.11	0.15
$F_{ng} + F_{rc1}$	0.23	0.11	0.15
$F_{ng} + F_{rc2}$	0.21	0.10	0.14
$F_{ng} + F_{da}$	0.19	0.08	0.11
$F_{ng} + F_{end}$	0.19	0.10	0.13
F_{all}	0.24	0.11	0.15

表 4.6: 平衡データにおける素性タイプの有効性の評価

素性セット	P	R	F
F_{ng}	0.80	0.73	0.76
$F_{ng} + F_{len}$	0.81	0.73	0.77
$F_{ng} + F_{tu}$	0.81	0.75	0.78
$F_{ng} + F_{rw1}$	0.81	0.73	0.77
$F_{ng} + F_{rw2}$	0.81	0.73	0.77
$F_{ng} + F_{rc1}$	0.81	0.72	0.76
$F_{ng} + F_{rc2}$	0.81	0.73	0.77
$F_{ng} + F_{da}$	0.81	0.73	0.77
$F_{ng} + F_{end}$	0.82	0.74	0.78
F_{all}	0.83	0.77	0.80

表 4.7: オリジナルデータにおける閾値 T_{comb} の最適化

	F_{ng}			F_{all}		
T_{comb}	P	R	F	P	R	F
100	0.25	0.10	0.14	0.19	0.09	0.12
110	0.26	0.10	0.14	0.19	0.09	0.12
120	0.25	0.10	0.14	0.19	0.09	0.12
130	0.25	0.10	0.14	0.19	0.09	0.12
140	0.26	0.10	0.14	0.22	0.11	0.14
150	0.23	0.10	0.14	0.20	0.10	0.13
160	0.24	0.10	0.14	0.20	0.10	0.13
170	0.24	0.10	0.14	0.17	0.08	0.11
180	0.17	0.09	0.11	0.11	0.06	0.08
190	0.15	0.08	0.10	0.11	0.06	0.07
200	0.14	0.07	0.09	0.12	0.07	0.09

タイプは，表 4.5，表 4.6 の両方でベースラインと比べて F 値が改善したことから，共感の判定に有効な素性タイプといえる．一方，他の素性タイプにおいては，オリジナルデータと平衡データとでベースラインからの F 値の増減が一致していないため，その有効性ははっきりしない．

4.3.4 組み合わせ素性の評価

組み合わせ素性の有効性を評価する．この実験では単語 n-gram 素性 (F_{ng}) と全ての素性タイプ (F_{all}) の二つの素性タイプのセットに組み合わせ素性を追加し，共感発話の推定結果を評価する．

組み合わせ素性の素性選択のための閾値 T_{comb} の最適化を開発データ上で行う．オリジナルデータならびに平衡データで T_{comb} の値を変動させたときの F 値の変化をそれぞれ表 4.7，表 4.8 に示す．平衡データでは，最初は 100 から 200 までの範囲で T_{comb} を変動させたが，200 以上に設定すると F 値が向上することが予測されたので，範囲を 100 から 300 に拡張した．オリジナルデータでは，T_{comb} を 140 にしたときが F_{ng}，F_{all} ともに F 値が最大となった．一方，平衡データでは，F_{ng} の場合は T_{comb}=280，F_{all} は T_{comb}=260 のときが最適な値となった．

表 4.9 と表 4.10 はそれぞれオリジナルデータ，平衡データにおける組み合わせ素性の有無による分類結果の比較を示している．$F_{ng} + COMB$ は単語 n-gram を素性タイプセットとしたときに組み合わせ素性を用いた手法，$F_{all} + COMB$ は全素性タイプとそれらの組み合わせ素性を用いた手法の結果である．表 4.9 より，組み合わせ素性を追加することで素性タイプ集合 F_{all} における精度，再現率が改善

表 4.8: 平衡データにおける閾値 T_{comb} の最適化

	F_{ng}			F_{all}		
T_{comb}	P	R	F	P	R	F
100	0.79	0.66	0.72	0.83	0.67	0.74
110	0.80	0.66	0.72	0.83	0.66	0.74
120	0.80	0.67	0.73	0.82	0.69	0.75
130	0.80	0.67	0.73	0.83	0.68	0.75
140	0.80	0.68	0.74	0.82	0.68	0.75
150	0.79	0.67	0.73	0.82	0.67	0.73
160	0.79	0.67	0.73	0.82	0.67	0.74
170	0.79	0.67	0.73	0.81	0.68	0.74
180	0.80	0.69	0.74	0.81	0.69	0.74
190	0.80	0.70	0.74	0.81	0.70	0.75
200	0.79	0.70	0.74	0.81	0.70	0.75
210	0.79	0.70	0.74	0.80	0.70	0.75
220	0.79	0.70	0.74	0.81	0.70	0.75
230	0.79	0.70	0.74	0.80	0.70	0.75
240	0.79	0.70	0.74	0.80	0.71	0.75
250	0.79	0.70	0.74	0.80	0.71	0.75
260	0.79	0.70	0.74	0.80	0.72	0.76
270	0.79	0.70	0.74	0.80	0.72	0.76
280	0.79	0.70	0.75	0.80	0.71	0.75
290	0.78	0.69	0.73	0.80	0.71	0.75
300	0.78	0.69	0.74	0.80	0.71	0.75

されたことが確認された．一方で，単語 n-gram 素性における組み合わせ素性の追加は，精度を向上させたが，再現率と F 値を低下させた．これらの結果から，組み合わせ素性は提案した素性タイプ集合では有効に働くが，単語 n-gram のみを素性としたベースラインでは有効ではないといえる．

表 4.9 では，$F_{all} + COMB$ は $F_{ng} + COMB$ と比べて精度は低いが，再現率は上回り，F 値も高い．一方，表 4.10 においては，精度，再現率，F 値の全ての指標で，$F_{all} + COMB$ は $F_{ng} + COMB$ を上回る．したがって，組み合わせ素性を使用するという条件の下でも，本研究で提案する素性タイプの有効性が確認できた．

表 4.9: オリジナルデータにおける組み合わせ素性の評価

素性セット	P	R	F
F_{ng}	0.23	0.11	0.15
F_{ng}+$COMB$	0.31	0.09	0.14
F_{all}	0.24	0.11	0.15
F_{all}+$COMB$	0.28	0.13	0.18

表 4.10: 平衡データにおける組み合わせ素性の評価

素性セット	P	R	F
F_{ng}	0.80	0.73	0.76
F_{ng}+$COMB$	0.80	0.73	0.77
F_{all}	0.83	0.77	0.80
F_{all}+$COMB$	0.81	0.76	0.80

4.3.5 負例フィルタリングの評価

負例フィルタリング手法の評価はオリジナルデータを用いて行う．まず，パラメータ S_{fil} の最適化を行う．S_{fil} は，負例が冗長かを判定する際に用いる発話間の類似度の閾値で，S_{fil} を小さく設定するほどより多くの負例を削除する．S_{fil} を変動させたときの開発データの共感推定の結果を表 4.11 に示す．F 値が最大となったのは S_{fil} の値が 0.3 のときであった．

この実験では，(1) 負例のフィルタリングなし，(2) 提案手法による負例のフィルタリング，(3) ランダムに負例を削除するフィリタリングの 3 つの手法を比較する．提案手法では，25,174 個の負例が冗長な負例として訓練データから削除された．(3) のランダムフィルタリングでは，これと同数の負例をランダムに削除する手法である．ただし，削除するデータがランダムに選択されることの評価への影響を排除するため，削除する負例をランダムに決める試行を 5 回繰り返し，その平均値を実験結果として示す．

表 4.12 に実験結果を示す．フィルタリングにより再現率が向上し，精度が低下した．これは，訓練データ中の負例が減少するほど分類器は発話を共感と判断しやすくなるため，自然な結果である．負例のフィルタリングにより F 値が向上していることから，正例と負例の数に偏りが見られるデータに対して，負例のフィルタリングは共感発話の分類性能の向上に貢献すると言える．しかし，提案手法の F 値は，負例をランダムに選択する手法の F 値よりも低い．すなわち，発話間の類似度計算に基づき冗長な負例を削除する方法の有効性は認められない．この原因は不明である．今後，この原因を解明し，負例のフィルタリングのアルゴリズムを改良する必要がある．

表 4.11: パラメータ S_{fil} の最適化

S_{fil}	P	R	F
0.9	0.08	0.05	0.06
0.8	0.10	0.06	0.07
0.7	0.09	0.06	0.07
0.6	0.08	0.05	0.06
0.5	0.09	0.07	0.08
0.4	0.07	0.08	0.08
0.3	0.08	0.25	0.13
0.2	0.06	0.45	0.11
0.1	0.04	0.55	0.07

表 4.12: 負例のフィルタリング手法の評価

	P	R	F
フィルタリングなし	0.28	0.13	0.18
提案手法	0.23	0.16	0.19
ランダムなフィルタリング	0.25	0.18	0.22

4.3.6 エラー分析

推定誤りの主な原因を解明するために，エラー分析を実施した．まず，false positive の誤り（共感を示す発話を非共感と推定する誤り）と false negative（非共感を示す発話を共感と推定する誤り）の多くは，前の発話が長いときに発生することがわかった．それらの場合では前の発話は多くの文を含むが，実際に推定対象の発話と関係があるのは最後の文一つだけと考えられる．そのため，前の発話から得られる素性の多くは，推定対象の発話とは無関係であり，推定誤りを引き起こしているものと考えられる．この問題を解決するためには，推定対象の発話と前の発話の結束性を考慮する必要がある．言い換えれば，長い前の発話の中から現在の発話と関係のある一文を選択する手法を導入する必要がある．また，推定対象の発話と前の発話の発話長が短すぎる場合にも誤りが多く見られた．これらの推定誤りは素性が足りていないことが原因と考えられる．提案手法では素性選択も実施しているため，短い発話では単語 n-gram の素性さえ抽出されないことがある．この問題に対しては，素性選択の対象を bi-gram と tri-gram に限定し，uni-gram の素性は全て残す方法が解決策として考えられる．

また，いくつかの false negative の誤りは，素性タイプ F_{end} が原因と考えられる．特定の文末表現は共感を示すために使われるが，それは常に共感を示しているわ

けではない．特に，一般に共感を示すと考えられる文末表現が，推定対象とその前の発話の両方が短いときには共感を示していない場合が多かった．このとき，抽出される素性の数が少ないこともあり，文末表現の特徴によって非共感発話が誤って共感発話と分類されていた．したがって，真に共感を示す文末表現とその出現条件を精査し，これを踏まえて文末表現の特徴の抽出方法を再考する必要がある．

4.4 まとめ

本章では，与えられた発話が相手への共感を示しているかを推定する手法について述べた．推定手法は教師あり機械学習に基づき，共感判定のための素性タイプを提案した．評価実験の結果，以下の 3 つの知見が得られた．(1) 提案手法の有効性を確認した．特に組み合わせ素性を用いたとき，ベースラインと比べて共感推定の F 値が大きく向上した．(2) 提案した素性タイプの中で，話者交代，自立語の繰り返しが共感発話の推定に有効であった．(3) 負例のフィルタリングは共感推定の F 値の向上に寄与することを示した．

正例と負例の分布に著しい偏りがあるオリジナルデータでは，共感推定の F 値は依然として低い．本論文では冗長な負例を削除するフィルタリング手法を提案したが，ランダムに負例を削除する場合よりも結果が悪かった．しかし，ランダムに負例を削除した場合でも共感推定の F 値は向上したことから，負例を削除するというアプローチは妥当であると考えられる．よりよい負例のフィルタリング手法を探究することが今後の重要な課題である．

第 5 章

ウェブからの人物の逸話の獲得

1.3 節では，システム主導の対話を実現するために，ユーザに提供する新しい話題に関連する一連の文を順次生成するアプローチを示した．また，ユーザに提供する話題として人物の逸話が適していることを述べた．

本章では，自由対話システムが対話を主導する際に提供する話題のデータベースを事前に構築することを目的とし，特定の人物の逸話をウェブから自動的に収集する手法を提案する．まず，5.1 節で，人物の逸話を用いた対話システムの構想について述べ，以降の節で人物の逸話を自動獲得する手法を詳述する．5.2 節では提案手法の概要を述べる．5.3 節では逸話の候補を抽出する手法について述べる．5.4 節では，獲得した逸話の候補文から不適切なものを削除するためのフィルタリングについて述べる．5.5 節では提案手法の評価実験について述べる．5.6 節では一つの情報源から逸話を収集した場合と比較し，本手法の意義を述べる．最後に 5.7 節で本章の成果を総括する．

5.1　人物の逸話を用いたシステム主導対話

本節では，人物の逸話のデータベースが用意されているとき，システム主導の対話を実現する手法の構想について述べる．

いま，ユーザとシステムがある人物に関して対話を進めていたとする．自由対話システムは現在の対話の話題を推定し (図 1.1 の話題推定の処理)，話題となっている人物，または話題と関係のある人物を決定する．例えば，ユーザとシステムがイチローについて話をしているときは，イチローの逸話を紹介する．また，サッカーについて雑談しているときは,「メッシ」や「本田」のようなサッカー選手の逸話を紹介する．対話システムが逸話を話し始めるタイミングは，ユーザ状態 (図 1.1 のユーザ状態推定の処理) から決定する．ユーザが現在の話題について，自分の知識を披露し終えた時や，ユーザが現在の話題に飽き始めている時，対話システムは話題の人物，または話題に関連した人物について「織田信長といえば，その後に天下を治めた豊臣秀吉は〜」のようにして逸話を話し始めるものとする．対

話システムが逸話を複数の発話に分けて提供し，一方ユーザは逸話の聞き役にまわることで，システム主導の対話が実現できると考えられる．

しかし，自由対話では様々な話題が取り上げられることを考慮すると，様々な人物についての逸話を対話システムのデータベースに収集する必要があるが，手作業での収集は非常に時間と労力がかかる．そこで，本研究では，膨大なコーパスであるウェブから逸話を自動的に獲得する手法を提案する．

5.2 人物の逸話の獲得手法の概要

本研究における人物の逸話の定義は図 5.1 に示す三つの要件を満たすテキストとする．人物の記録，賞罰，経歴など，人物に関する事実を説明したテキストは逸話とはみなさない．ただし，人物の事実と合わせて図 5.1 に示す条件を満たすテキストが付随する場合は，それら全体を逸話とみなす．

(a) その人物の性格や特徴を表わす話．
(b) あまり知られていない．
(c) 自由対話において，ユーザの興味を引く．

図 5.1: 逸話の定義

図 5.2 はウェブページにおける逸話の例である．この例では，赤い枠で囲まれたテキストがベートーベンの逸話となっている．

提案手法における処理の流れを図 5.3 に示す．まず，逸話を取得する対象となる人物名をクエリとしてウェブ検索を行い，人物名を含むウェブページの HTML ファイルを収集する．次に，HTML ファイルの中から逸話の候補となるテキストを抽出する．ウェブ検索ならびに逸話候補抽出処理の詳細は 5.3 節で述べる．最後に，抽出した候補の中から不適切なものを除外するフィルタリングの処理を適用する．フィルタリングの詳細は 5.4 節で述べる．

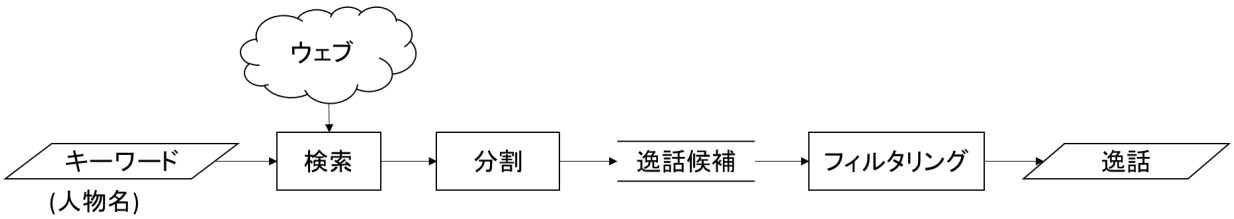

図 5.2: ウェブページ上の逸話の例

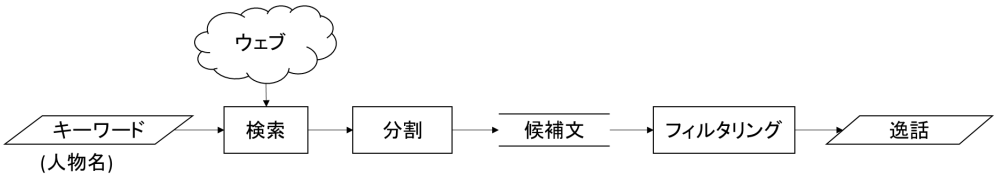

図 5.3: 提案手法の流れ

5.3 逸話候補の抽出

まず，式(5.1)をクエリとして，すなわち対象となる人物名と「逸話」というキーワードの組をクエリとして，BingAPI[1]による検索を行い，その人物に関するウェブページとして検索結果の上位 100 ページの HTML テキストを取得する．

$$[\text{人物名}] \ \& \ \text{逸話} \tag{5.1}$$

取得したウェブページについて，その HTML ソースファイルを解析し，Document Object Model (DOM) による木構造 (DOM ツリー) を得る．DOM ツリーとは，HTML ファイルにおける HTML タグの階層構造を木構造で表現したものである．DOM ツリーにおける内部ノードもしくは葉は DOM ノードと呼ばれる．DOM ノードは 1 つの HTML タグ，もしくは HTML タグで囲まれたテキストの断片に対応する．

次に，HTML テキストをいくつかのパッセージに分割する．ここでのパッセージの大まかな定義は，ひとつのトピックについて言及した 1 つまたは複数の段落であるとする．HTML タグはブロック要素 (⟨div⟩，⟨h1⟩，⟨p⟩ など) とインライン要素 (⟨span⟩，⟨a⟩，⟨b⟩ など) の二種類のタグに分類される．提案手法では，最小のブロック要素に対応する DOM ノードを検出し，その DOM ノードが支配するテキストをパッセージとして取り出す.「最小のブロック要素」とは，自身がブロック要素の HTML タグに対応し，かつその子孫にブロック要素を含まない DOM ノードである．

上記の手法で抽出されたパッセージはテキスト長が長く，2 つ以上の逸話を含む場合がある．そのため，以下の手続きにしたがってパッセージを分割する．検出された最小のブロック要素に対応する DOM ノード内のテキストを $\{t_1, \cdots, t_n\}$ とおく．t_i は HTML タグで分割されたテキストを示す．t_i が以下の条件のいずれかを満たすとき，その t_i をセパレータとしてパッセージを分割する．

- t_i は空白またはタブのみを含む.
- t_i の文字数が 5 以下である．この場合，t_i は完全な文ではなく，この前後で異なるパッセージが出現していると考えられる．

上記の手続きによって取得されたパッセージを逸話の候補とする．

[1]https://datamarket.azure.com/dataset/bing/search

5.4 フィルタリング

5.4.1 フィルタリングのためのルール

HTML テキストから抽出した逸話候補には逸話でないテキストが大量に混ざっており，これをフィルタリングする必要がある．本研究では，逸話でないパッセージを除外するためのルールとして，以下の $\mathbf{r_1}$ から $\mathbf{r_{11}}$ を用いる．これらのルールを全て適用しても除外されないパッセージを最終的な逸話として出力する．

$\mathbf{r_1}$: 見出しタグ

ウェブページにおける見出しは，そのウェブページの内容を端的に表わしていると考えられる．もし，見出しに「逸話」というキーワードが含まれていれば，その見出し以下のテキストには人物の逸話が書かれている可能性が高い．そこで，逸話候補の DOM ノードが (r1:a) の条件を満たさないとき，その逸話候補を除外する．また，(r1:a) の条件を満たすときでも，(r1:b)，(r1:c) のいずれかの条件を満たさないときは，その逸話候補を除外する．

(r1:a) n_a または n_s が見出しタグ ($\langle h1\rangle, \langle h2\rangle$ など) に対応する
(r1:b) ht のテキストが人物名を含む
(r1:c) ht のテキストが「逸話」または「エピソード」を含む

n_a および n_s は，それぞれ逸話候補の DOM ノードの先祖ノード，前に出現する兄弟ノードを表わす．また，ht は (r1:a) の条件を満たす見出しタグを表わす．

図 5.2 に示したウェブページを例にこのルールの働きを説明する．同図の点線で囲まれた見出しテキスト「ベートーベンの逸話」は，⟨h2⟩ タグでマークアップされており，かつその DOM ノードは赤線で囲まれたテキストの DOM ノードの先祖に該当する．人物名も「逸話」というキーワードも含み，$\mathbf{r_1}$ の条件を満たさないため，除外されずに逸話として残される．もし，h2 タグのテキストが人物名も「逸話」「エピソード」というキーワードも含まないときは，この逸話候補は削除される．

$\mathbf{r_2}$: テキストの主語

逸話候補のパッセージが人物に関する内容を表わしているかをチェックするため，人物名が文の主語となっているかを確認する．具体的には，パッセージの中に「⟨ 人物名 ⟩ が」や「⟨ 人物名 ⟩ は」という句を含む文が存在しないとき，その逸話候補を除外する．

$\mathbf{r_3}$: 一人称の単語

一人称が主語となる文は，ウェブページの著者の意見や感想が述べられており，人物の逸話ではないと考えられる．このルールは，パッセージの中に一

人称の代名詞 (私，僕，うち，俺) が存在するとき，その逸話候補を除外する．但し，一人称の代名詞が引用を表わす括弧 (‘「’ と ‘」’) 内に出現したとき，それは人物の台詞であるとみなして，例外的に除外しない．例えば,「私はベートーベンが好きだ」という文はウェブページの作成者の意見を表わしていると考えられるため除外する．一方,「ベートーベンは『私は音楽を愛している』と言った」という文を含む逸話候補は例外として除外しない．

r_4: 主観的表現

「〜と思う」といった文は，ウェブページの著者の意見や感想を述べていると考えられるため，人物の逸話とはみなせない．そこで，パッセージに「思う」という動詞が含まれるとき，その逸話候補を除外する．ただし，r_3 と同様に,「思う」が引用を表わす括弧内に出現したときは例外的に除外しない．

r_5: 依頼の表現

「お願い」「教えて」といった動詞が述語となっている文は，ウェブページの著者が何かを依頼していると考えられ，逸話とはみなせない．このルールは，依頼の表現「お願い」「教えて」が述語となっている文がパッセージに存在するとき，その逸話候補を除外する．日本語では述語は文末の近くに現れる．ここでは依頼表現が述語であるかを文の位置から判断する．依頼表現の位置を p，文の長さを l とし，$p/l \geq 0.8$ であれば，その依頼表現は文の述語であると判定する．

r_6: まとめ表現

「エピソード」を含み，文の後ろに「まとめる」「伝える」「紹介」「披露」のいずれかを含む逸話候補は除外する．例えば,「〜のエピソードを紹介します」のような導入文は，そのウェブページに人物の逸話が存在することを示唆するが，その文を含むパッセージ自体は逸話ではない．

r_7: 最小テキスト長

本論文では，ウェブから獲得した逸話は，対話システムからユーザに話題を提供し，システム主導の対話を実現するために用いることを想定している．また，この際，対話システムは逸話の文をひとつずつ順に生成することで対話を主導する．したがって，獲得する逸話は複数の文から構成され，ある程度の長さを持つことが要求される．そこで，テキストの長さが 50 文字未満の逸話候補は除外する．

r_8: 最大発話数

逸話はある程度の長さが必要となる一方で，あまりに長い逸話は，対話システムが延々と話を続ける事態を生じさせるため，システムがユーザに提供する話題としてふさわしくない．そこで，6 文以上で構成される逸話候補を除外する．

r_9: 非文

非文を含む逸話候補を除外する．ここでは，文末が“。”や“」”でない文は非文とみなす.

r_{10}: 先頭指示語の有無

指示詞がパッセージの先頭に出現するとき，その指示詞が指す対象はパッセージの前に存在すると考えられる．このようなパッセージを逸話として抽出したとき，指示詞が指す対象はパッセージに含まれていないため，パッセージを読むだけでは内容が理解できない可能性が高い．このルールでは,「この」という指示詞がパッセージの先頭に出現するとき，その逸話候補を除外する.

r_{11}: リンクタグ

逸話候補となるパッセージが ⟨a⟩ タグの中に含まれる場合，そのパッセージは広告リンクの内容を表わしていると考えられる．したがって，逸話候補の DOM ノードの先祖に ⟨a⟩ タグが存在するとき，その逸話候補を除外する.

5.4.2 抽出逸話数の制限

これまで述べた手法で逸話を獲得した予備実験では，逸話でないパッセージを誤って逸話として抽出する誤りが多かった．特に，1 つのウェブページから数多くのパッセージが逸話として誤抽出され，これが逸話抽出の精度を大きく低下させる要因となっていた．一方，1 つのウェブページに記載されている逸話はそれほど多くないと考えられる.

そこで，1 つのウェブページから抽出される逸話数に制限をかける手法を提案する．具体的には，1 つのウェブページから抽出された逸話の数を n とし，それが閾値 T_n よりも大きいとき，そのウェブページから抽出した全ての逸話を除外する．ウェブページによっては，ある人物の複数の逸話が並べて書かれていることもある．抽出逸話数に制限をかけることによって，このようなウェブページからは逸話が抽出されなくなる．言い換えれば，逸話抽出の再現率が低下することが予想される．しかし，自由対話システムのための知識として逸話を抽出する場合，再現率より精度が重視される．なぜなら，ウェブ上に存在する全ての逸話を網羅的に獲得する必要はない一方で，獲得した逸話に誤りが少ないことが望まれるからである．本研究では，1 つのウェブページから抽出される逸話の数に制限を設けることによって精度を向上させることを狙う.

5.5 評価

5.5.1 実験設定

提案手法による逸話獲得の性能を評価するため，表 5.1 に示す 5 人の人物を評価対象とした．それぞれの人物に対し，100 個のウェブページを取得した．この際に用いたクエリは，式 (5.1) に示した通り，人物名と「逸話」の組である．ただし，人物名は，人物の姓と名前を or で連結した論理式で表現する．クエリとして用いた人物名を表 5.1 の 2 列目に示す．ただし,「イチロー」は名前だけで記述されることが多いので，名前のみを人物名とした．次に，取得した 100 個のウェブページの中から，逸話に該当するパッセージを人手でマークアップした．正解の逸話のマークアップ作業は一人の作業者が実施した．表 5.1 の 3 列目に，それぞれの人物に対する正解の逸話数を示す．なお．同じ内容の逸話が重複して現われることがある．表 5.1 の正解逸話数はのべ数である．すなわち，同じ内容の逸話が複数存在することを許している．

表 5.1: データセット

人物	クエリ	逸話数
イチロー	イチロー	60
織田信長	織田 **or** 信長	85
ダヴィンチ	レオナルド **or** ヴィンチ	104
ベートーベン	ルートヴィヒ **or** ベートーベン	104
モーツァルト	ヴォルフガング **or** アマデウス **or** モーツァルト	65

逸話抽出の結果は，精度，再現率によって評価する．その定義を式 (5.2)，(5.3) に示す.

$$\text{精度} = \frac{\text{抽出された正解の逸話の数}}{\text{抽出された逸話の総数}} \tag{5.2}$$

$$\text{再現率} = \frac{\text{抽出された正解の逸話の数}}{\text{正解の逸話の総数}} \tag{5.3}$$

提案手法が正解の逸話を抽出できたかどうかは以下の基準により判定する．抽出された逸話 (パッセージ) の範囲が正解の逸話のそれと完全に一致するとき，その正解逸話を抽出できたとみなす．一方，抽出された逸話の範囲が正解の逸話を完全に包含するときも，その正解の逸話を抽出できたとみなす．抽出されたパッセージの中には逸話ではない文が含まれることになるが，本実験ではこの種の誤りは許容する．一方，抽出された逸話の範囲が正解の逸話の一部しか含まないとき，不正解とみなす．すなわち，逸話として完全なテキストを取得できないときは抽出失敗とする．また，提案手法によって抽出されたパッセージが複数の正解逸話を含むときがある．このとき，含まれる正解の逸話の数に関わらず，正解逸話をひとつだけ抽出できたとみなし，式 (5.2)，(5.3) における分子に 1 を足す.

表 5.2: 逸話抽出の結果

人物	ベースライン		提案手法	
	精度	再現率	精度	再現率
イチロー	0.179	0.083	0.429	0.056
織田信長	0.049	0.035	0.375	0.141
ダヴィンチ	0.692	0.173	0.238	0.048
ベートーベン	0.125	0.029	1.000	0.029
モーツァルト	0.345	0.154	0.167	0.015
マイクロ平均	0.232	0.093	0.348	0.057

5.5.2 結果

表 5.2 に提案手法とベースラインによる逸話抽出の精度と再現率を示す．ベースラインは，図 5.3 におけるフィルタリングの処理のところで，パッセージが人物名と「逸話」というキーワードの両方を含むときに逸話として検出するという単純な手法である．一方，提案手法では，5.4.1 項で述べた 11 個のルールを用いてパッセージのフィルタリングを行う．5 人の人物のマイクロ平均を比較すると，提案手法の精度はベースラインと比べて 11 ポイント高く，再現率は 4 ポイント低かった．前述したように，対話システムのための逸話抽出では再現率より精度が重視されるので，望ましい結果が得られている．

提案手法による逸話抽出の精度はベースラインを上回っているが，マイクロ平均で 34.8%と十分に高いとは言えない．精度が低い理由として，人物の逸話と単なる事実のテキストがフィルタリングによって十分に区別できていないことが考えられる．「イチローはメジャーリーガーである」という文は，対象となる人物に関連してはいるが，人が興味を引くような内容は含まれておらず，ただ事実を述べているだけである．このように，人物に関連してはいるが逸話ではないというテキストが数多く抽出されており，このことが精度が低い主な要因となっている．特に逸話の定義（図 5.1）の (b) 人にあまり知られていない，(c) 人の興味を引く，という観点から，逸話とそれ以外のテキストを区別することが困難である．このような誤りを除くためには，システムがテキストの内容を深く理解することが必要である．

提案手法の再現率も 1.5%〜14.1%と低い．その主な要因を調べたところ，複数の正解の逸話がひとつのパッセージとして取り出されていることがわかった．今回の実験では，抽出したパッセージに複数の逸話が含まれているときは，1 つの逸話が抽出されたとみなしているため，見かけ上再現率が低下する．この問題は，パッセージ分割のアルゴリズムを改良することで改善できる．ウェブページは様々なスタイルで記述されるため，パッセージの一般的な分割方法を探究することは難しい問題ではあるが，1 つの逸話を 1 つのパッセージとして抽出するためには不可

欠である．

また，それぞれの人物ごとの結果を比較すると，例えばベートーベンの逸話抽出精度は100%である一方，モーツァルトの逸話抽出精度は17%と人物によって大きく異なる．さらに，レオナルドとモーツァルトの結果はベースラインを下回っている．人物によって逸話の内容は当然異なり，また逸話の書かれ方も異なるが，提案手法ではその点を考慮していないためと考えられる．スポーツ選手，ミュージシャン，歴史上の人物など，人物のタイプに応じて適切な逸話抽出の手法を探究することが対策として考えられる．

次に，5.4.2 項で述べた，1 つのウェブページから抽出する逸話数に制限をかける手法を評価する．ここでは，ウェブページ当たりの抽出逸話数の閾値 T_n を 1，2，3 に設定したとき，および制限をかけないときの逸話抽出の精度を比較した．結果を表 5.3 に示す．$T_n = 1$ のとき，つまり 1 つのウェブページから 2 つ以上の逸話が抽出されたときにそれらを全て除外した場合，精度のマイクロ平均は 47.1%となり，1 つのウェブページから抽出される逸話数に制限を設けない場合に比べて精度がおよそ 12 ポイント向上した．また，再現率の結果を同様に表 5.4 に示す．T_n を 3 から 1 に減らしていくと，再現率も低下する．$T_n = 1$ のとき，再現率のマイクロ平均は 2.1%となり，制限を設けない場合に比べて，再現率はおよそ 4 ポイント低下した．しかし，再現率の低下よりも精度の向上の方が大きい．この結果から，人物に関連したテキストではあるが逸話ではないものを除く手法として，抽出される逸話の総数は少なくなるが，同一ウェブページから抽出した逸話の数に制限を設けることは有効であるとわかった．

5.6 Wikipedia からの逸話抽出の検討

本研究ではウェブ上のあらゆるページから人物の逸話を自動的に検出することを試みた．一方，人物の逸話をまとめたウェブサイトのみを獲得の対象とするアプローチも考えられる．特定のウェブサイトのみを対象とするならば，逸話に該当するパッセージを抽出することは容易である．人物に関する様々な情報が記載された代表的なウェブサイトとして Wikipedia が挙げられる．本節では，Wikipedia から逸話を抽出する手法の妥当性を検証する．

今回の検証では，表 5.5 に示す 5 分野 14 名の人物を対象とした．これらは 5.5 節の実験で用いた 5 名の人物を含む．これらの人物の日本語 Wikipedia のエントリを調べ，その中に書かれている逸話の数を調査した．その結果，14 名の人物のうち，Wikipedia 中で逸話のセクションが存在するのは 5 名であった．そのうち，モーツァルトは 4 個の逸話，イチローは 19 個の逸話の記述があった．

この結果から，Wikipedia では明示的に逸話として記述されているテキストは少なく，Wikipedia を情報源とするだけでは多くの人物についての逸話を獲得することは困難であることがわかった．したがって，本研究で取り組んだように，ウェブ上から広く逸話を検索・獲得する技術の確立が望まれる．なお，Wikipedia の中に

表 5.3: 閾値 T_n を変動させたときの逸話抽出の精度

人物	$T_n = 1$	$T_n = 2$	$T_n = 3$	制限なし
イチロー	0.400	0.500	0.429	0.429
織田信長	0.375	0.429	0.412	0.375
ダヴィンチ	1.000	1.000	0.250	0.238
ベートーベン	1.000	1.000	1.000	1.000
モーツァルト	0.500	0.167	0.167	0.167
マイクロ平均	0.471	0.484	0.432	0.348

表 5.4: 閾値 T_n を変動させたときの逸話抽出の再現率

人物	$T_n = 1$	$T_n = 2$	$T_n = 3$	制限なし
イチロー	0.033	0.050	0.050	0.056
織田信長	0.035	0.082	0.094	0.141
ダヴィンチ	0.010	0.010	0.010	0.048
ベートーベン	0.010	0.029	0.029	0.029
モーツァルト	0.015	0.015	0.015	0.015
マイクロ平均	0.021	0.037	0.040	0.057

表 5.5: 調査対象とした人物の一覧

分野	人物
医療	シュヴァイツァー，ナイチンゲール
音楽	モーツァルト，ベートーベン
歴史	ナポレオン，織田信長
スポーツ	イチロー，マイケル・ジョーダン
アニメ	ケンシロウ，ドラえもん
芸術	ダ・ヴィンチ，ゴッホ
芸能	石原裕次郎，ブルース・リー

は，逸話と明示されていなくても，逸話とみなせるパッセージが存在する場合があった．しかし，これらの逸話を抽出するためには，提案手法のような逸話の自動獲得技術が必要である．

5.7 まとめ

本章では，ウェブから指定された人物の逸話を抽出する手法を提案した．抽出された逸話は，対話システムが対話を主導する際に提供する話題として利用できる．提案手法では，まず，人物名と「逸話」の組をクエリとしてウェブ検索を行う．次に，得られたウェブページの DOM ツリーを解析し，ウェブページ上のテキストをいくつかのパッセージに分割し，個々のパッセージを逸話の候補とする．最後に，逸話候補の中から逸話でないものを除外するために，ルールベースのフィルタリングを適用する．5 名の人物を対象とした実験の結果，提案手法はベースラインに比べて逸話抽出の精度が 11 ポイント向上したことを確認した．さらに，1 つのウェブページから取得される逸話の数に制限をかけることにより，精度はさらに 14 ポイント向上した．エラー分析の結果，人物に関係しているが逸話ではないテキストを識別することが難しいことがわかった．この識別には深い自然言語理解が必要であるが，本論文では比較的単純な表層上の手がかりをもとに逸話を抽出する方法の効果を実験的に検証した．また，パッセージ分割の誤りも逸話抽出の誤りの原因となることがわかった．特に大量のウェブページから逸話を抽出するときは，パッセージ分割の誤りは無視できないため，改善を要する．

今後は，構文解析結果や意味解析結果を利用して，フィルタリングのためのルールを改善することが挙げられる．逸話によく出現する構文的パータン，意味的パターンを明らかにし，これを逸話以外のテキストを除外するルールとして実装する．逸話かそうでないかを区別するために機械学習の手法を用いることも検討に値する．パッセージ分割の性能を改善するために，ウェブページを DOM ツリーの構造によっていくつかのタイプに分類し，そのタイプに応じたパッセージ分割の手法を探究するアプローチが考えられる．

逸話の定義は再考が必要である．図 5.1 に示した 3 つの条件のうち，(b) と (c) はかなり主観的であり，人によって判断が分かれると考えられる．(b) の条件「あまり知られていない」については，ある逸話が広く世間に知られているかを人が客観的に判断するのは難しい．これを客観的に判断する方法としては，同じ逸話が多くのウェブページ上に出現する場合には著名であると判定したり，検索エンジンにおける順位が低い場合には著名ではないと判定することが考えられる．しかし，上記の基準によって本当に逸話が有名か否かを判断できるかは詳細に検討する必要がある．また，(c) の条件「人の興味を引く」についても，人によって判断が分かれやすいと考えられる．例えば，イチローの熱心なファンはイチローの多くの逸話を知っていると思われるが，そのような人はイチローの既知の逸話に興味を持たないだろう．また，野球に興味がない人にとっては，やはりイチローの

逸話を知りたいとは思わないだろう．逸話の定義が人の主観に強く依存しているとき，そのような曖昧なパッセージをウェブから自動的に獲得する手法を開発するのは難しい．今後は，自由対話システムでどのように利用するかも考慮に入れ，逸話の客観的な定義を設定する必要がある．

第 6 章

結論

6.1 本論文の貢献

本論文では，1 章で述べたように，自由対話システムにおける重要な研究課題，すなわち対話行為の自動分類，共感の推定，人物の逸話の自動獲得の 3 つの研究課題に取り組んだ．以下，それぞれの研究課題について，本論文の成果や得られた知見をまとめる．

対話行為自動推定

これまでの教師あり機械学習による対話行為自動推定手法では，複数の対話行為を判別する分類器の学習を一つの素性タイプのセットにより行うことが多かった．そのため，個々の対話行為から見ると，その対話行為の判別に最適な素性タイプのセットが用いられていないため，正解率が低くなる．このことが全体の対話行為推定の正解率の低下を招いていた．

本研究では，この問題に対し，対話行為毎に発話がその対話行為を持つか否かを判定する二値分類器を学習し，複数の対話行為に対する二値分類器の判定結果から入力発話の対話行為を決定する手法を提案した．提案手法は，個々の対話行為毎に発話がその対話行為に該当するかを判定する第 1 段階と，第 1 段階の結果から最終的に最も適切な対話行為を選択する第 2 段階から構成される．提案手法では，第 1 段階で用いる分類器を学習する際，対話行為毎に最適な素性タイプのセットを実験的に決定する．第 2 段階では，個々の分類器の判定の信頼度を比較して最終的な対話行為を選択する手法の他に，更に推定精度を向上させるために，信頼度を学習素性とした分類器を学習する手法，判定の信頼度に重み付けを行う手法，判別の難しい対話行為の組に対してそれらのいずれかを選択する分類器を学習する手法を提案した．

評価実験の結果，対話行為を区別せずに素性タイプの選択を行う手法と比べて，提案手法の対話行為推定の F 値は 0.6 ポイント高かった．この差は統計的に有意であることが確認された．また，以下の知見を得た．

- 有効な素性タイプのセットは対話行為によって異なることを実験的に確認し，対話行為毎に素性タイプの最適化を行う提案手法のアプローチが有望であること
- 過去の対話行為を素性とするとき，その最適な長さは対話行為毎に異なること
- 第 2 段階において，本論文が提案した手法がいずれも対話行為推定の F 値の向上に貢献したこと
- 自由対話システムでは様々なトピックが話題になることから，対話によって有効な素性タイプが異なる可能性があり，そのことが素性タイプの最適化を実験的に行う提案手法のアプローチの問題点となっていること

共感の推定

本研究では，与えられた発話が相手への共感を示しているかを推定する手法を提案した．対話コーパスを分析し，共感の有無を判定するための学習素性をいくつか考案した．また，対話における共感発話と非共感発話の分布に著しい偏りがあるため，訓練データから冗長な負例を削除する手法を提案した．

評価実験の結果，提案手法による共感推定の F 値は 0.18 となり，ベースラインを 3 ポイント上回った．また，以下の知見を得た．

- 組み合わせ素性を用いたとき，ベースラインと比べて共感推定の F 値が大きく向上すること
- 提案した素性タイプの中で，話者交代，自立語の繰り返しが共感発話の推定に有効であること
- 負例のフィルタリングは共感推定の F 値の向上に寄与すること

ウェブからの人物の逸話の獲得

本研究では，指定された人物の逸話をウェブから抽出する手法を提案した．ウェブ検索で得られたウェブページの DOM ツリーを解析し，ウェブページ上のテキストをいくつかのパッセージに分割し，個々のパッセージを逸話の候補とする．次に，本研究で作成した 11 個のルールを用いて逸話候補の中から逸話でないものを除外する．

評価実験では 5 名の人物を対象とした逸話抽出を行った．提案手法は表層的な手がかりだけに基づく比較的単純なものであるが，その効果を実験的に検証した．その結果，以下の知見が得られた．

- 単純なキーワードマッチングを用いたベースラインに比べて，提案手法による逸話抽出の精度が 11 ポイント向上した．

- 1 つのウェブページから取得される逸話の数に制限をかけることにより，精度が 14 ポイント向上した．
- 提案手法では，人物に関係しているが逸話ではないテキストを識別することが難しいことがわかった．

6.2 今後の課題

対話行為の推定に関する今後の課題としては,「フィラー」「確認」「要求」の対話行為推定の F 値を改善することが挙げられる．本研究の動機は,「自己開示」のような推定精度が比較的高い対話行為がある一方，推定の F 値が低い対話行為も存在するという問題を解決することであった．しかし，上記の 3 つの対話行為についてはその目標が達成されていない．したがって，今後まず取り組むべき課題としたい．また，本論文の評価実験で得られた重要な知見は，対話行為の現われ方は対話の内容に依存するということである．そのため，一般に訓練データとテストデータでは対話の内容が異なるが，そのことが対話行為の推定を難しくしている大きな要因となっている．訓練データとテストデータの内容が異なるために機械学習の性能が低下する問題は広く知られており，それを解決するために領域適応と呼ばれる手法が研究されている．よって，領域適応の技術を対話行為の自動推定に応用することで，対話行為推定の F 値が改善されることが期待できる．3.6.6 項で述べたように，対話行為毎に適切な特徴のセットを設定するという提案手法のアプローチの妥当性を検証するためには更なる実験が必要である．また，自然言語処理分野でも近年盛んに利用されるようになった深層学習 [25] との比較も重要である．

共感推定については，共感発話の数が非共感発話と比べて一般に著しく小さいことが最も重要な問題である．言い換えれば，訓練データにおける正例と負例の不均衡を是正することが重要である．本研究では，アンダーサンプリングが共感推定の F 値の向上に貢献することを確認した．しかし，本研究で提案した負例のフィルタリング手法は，互いに類似している冗長な負例を訓練データから削除するというものであったが，ランダムに削除する負例を選択する手法よりも結果が悪かった．つまり，アンダーサンプリングをどのように実現するかについては，まだ探究の余地が残されている．また，訓練データにおける正例と負例の数が著しく異なるという問題は，共感推定以外にもよく起こる．共感推定に限らず，一般の分類問題にも適用可能な負例のフィルタリング手法の探究にも取り組みたい．

ウェブからの人物の逸話の獲得に関して解決するべき重要な課題は，パッセージ分割の誤りの削減と，逸話のフィルタリングの改善である．ウェブページのパッセージ分割は，一般に難しい課題である．提案手法は DOM ツリーや HTML タグのみを手がかりとしているが，パッセージ分割の目的は同じトピックについて言及した文のまとまりを検出することであるため，テキストの内容を解析して利用

する手法を検討する必要がある．逸話のフィルタリングも難しい課題である．特に，人物に関連のあるパッセージが，逸話であるのか，単に事実を述べたものであるのかを判定するのは，パッセージの意味理解が必要不可欠である．一方，テキストをどのように意味解析すれば逸話か否かを判定できるかは，現在では不明のままである．このように逸話の自動獲得はチャレンジングなタスクであるが，今後取り組んでいきたい．

また，本研究では自由対話システムにおける 3 つの要素技術の開発に取り組んだが，これらを自由対話システムに組み込んだとき，対話システムの質の向上にどの程度寄与するかは確認していない．よって，今後は自由対話システム全体を実装し，その性能を評価する必要がある．ユーザとシステムとで対話を行い，その質や満足度をユーザに主観的に評価してもらう被験者実験を行う予定である．自由対話システムの評価を通じて，本研究で取り組んだ要素技術についても，今後解決するべき新たな課題が見つかるものと考えられる．

謝辞

本研究を行うにあたり，多くの方々に協力して頂いたので，ここに感謝の意を表す．はじめに，担当教員として多くの助言，指導を頂いた北陸先端科学技術大学院大学准教授白井清昭先生に深謝する．次に，本論文を審査して下さった北陸先端科学技術大学院大学東条敏教授，池田心准教授，長谷川忍教授，(株）ホンダ・リサーチ・インスティチュート・ジャパン　シニア・リサーチャの船越孝太郎氏に深く感謝する．池田心准教授には副テーマの指導教員としても多くの助言をいただいた．深謝する．最後に研究に協力して頂いた全ての方に感謝の意を表し，謝辞とする．

参考文献

[1] Timothy W. Bickmore and Rosalind W. Picard. Establishing and maintaining long-term human-computer relationships. *ACM Trans. Comput.-Hum. Interact.*, Vol. 12, No. 2, pp. 293–327, 2005.

[2] Leo Breiman. Bagging predictors. *Machine Learning*, Vol. 24, No. 2, pp. 123–140, 1996.

[3] Leo Breiman. Random forests. *Machine Learning*, Vol. 45, No. 1, pp. 5–32, 2001.

[4] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. Smote: Synthetic minority over-sampling technique. *J. Artif. Int. Res.*, Vol. 16, No. 1, pp. 321–357, 2002.

[5] Chang Chih-Chung and Lin Chih-Jen. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, Vol. 2, No. 3, pp. 27:1–27:27, 2011.

[6] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. LIBLINEAR: A library for large linear classification. *The Journal of Machine Learning Research*, Vol. 9, pp. 1871–1874, 2008.

[7] 福田正治. 共感と感情コミュニケーション (1) 共感の基礎. 研究紀要, No. 36, pp. 45–58, 2008.

[8] Takayuki Hasegawa, Nobuhiro Kaji, Naoki Yoshinaga, and Masashi Toyoda. Predicting and eliciting addressee's emotion in online dialogue. In *The 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 964–972, 2013.

[9] Ryuichiro Higashinaka, Kenji Imamura, Toyomi Meguro, Chiaki Miyazaki, Nozomi Kobayashi, Hiroaki Sugiyama, Toru Hirano, Toshiro Makino, and Yoshihiro Matsuo. Towards an open-domain conversational system fully based on natural language processing. In *Proceedings of the 25th International Conference on Computational Linguistics*, pp. 928–939, 2014.

[10] Ryuichiro Higashinaka, Nozomi Kobayashi, Toru Hirano, Chiaki Miyazaki, Toyomi Meguro, Toshiro Makino, and Yoshihiro Matsuo. Syntactic filtering and

content-based retrieval of Twitter sentences for the generation of system utterances in dialogue systems. In *Proceedings of the 5th International Workshop on Spoken Dialog Systems*, pp. 113–123, 2014.

[11] Ryuichiro Higashinaka, Toyomi Meguro, Hiroaki Sugiyama, Toshiro Makino, and Yoshihiro Matsuo. On the difficulty of improving hand-crafted rules in chat-oriented dialogue systems. In *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pp. 1014–1018, 2015.

[12] Shinsuke Higuchi, Rafal Rzepka, and Kenji Araki. A casual conversation system using modality and word associations retrieved from the web. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, pp. 382–390, 2008.

[13] 土方嘉徳. 利用者の好みをとらえ活かす-嗜好抽出技術の最前線-：1. 嗜好抽出・情報推薦の基礎理論 1）嗜好抽出と情報推薦技術. 情報処理, Vol. 48, No. 9, pp. 957–965, 2007.

[14] 平野徹, 小林のぞみ, 東中竜一郎, 牧野俊朗, 松尾義博. パーソナライズ可能な対話システムのためのユーザ情報抽出. 人工知能学会論文誌, Vol. 31, No. 1, pp. DSF–B_1–10, 2016.

[15] 張恵芳. 自然会話における「ヨネ」の意味類型と表現機能. 言語学論叢, Vol. 28, pp. 17–32, 2009.

[16] Nobuo Inui, Toshiaki Ebe, Bipin Indurkhya, and Yoshiyuki Kotani. A case-based natural language dialogue system using dialogue act. In *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, pp. 193–198, 2001.

[17] 磯村直樹, 鳥海不二夫, 石井健一郎. 対話エージェント評価におけるタグ付与の自動化. 電子情報通信学会論文誌. A, 基礎・境界, Vol. 92, No. 11, pp. 795–805, 2009.

[18] 伊東昌子, 永田良太. 談話場における相互行為の構築に関わる文末詞の修辞機能. 認知科学, Vol. 14, No. 3, pp. 282–291, 2007.

[19] Dan Jurafsky, Elizabeth Shriberg, and Debra Biasca. Switchboard swbd-damsl shallow-discourse-function annotation coders manual. *Institute of Cognitive Science Technical Report*, pp. 97–102, 1997.

[20] Daniel Jurafsky, Elizabeth Shriberg, and Debra Biasca. Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual. Technical report, Institute of Cognitive Science Technical Report, 1997.

[21] 柏岡秀紀, 翠輝久, 水上悦雄, 杉浦孔明, 岩橋直人, 堀智織. 観光案内への音声対話システムの活用. デジタルプラクティス, Vol. 3, No. 4, pp. 254–261, 2012.

[22] 河原達也, 川島宏彰, 平山高嗣, 松山隆司. 対話を通じてユーザの意図・興味を探り情報検索・提示する情報コンシェルジェ. 情報処理, Vol. 49, No. 18, pp. 912–918, 2008.

[23] 河内彩香. 日本語の雑談の談話における話題展開機能と型. 早稲田大学日本語教育研究, Vol. 3, pp. 41–55, 2003.

[24] Su Nam Kim, Lawrence Cavedon, and Timothy Baldwin. Classifying dialogue acts in one-on-one live chats. In *Proceedings of EMNLP*, pp. 862–871, 2010.

[25] Yoon Kim. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1746–1751, 2014.

[26] 桐山伸也, 広瀬啓吉, 峯松信明. 話題知識を導入した文献検索音声対話システム. 電子情報通信学会論文誌. D-2, Vol. 85, No. 5, pp. 863–876, 2002.

[27] 小林峻也, 萩原将文. ユーザの嗜好や人間関係を考慮する非タスク指向型対話システム. 人工知能学会論文誌, Vol. 31, No. 1, pp. DSF–A_1–10, 2016.

[28] 国立国語研究所. 日本語自然会話書き起こしコーパス (旧名大会話コーパス), 2001. 科学研究費基盤研究 (B)(2) 「日本語学習辞書編纂に向けた電子化コーパス利用によるコロケーション研究」(平成 13 年度〜15 年度), https://nknet.ninjal.ac.jp/nknet/ndata/nuc/.

[29] Iolanda Leite, André Pereira, Samuel Mascarenhas, Carlos Martinho, Rui Prada, and Ana Paiva. The influence of empathy in human-robot relations. *International Journal of Human-Computer Studies*, Vol. 71, No. 3, pp. 250 – 260, 2013.

[30] 前田英作, 南泰浩, 堂坂浩二. 人ロボット共生におけるコミュニケーション戦略の生成. 日本ロボット学会誌, Vol. 29, No. 10, pp. 887–890, 2011.

[31] 目黒豊美, 東中竜一郎, 杉山弘晃, 南泰浩. 意味属性パターンを用いたマイクロブログ中の発言に対する自動対話行為付与. 研究報告音声言語情報処理（SLP）, Vol. 2013, No. 1, pp. 1–6, 2013.

[32] Toyomi Meguro, Yasuhiro Minami, Ryuichiro Higashinaka, and Kohji Dohsaka. Learning to control listening-oriented dialogue using partially observable markov decision processes. *ACM Trans. Speech Lang. Process.*, Vol. 10, No. 4, pp. 1–20, 2014.

[33] Dmitrijs Milajevs and Matthew Purver. Investigating the contribution of distributional semantic information for dialogue act classification. In *Proceedings of the 2nd Workshop on Continuous Vector Space Models and their Compositionality (CVSC)*, pp. 40–47, 2014.

[34] 南泰浩, 東中竜一郎, 堂坂浩二, 目黒豊美, 森啓, 前田英作. 対話行為タイプ列 Trigram による行動予測確率に基づく POMDP 対話制御. 電子情報通信学会論文誌. A, 基礎・境界, Vol. 95, No. 1, pp. 2–15, 2012.

[35] 翠輝久, 河原達也, 正司哲朗, 美濃導彦. 質問応答・情報推薦機能を備えた音声による情報案内システム. 情報処理学会論文誌, Vol. 48, No. 12, pp. 3602–3611, 2007.

[36] 翠輝久, 水上悦雄, 志賀芳則, 川本真一, 河井恒, 中村哲. ユーザの相づち・うなずきを喚起する音声対話システム (対話生成, ＜特集＞人とエージェントのインタラクション論文). 電子情報通信学会論文誌. A, 基礎・境界, Vol. 95, No. 1, pp. 16–26, 2012.

[37] 水上雅博, Lasguido Nio, 木付英士, 野村敏男, Graham Neubig, 吉野幸一郎, Sakriani Sakti, 戸田智基, 中村哲. 快適度推定に基づく用例ベース対話システム. 人工知能学会論文誌, Vol. 31, No. 1, pp. DSF–C_1–12, 2016.

[38] 水野淳太, 乾健太郎, 松本裕治. ウェブニュースを利用した雑談対話システム. 言語・音声理解と対話処理研究会 (SIG-SLUD), 人工知能学会研究会資料, SIG-SLUD, 第 55 巻, pp. 1–6, 2009.

[39] 西田公昭. 対話者の会話行為が会話方略ならびに対人認知に及ぼす効果. *The Japanese Journal of Phychology*, Vol. 63, No. 5, pp. 319–325, 1992.

[40] 大西可奈子, 吉村健. コンピュータとの自然な会話を実現する雑談対話技術. NTT DoCoMo テクニカル・ジャーナル, Vol. 21, No. 4, pp. 17–21, 2014.

[41] 大竹裕也, 萩原将文. 高齢者のための発話意図を考慮した対話システム. 日本感性工学会論文誌, Vol. 11, No. 2, pp. 207–214, 2012.

[42] 関野嵩浩, 井上雅史. 発話に対する拡張談話タグ付与. 第 6 回情報処理学会東北支部研究会報告, 2010.

[43] 柴田雅博, 冨浦洋一, 西口友美. 雑談自由対話を実現するための WWW 上の文書からの妥当な候補文選択手法. 人工知能学会論文誌, Vol. 24, No. 6, pp. 507–519, 2009.

[44] Andreas Stolcke. SRILM – an extensible language modeling toolkit. In *Proceedings of ICSLP*, pp. 901–904, 2002.

[45] Hiroaki Sugiyama, Toyomi Meguro, Ryuichiro Higashinaka, and Yasuhiro Minami. Open-domain utterance generation for conversational dialogue systems using web-scale dependency structures. In *Proceedings of the 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 334–338, 2013.

[46] 角薫, 角康之, 間瀬健二, 中須賀真一, 堀浩一. 個人の概念空間を利用した興味の推定による情報提供(知能情報メディア論文特集). 電子情報通信学会論文誌. D-II, 情報・システム, II-パターン処理, Vol. 82, No. 10, pp. 1634–1644, 1999.

[47] Kazuko Takahashi, Hiroya Takamura, and Manabu Okumura. Estimation of class membership probabilities in the document classification. In *Proceedings of the 11th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD)*, pp. 284–295, 2007.

[48] Shota Takeuchi, Tobias Cincarek, Hiromi Kawanami, Hiroshi Saruwatari, and Kiyohiro Shikano. Construction and optimization of a question and answer database for a real-environment speech-oriented guidance system. In *Proceedings of COCOSDA*, 2007.

[49] Joseph Weizenbaum. Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, Vol. 9, No. 1, pp. 36–45, 1966.

[50] Bo Xiao, Dogan Can, Panayiotis G. Georgiou, David Atkins, and Shrikanth S. Narayanan. Analyzing the language of therapist empathy in motivational interview based psychotherapy. In *Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, 2012.

[51] Rui Yan, Yiping Song, and Hua Wu. Learning to respond with deep neural networks for retrieval-based human-computer conversation system. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '16, pp. 55–64, 2016.

[52] Show-Jane Yen and Yue-Shi Lee. Cluster-based under-sampling approaches for imbalanced data distributions. *Expert Systems with Applications*, Vol. 36, No. 3, Part 1, pp. 5718 – 5727, 2009.

[53] Koichiro Yoshino, Shinsuke Mori, and Tatsuya Kawahara. Spoken dialogue system based on information extraction using similarity of predicate argument structures. In *Proceedings of the 12th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 59–66, 2011.

[54] Yang Zhaojun, Li Baichuan, Zhu Yi, King Irwin, Levow Gina, and Meng Helen. Collaborative filtering model for user satisfaction prediction in spoken dialog system evaluation. In *Spoken Language Technology Workshop (SLT), 2010 IEEE*, pp. 472–477, 2010.

[55] Zhaohui Zheng, Xiaoyun Wu, and Rohini Srihari. Feature selection for text categorization on imbalanced data. *SIGKDD Explor. Newsl.*, Vol. 6, No. 1, pp. 80–89, 2004.

[56] 国立国語研究所. 分類語彙表. 大日本図書, 2004.

第 A章

対話行為推定の実験結果の補足

A.1　素性タイプ選択の過程

9 つの対話行為のそれぞれについて，図 3.2 のアルゴリズムで素性タイプを選択する過程を表 A.1 から A.9 に示す．これらの表では，各ステップで残された素性タイプをチェックで示している．素性タイプ選択に要する反復回数は対話行為毎に異なるが，1〜8 回の反復で収束している．

表 A.1: 「自己開示」の分類器における素性タイプの選択過程

step	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{18}	f_{19}	f_{20}	f_{21}	f_{22}	f_{23}	f_{24}	f_{25}	f_{26}	f_{27}	f_{28}
1	✓			✓	✓	✓	✓		✓	✓	✓	✓	✓	✓				✓	✓	✓	✓		✓	✓	✓	✓	✓	✓
2	✓				✓	✓	✓		✓		✓		✓	✓				✓	✓	✓	✓		✓	✓	✓	✓	✓	✓
3	✓				✓	✓					✓		✓					✓	✓	✓	✓						✓	✓
4	✓				✓	✓	✓		✓		✓		✓	✓				✓	✓	✓	✓		✓	✓	✓		✓	✓
5	✓				✓	✓	✓		✓				✓	✓				✓	✓	✓	✓		✓				✓	✓
6	✓				✓	✓	✓		✓				✓	✓				✓	✓	✓	✓		✓	✓	✓		✓	✓

表 A.2: 「質問 (YesNo)」の分類器における素性タイプの選択過程

step	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{18}	f_{19}	f_{20}	f_{21}	f_{22}	f_{23}	f_{24}	f_{25}	f_{26}	f_{27}	f_{28}
1	✓		✓		✓		✓				✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓		✓	✓	✓
2	✓												✓	✓		✓	✓		✓	✓	✓					✓	✓	✓
3	✓		✓		✓		✓				✓	✓	✓	✓	✓	✓	✓		✓		✓	✓		✓		✓	✓	✓
4	✓				✓		✓				✓	✓	✓	✓	✓		✓		✓		✓	✓		✓		✓		✓
5	✓				✓		✓				✓	✓	✓	✓	✓	✓	✓		✓		✓	✓		✓		✓	✓	✓
6	✓				✓		✓				✓	✓	✓	✓	✓	✓	✓		✓		✓	✓				✓	✓	✓
7	✓				✓		✓				✓			✓	✓		✓		✓		✓	✓				✓		✓
8	✓				✓		✓				✓	✓		✓	✓		✓		✓		✓	✓				✓	✓	✓

表 A.3: 「質問 (What)」の分類器における素性タイプの選択過程

step	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{18}	f_{19}	f_{20}	f_{21}	f_{22}	f_{23}	f_{24}	f_{25}	f_{26}	f_{27}	f_{28}
1	✓		✓		✓	✓	✓				✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓			✓
2	✓		✓		✓						✓		✓	✓	✓	✓	✓			✓	✓	✓		✓				✓

表 A.4: 「応答 (YesNo)」の分類器における素性タイプの選択過程

step	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{18}	f_{19}	f_{20}	f_{21}	f_{22}	f_{23}	f_{24}	f_{25}	f_{26}	f_{27}	f_{28}
1	✓		✓		✓		✓	✓				✓	✓			✓	✓	✓		✓	✓	✓	✓	✓	✓			✓
2			✓					✓				✓	✓			✓	✓	✓			✓	✓	✓	✓	✓			✓
3	✓		✓				✓	✓				✓	✓			✓	✓	✓		✓	✓	✓	✓	✓	✓			✓
4	✓		✓				✓						✓			✓	✓	✓			✓		✓	✓	✓			✓
5	✓		✓				✓						✓			✓	✓	✓			✓							✓

表 A.5: 「応答 (平叙)」の分類器における素性タイプの選択過程

step	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{18}	f_{19}	f_{20}	f_{21}	f_{22}	f_{23}	f_{24}	f_{25}	f_{26}	f_{27}	f_{28}
1	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓		✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
2	✓					✓		✓	✓	✓	✓	✓				✓		✓	✓	✓	✓	✓	✓	✓	✓			✓
3	✓					✓			✓	✓	✓	✓				✓		✓	✓	✓		✓	✓	✓				✓

表 A.6: 「あいづち」の分類器における素性タイプの選択過程

step	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{18}	f_{19}	f_{20}	f_{21}	f_{22}	f_{23}	f_{24}	f_{25}	f_{26}	f_{27}	f_{28}
1	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓		✓	✓	✓	✓
2	✓		✓		✓	✓		✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓		✓		✓	✓		
3					✓			✓	✓	✓	✓	✓				✓	✓	✓	✓	✓	✓				✓			
4	✓		✓		✓			✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓		✓		✓	✓		
5	✓		✓		✓					✓	✓	✓			✓			✓	✓	✓	✓		✓		✓	✓		

表 A.7: 「フィラー」の分類器における素性タイプの選択過程

step	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{18}	f_{19}	f_{20}	f_{21}	f_{22}	f_{23}	f_{24}	f_{25}	f_{26}	f_{27}	f_{28}
1	✓		✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
2	✓		✓	✓	✓							✓	✓		✓	✓	✓	✓	✓	✓	✓	✓			✓	✓		✓
3	✓		✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
4	✓			✓	✓	✓	✓				✓	✓	✓	✓		✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
5	✓			✓	✓	✓					✓		✓	✓		✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
6	✓			✓	✓	✓					✓		✓	✓		✓	✓	✓	✓	✓		✓	✓			✓		✓

表 A.8: 「確認」の分類器における素性タイプの選択過程

step	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{18}	f_{19}	f_{20}	f_{21}	f_{22}	f_{23}	f_{24}	f_{25}	f_{26}	f_{27}	f_{28}
1	✓		✓		✓	✓	✓			✓	✓		✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓		✓	✓	✓
2	✓		✓		✓	✓	✓			✓			✓	✓	✓	✓	✓		✓	✓	✓	✓		✓		✓		✓
3	✓		✓		✓	✓				✓			✓	✓	✓	✓			✓	✓	✓	✓		✓		✓		✓
4	✓		✓		✓	✓				✓			✓		✓	✓				✓				✓		✓		
5	✓		✓		✓	✓				✓					✓					✓				✓		✓		

表 A.9: 「要求」の分類器における素性タイプの選択過程

step	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{18}	f_{19}	f_{20}	f_{21}	f_{22}	f_{23}	f_{24}	f_{25}	f_{26}	f_{27}	f_{28}
1	✓			✓	✓	✓	✓		✓		✓	✓	✓	✓	✓	✓	✓		✓		✓	✓		✓	✓	✓	✓	

A.2 対話行為推定の対応表

提案手法のうち最も F 値の高い Pro_b が選択した対話行為と正解の対話行為の対応表を表 A.10 に示す.

表 A.10: 正解の対話行為と Pro_b の出力との対応表

(正解) \ (Pro_b の出力)	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
d_1	4622	22	24	2	94	65	18	15	10
d_2	93	466	28	0	2	1	1	28	0
d_3	55	37	252	0	5	3	2	3	0
d_4	5	0	0	184	9	8	3	0	0
d_5	106	5	1	6	655	2	4	2	0
d_6	149	1	2	13	6	649	100	9	0
d_7	73	1	6	5	9	79	203	0	0
d_8	213	79	3	0	17	15	1	128	0
d_9	67	2	0	0	11	0	0	0	18

d_1:自己開示，d_2:質問 (YesNo)，d_3:質問 (What)，d_4:応答 (YesNo)，
d_5:応答 (平叙)，d_6:あいづち，d_7:フィラー，d_8:確認，d_9:要求

本研究に関する発表論文

ジャーナル論文 (査読あり)

[1] 福岡知隆，白井清昭： 対話行為に固有の特徴を考慮した自由対話システムにおける対話行為推定, 自然言語処理，24(4)，pp.523-546，2017.

国際会議論文 (査読あり)

[2] Kiyoaki Shirai, Tomotaka Fukuoka. “Acquisition of Anecdote from Web for Free Conversation System”. Asean Japan Joint-Workshop on Computational Linguistics & Informatics, 2017.

[3] Tomotaka Fukuoka, Kiyoaki Shirai. “Identification of Sympathy in Free Conversation”. The 29th Pacific Asia Conference on Language, Information and Computation (PACLIC 29), pp.1-9, 2015.

国内発表 (査読なし)

[4] 福岡知隆，白井清昭： 対話行為毎の特徴に応じた対話行為自動推定, 言語処理学会第 22 回年次大会, pp.1121-1124, 2016.

[5] 福岡知隆，白井清昭： チャットシステムのための共感発話の推定，言語処理学会第 20 回年次大会, pp.765-768, 2014.

[6] 福岡知隆，白井清昭: 自由対話システムにおける対話行為の自動推定 - 個々の対話行為に応じた学習素性の設計，NLP 若手の会 第 9 回シンポジウム, 2014.