

会話の表層構造に着目した対話エージェントの設計

Design of the agent system focused on the surface structure of conversation

小林 司† 森井 昌克†
Tsukasa Kobayashi Masakatu Morii

1 はじめに

近年、高機能携帯端末の普及や音声認識・音声合成技術の向上により自然言語による対話インターフェースが注目を集めている。自然言語は人間がコミュニケーションに用いる基本的な手段である。そのため自然言語の利用はユーザにとっての利便性を向上させるが、ユーザがインターフェースに対し人間同様の受け答えを期待する可能性がある。そのため対話インターフェースの中には、人間を模した返答を行うものが多く見られる。このような人間を模した会話を行うプログラムを対話エージェントと呼ぶ。

多くの対話エージェントはインターフェースの一部を成している。そのため、携帯端末の操作や予約の受付など、特定のタスク実行に関する会話を重視している。そのため、特定のタスクに関して会話を行う対話エージェントは多く研究されている。一方、特定のタスクを持たない会話である雑談については、あまり研究が行われていない。しかし人間が会話を行う際は、特定のタスクについてのみ話すとは限らない。むしろ店の店員など特定のタスクを持つ人間であっても、その他の話題について雑談を行うことが少なくない。この雑談には相手に親しみを感じさせる効果があると考えられる。そのため、対話エージェントが雑談を行うことは、ユーザに親しみを感じさせる上で重要であると考えられる。

そこで本研究では対話エージェントに雑談を行わせる方法を考える。多くの対話エージェント設計段階で想定されていない話題に関する会話を苦手としており、想定外の話題が現れやすい雑談を行うことは困難である。しかし人間は想定外の話題に対しても、相手に話を合わせることで会話を継続することができる。対話エージェントも人間のように相手に話を合わせることで出来れば、想定外の話題に対しても適切な返答を行うことができると考えられる。本研究では会話の表層構造を表す DAMSL タグを利用して、相手に話を合わせるような応答を行う手法を提案する。提案手法は辞書を用いる対話エージェントを対象とする。これらの対話エージェントは、辞書から一つの発言を選択することで応答を行う。提案手法はこの応答選択に DAMSL タグを利用する。提案手法を用いることで、辞書を用いる対話エージェントの雑談機能を拡張できると考えられる。

本稿ではまず 2 章にて対話エージェント全般について述べる。その後 3 章にて DAMSL について述べる。4 章では提案手法である DAMSL タグを用いた応答選択手法について述べる。5 章では提案手法に対する階層型 DAMSL の導入について述べる。6 章では本稿の内容をまとめる。

2 対話エージェント

本章では対話エージェントについて述べる。対話エージェントは入力されたユーザの発言に対して、自然言語による返答を行うプログラムである。入力される発言に制限はなく、あらゆる入力に対して返答を行うことが期待される。入出力を音声によって行う音声対話エージェントも存在するが、本研究ではテキストによる入出力を前提とする。対話エージェントは返答の

生成方法により幾つかの種類が存在するが、最も広く用いられているのは辞書型対話エージェントである。

辞書型対話エージェントは、あらかじめ作成された辞書内の発言を用いて返答を行う。辞書にはキーワードとそれに対応する返答用発言が登録されている。エージェントはユーザから発言を入力された時に、ユーザの発言中にいずれかのキーワードが含まれるかどうかを検査する。そしてキーワードが含まれた場合は、そのキーワードに対応する返答を行う。辞書型対話エージェントはあらかじめ人手で作成した辞書に依存して会話を行う。そのため設計者の意図したとおりに動作する可能性が高く、利用しやすい特徴がある。

辞書型対話エージェントの性能は、辞書の構成に大きく左右される。辞書の質は主に以下の要素から決定される。

- キーワードと発言の組み合わせが適切であるか
- キーワードが会話中に現れる発言を網羅できるか
- 十分な数のキーワードと発言が存在するか

このうち会話の網羅については、用いるキーワードの長さが大きく影響する。長いキーワードを用いることで、状況に応じた適切な返答を行うことができる。しかし長いキーワードを用いる場合、ユーザの発言中にキーワードが含まれず返答が行えないという問題が発生しやすい。一方短いキーワードを用いればより確実に返答を行うことができる。しかし状況と合わない返答を行う可能性が高くなる。またキーワードに対する適切な返答の設定も困難となる。したがって辞書を用いて自然な会話を行うためには、適切な長さのキーワードと返答をバランスよく含む辞書を作成する必要がある。しかしこれは辞書作成者に高い技術や経験と大きな労力が求められるため、容易ではない。

またユーザの発言中に複数のキーワードが含まれた場合には、いずれのキーワードに対応する発言を行うか選択する必要がある。多くのエージェントでは、短いキーワードと長いキーワードが共に含まれた場合は長いキーワードを優先する。これは長いキーワードに対する発言の方が、より状況にあっていると考えられるためである。しかしキーワードの重要性は必ずしも長さだけから決定できるわけではない。またキーワードの長さが等しい場合には、上記の選択手法は利用できない。この場合には、無作為にキーワードを選択して応答を決定する等の手段を取らざるを得ない。

一般に、雑談では、現れる話題を前もって推測することが困難である。そのため話題に合わせた詳細なキーワードと返答を用意することができない。したがって主に短いキーワードを用いて会話を行うことになるが、この場合はキーワードの重複が発生しやすくなるため、適切な応答選択が困難となる。そのため、多くの対話エージェントは雑談を苦手としている。

短いキーワードを活用して会話を行うため、発言の内容ではなく形式に着目する対話エージェントも古くより存在する [1]。しかしこれらのエージェントの応答手法は言語の構造や設計者の経験則に基づいており、多くの対話エージェントで容易に利用できるものではない。

3 DAMSL

本研究では会話の表層構造分析のために提案された手法である DAMSL を用いる。3.1 節では基本的な DAMSL である SWBD-DAMSL について述べる。また 3.2 節で DAMSL の対話

† 神戸大学大学院工学研究科, Graduate School of Engineering, Kobe University

エージェントへの適用について述べる。

3.1 SWBD-DAMSL

SWBD-DAMSL(Switchboard-DAMSL)[2] は会話文の表層的な構造分析のために提案された手法である。SWBD-DAMSL では会話中の各発言に対し、DAMSL タグを割り振る。DAMSL タグは各発言の意味ではなく、会話における役割に応じて決定される。タグには相槌や Yes/No 疑問文など 40 種類以上が存在する。ただし一般的な会話文における各タグの出現率には大きな偏りがあり、会話中に非常に多く現れるものほとんど現れないものが存在する。会話中の全発言にタグを割り振りその遷移関係や構成比などを調べることで、対象の会話を分析することができる。

SWBD-DAMSL は米国で提案された手法であるため、英語の会話文を対象としている。そのため DAMSL タグの中には他の言語では適切でないものも存在する。しかし本手法のアイデアは、日本語を含め他の多くの言語でも有効だと考えられる。

3.2 対話エージェントへの応用

機械的に発言の意味理解を行うことは非常に困難であるため、ほとんどの対話エージェントは発言の意味理解無しに返答を行なっている。そのために文末表現など発言の形式に関する表現をキーワードとし、それに応じた返答を行うことも多い。これは各発言の意味を無視して表層的な役割に注目するという DAMSL の特徴と類似しているといえる。

多くの対話エージェントは辞書に強く依存するため、応答手法は辞書の各製作者の個性に影響され、一般性は非常に低い。しかし DAMSL を用いることで、対話エージェントの応答手法を一般化できると考えられる。

そのため、対話エージェントに DAMSL を適用する研究も既に行われている。磯村らは、DAMSL を用いて対話エージェントの評価を行なっている [3]。磯村らはこの研究で、対話エージェントの評価手法を提案している。この手法では、まず対話エージェントの行った会話に対して DAMSL タグを付与する。これを人間同士の会話に DAMSL タグを付与したものと比較し、タグの遷移関係の類似度を対話エージェントの性能としている。

この手法では DAMSL を対話エージェントの評価に用いているが、同様の手法で対話エージェントの応答選択に DAMSL を用いることはできない。これは、対話エージェントのリアルタイム性のためである。

会話の評価を行う場合、タグの付与は会話終了後に行う。したがってタグの付与に会話系列全体を用いることができる。しかし対話エージェントの応答選択に DAMSL を利用する場合、ユーザの発言に対してすぐにタグを付与する必要がある。そのため、会話系列全体を利用することはできない。さらにタグ付与を高速に行う必要もある。したがって DAMSL を用いて対話エージェントを設計するためには、異なる手法が求められる。

4 雑談を行う対話エージェント

本研究の目標は、任意の雑談を行う対話エージェントの設計である。そのために、人間が相手に話を合わせる方法を模倣する。まず、人間がどのように相手に話を合わせているか考える。人間は相手の発言中に知らない単語や聞き取れない単語が存在した場合でも、話を続けることがある。中には、不明な単語について、意味を尋ねる、推測するなどの対応を行うこともあるが、それが困難な時も多い。その場合、相手の発言の内容が不明なまま返答を行っている。この場合は、相手の主張に対して同意を示す、質問に対してはひとまず肯定する、どちらとも取れない答えを返すなどの対応を行うことが多い。これは、相手が期待している返答の形式を推測し、それに応じた返答を行っていると考えられる。

ここで期待している返答の形式とは、相手の主張に対する同意や、相手の質問に対する肯定、否定などを表す。これら返答

表 1 提案手法の DAMSL タグ一覧

DAMSL タグ (略称)	訓練コーパス内出現率 (%)
能動的通常発言 (NS)	37.2
指示, 提案 (ND)	2.5
5W1H の疑問文 (QW)	5.0
Yes/No 疑問文 (QY)	9.2
同意を求める発言 (QA)	2.7
その他の疑問文 (QO)	1.8
相槌 (NB)	10.3
相槌的疑問文 (QB)	3.2
感謝 (GT)	1.4
謝罪 (GA)	1.0
あいさつ (GF)	10.8
肯定 (RY)	3.3
否定 (RN)	2.2
曖昧な答え (RF)	0.9
相手への同意 (RA)	2.6
理解・納得 (RU)	1.9
その他 (OT)	3.9

の形式は、DAMSL と非常に類似しているといえる。そこで本研究では、DAMSL タグの遷移情報を用いることで相手に話を合わせて雑談を行う手法を提案する。4.1 節では提案手法について述べる。4.2 節では提案手法におけるタグ付与手法について述べ、4.3 節では提案手法の評価を行う。

4.1 DAMSL を利用した対話エージェントの設計

提案手法は辞書を用いる対話エージェントを対象としており、辞書からの応答選択に DAMSL を利用する。辞書にはキーワードと発言に加えて、その発言の DAMSL タグが記述されている。またあらかじめ人間同士の会話文から DAMSL タグの遷移確率を求め、遷移確率表として利用する。提案手法ではユーザの発言から DAMSL タグを求め、遷移元タグする。その後遷移元タグに対する返答として適切なタグを遷移確率表から求め、そのタグに従い返答する。なお、提案手法では日本語対話エージェント用に調整した DAMSL を用いる。提案手法で用いる DAMSL の一覧を表 1 に示す。

提案手法の流れは以下のとおりである。

- ユーザの発言に DAMSL タグ (遷移元タグ) を付与する
- 遷移確率表を参照し、遷移元タグからの遷移確率が高いタグ (遷移先タグ) を求める
- 辞書から遷移先タグを持つ発言を取り出し、返答とする。このとき、可能であればキーワードも考慮する

ここで遷移先タグは、人間同士の会話において遷移元タグの直後によく現れていたタグであり、これがユーザの期待する返答の形式であると考えられる。遷移先タグにしたがって返答を決定することで、ユーザに話を合わせる返答を実現する。以下に提案手法の詳細を述べる。

- ユーザの発言に対する DAMSL タグ付与

ユーザの発言に対して DAMSL タグを付与する。これは DAMSL タグの遷移関係を用いて、ユーザが期待している返答の形式を求めるためである。このタグ付与はユーザとの会話中に行う必要がある。そのため、サポートベクターマシン (以下 SVM) を用いた自動分類によりタグを付与する。SVM によるタグ付与については 4.2 節にて述べる。これにより付与したユーザ発言の DAMSL タグを遷移元タグとして扱う。

- 遷移確率表に基づく遷移先タグの判定

遷移確率表を参照し、遷移元タグから遷移する DAMSL タグを判定する。これを遷移先タグとして扱う。遷移先タグは、遷

移行タグの直後に現れる可能性の高いタグを表している。これはユーザ発言に対する返答としてふさわしいタグであると考えられる。

まず遷移確率表について述べる。遷移確率表はあらかじめ DAMSL タグを付与した人間同士の会話文より作成する。この会話文において、各タグの直後に現れているタグをそれぞれ求め、あるタグから別のタグへの遷移確率を求める。遷移確率の算出には隣接する発言のタグのみを考慮し、その前後の発言のタグは考慮しない。これにより求めた遷移確率を全て記録し、遷移確率表とする。

次に遷移先タグの判定について述べる。遷移先タグを判定する際には遷移確率表を参照し、遷移元タグから遷移する可能性のあるタグを求める。これらのタグに対し重み付き抽選を行い、選ばれたタグを遷移先タグとする。抽選の重みには遷移確率を用いているため、あるタグが遷移先タグとして選ばれる確率は遷移元タグからの遷移確率に等しくなる。なお、後述する返答文の決定において、選択された遷移先タグに対応する返答が存在しない場合があるため、遷移先タグは上位 3 個まで候補を選択している。

● 遷移先タグに基づく返答文の決定

遷移先タグを基にして、対話エージェントの行う返答を決定する。返答の発言は応答用の辞書より取得する。辞書はキーワード、タグ、発言から構成される。ここでキーワードとはユーザの発言中にこの文字列が含まれた場合、該当する発言を行うためのものである。これは会話中の特定の状況において、その状況に適した発言を行うために用いる。一方、辞書の中にはキーワードを持たない発言も登録する。これはより多くの状況で用いることのできる汎用性の高い発言である。また、タグは該当する発言の DAMSL タグを表している。発言は対話エージェントが返答として用いるためのものである。対話エージェントの中には、辞書の発言をそのまま用いるのではなく、一部の単語を置換する、順序を並べ替える等の処理を行うものもあるが、本研究ではこれらの処理はおこなっていない。しかし DAMSL タグが変化しない処理であれば、提案手法と併用することは可能である。

返答決定の流れは以下のとおりである。

1. 辞書より、遷移先タグと一致するタグを持つキーワード、発言の組を全て取り出す。
2. 取り出した組のうち、ユーザの発言中に含まれるキーワードを持つ組を選択する。
3. 2 で該当する組が複数存在する場合は最も長いキーワードを持つ組を選択する。
4. その後も該当する組が複数存在する場合は、無作為に 1 つの組を選択する。
5. 2-4 により 1 つの組を選択できた場合には、その組の発言を返答とする。
6. 2 で該当する組が存在しなかった場合、タグに関係なくユーザの発言中に含まれるキーワードを持つ組を選択する。
7. ユーザの発言中にいずれのキーワードも存在しなかった場合、キーワード無しで登録されている組のうち、遷移先タグと一致するタグを持つ発言を全て選択する。
8. 7 で該当する組が複数存在した場合、無作為に 1 つの発言を選択する。
9. 7-8 により 1 つの発言を選択できた場合には、その発言を返答とする。
10. 1 つも組が存在しなかった場合、遷移先タグを第 2 候補以降に変更する。

4.2 SVM を用いた DAMSL タグの自動付与手法

提案手法ではユーザの発言に対して、その場で DAMSL タグを求めて付与する必要がある。そのために SVM による自動分類を行う。本研究では SVM ライブラリとして libsvm[4] を用いている。SVM は 2 クラス分類を行う機械学習型の分類器だが、

提案手法ではこれを拡張して複数タグの分類に用いている。ここでは libSVM の多クラス分類機能を利用する。多クラス分類の別の手法についてはさらに 5.1 にて述べる。SVM ではあらかじめ人手によって分類された訓練データを利用する。入力された分類対象データを訓練データと比較し、訓練データにおけるいずれの分類に属する可能性が高いか判断する。

提案手法では、人手で DAMSL タグを付与された会話群が訓練データとなる。訓練データには、1 発言毎に 1 つの DAMSL タグを付与しておく。また SVM には発言そのものではなく、そこから抽出したパラメータを入力する。提案手法では直前のタグ、単語 1-gram、単語 2-gram、単語繰り返し数をパラメータとして用いる。以下で各パラメータについて述べる。

● 直前のタグ

直前のタグとは、直前に行われた発言に付与された DAMSL タグのことである。1 発言に付与される DAMSL タグは 1 つであるため、直前のタグは 1 つだけである。直前のタグのパラメータは、ある発言とその直前の発言の間に因果関係が存在することを前提としている。例えば、Yes か No を問う発言の直後には肯定や否定の発言が行われることが多いとすれば、この発言間には因果関係があると言える。これは本研究の前提条件であり、多くの場合成立すると考えられる。

● 単語 n-gram

単語 n-gram とは、n 個の連続する単語を 1 つの単位としたものである。提案手法では、あらかじめ訓練コーパス中の全発言から全単語 n-gram を数え上げ、各 n-gram の各分類における出現頻度を求めておく。これを用い対象の発言中の各 n-gram に対して、訓練コーパスにおける各分類での出現頻度を調べ、分類毎にそれらの和を求めてパラメータとしている。単語 n-gram のパラメータは、各分類毎に出現単語に特徴があることを前提としている。この前提が成り立つ時、出現単語の構成から対象の発言がいずれの分類に属するかを推測できると考えられる。提案手法では単語 1-gram と 2-gram を用いる。ただし助詞については直前の単語の一部として数えている。

● 単語繰り返し数

単語繰り返し数とは、直前の発言と重複する単語の数を指している。日本語においては、相手の発言中の単語を繰り返すことが特別な意味を持つ場合がある。例えば、「もうお昼食べた?」という疑問文に対して、「食べた。」と答えることで肯定の意を表すことができる。ここでは相手の発言中の動詞を繰り返すことで、相手の発言を肯定している。単語の繰り返しは単語 n-gram によって表すことのできない特徴であるため、別のパラメータとして設定している。

4.3 DAMSL を利用した対話エージェントの評価

提案手法により対話エージェントが適切に雑談を行えるようになるかどうか評価実験を行う。提案手法は辞書を用いる対話エージェントに対して、DAMSL タグを利用した応答選択を追加したものといえる。そこで DAMSL タグを利用する対話エージェントと、利用しない対話エージェントの 2 つを用意し、比較実験を行う。

以下では DAMSL タグを利用する対話エージェントを提案手法、利用しない対話エージェントを既存手法と呼ぶ。提案手法、既存手法ともに用いる辞書は同じであり、DAMSL タグの利用を除けばアルゴリズムも同一である。

評価実験では、提案手法と既存手法を 6 人の被験者に利用してもらい、以下の 4 つの質問に答えてもらった。また Web チャットルーム等から収集した 4240 個の発言から成る訓練コーパスを用いた。なお被験者に対しては、利用する 2 つのエージェントのうちいずれが提案手法であるかは示していない。

1. こちらの話を聞いているか
2. 快適に会話できるか
3. エージェントが人間らしく感じられるか
4. 総合的にどちらが優れているか

表 2 提案手法の評価結果

質問項目	提案手法	既存手法	同じ
1 話を聞く	2	3	1
2 快適に会話	2	2	2
3 人間らしさ	0	2	4
4 総合評価	3	3	0

ここで「こちらの話を聞いているか」という質問は、対話エージェントがユーザの発言に応じた返答を行えているかどうか確認するためのものである。提案手法はユーザに対して話を合わせるような応答を重視している。そのため提案手法においては最も重要な項目であると言える。

「快適に会話できるか」という質問は、対話エージェントとの会話によってストレスが発生しないか確認するためのものである。対話エージェントの主な利用用途はユーザ支援であるため、快適性は対話エージェントの価値に直結する。提案手法はユーザと話が噛み合わないという問題を避けられる半面、当たり障りのない発言が多いという印象を与えやすい。そのため、この質問により提案手法が本当に雑談を潤滑に行う助けになるかどうかを調べることができる。

「エージェントが人間らしく感じられるか」という質問は、対話エージェントが人間と同じような応答を行えるかどうか確認している。提案手法は話を合わせるという人間の会話方法を模倣しているため、既存手法よりも人間らしく感じられると期待できる。

「総合的にどちらが優れているか」という質問は、対話エージェント全体の印象を問うものである。対話エージェントに期待する要素はユーザ毎に異なるため、より多くのユーザの要望を満たすことができるという汎用性を確認しているとも言える。

なお各質問にはその項目において、提案手法と既存手法のいずれがより優れているかを選んでもらった。どちらも差が感じられない場合のために、どちらも同じという選択肢も設けている。

本実験の結果は表 2 の通りとなった。ここで「提案手法」、「既存手法」の値は、該当する質問項目において、その手法の対話エージェントの方が優れていると答えた人数である。「同じ」の値は、該当する質問項目において、2 つの対話エージェント間に有意の差は無いと答えた人数である。表 2 から、提案手法と既存手法の間に有意の差は見られなかった。この原因としては、以下が考えられる。

- 分類が正しく行われないことが多かった
- 各タグに含まれる発言が多すぎて、会話を制御できない

分類については、大部分の発言が「能動的通常発言 (NS)」か「あいさつ (GF)」として判定されていた。NS と GF は訓練コーパス内での出現率の高いタグであり、これが過剰に多様な特徴を内包してしまったため、他のタグに分類すべき発言まで NS として分類されてしまったと考えられる。

また各タグに含まれる発言については、一部のタグの範囲が広すぎるため、適切なタグが得られても適切な返答が行えない状況が見られた。例えば「どこに遊びに行くの?」という発言の適切なタグは「5W1H 疑問文」である。遷移確率表において、「5W1H 疑問文」からの遷移確率の高いタグは「能動的通常発言」であるため、「私は大学生です。」という「能動的通常発言」タグの発言は、適切なタグに基づく返答である。しかしこれは明らかに不自然な返答である。そのため、DAMSL タグによって適切な返答を行うためには、より詳細なタグが必要であると考えられる。

上記の問題を解決するために提案手法の改良を行う。改良した提案手法については、5 章で述べる。

5 階層分類型対話エージェント

4.3 節の評価実験では、提案手法によって適切な雑談が行えるようにはならなかった。この原因としては分類の精度が十分ではなかったことに加え、各タグに含まれる発言が多岐に渡り、遷移先タグが適切であっても適切な返答が行われないことが多かったためであると考えられる。上記の問題を解決するため、タグ項目及び分類手法を改良する。5.1 節ではタグ項目及び分類手法の改良について述べる。5.2 節では改良した対話エージェントの評価を行う。

5.1 タグ項目と分類手法の改良

改良前の提案手法では各タグが多く種類の発言を含んでいたため、会話を十分に制御できていなかった。この問題を解決するために、DAMSL タグの項目を改良する。改良後の DAMSL タグは表 5 の通りである。改良後のタグは、3 層の階層構造となっている。これは SVM の精度が不十分で厳密な分類が難しい場合に、タグ情報を部分的に得るためである。その他に、後述する分類の際に処理時間を軽減するためでもある。また表 5 のタグは改良前の表 1 のタグと比べ、5W1H に関するタグを中心にタグの個数が増加している。これは各タグに含まれる発言を減らし、遷移先タグの選択によって返答を適切に制御するためである。

次に分類手法の改良について述べる。改良前の提案手法では、多クラス分類の為に libSVM の多クラス分類機能を用いていた。しかしここでは 2 クラス分類を繰り返し、多数決を取ることで分類を行う。ここで 2 クラス分類とは、対象の発言が 2 つのタグのうちいずれに近い特徴を持つかを求めることを指す。ただしこの手法を単純に行う場合、全タグの組み合わせの数だけ分類を行う必要がある。これは処理時間の大幅な増加を招くため、階層構造を利用して処理時間の軽減を図る。

分類の流れは以下のとおりである。

1. 最上位階層から開始
2. 現在の階層で、共通の親を持つ全てのタグの組み合わせを作成
3. 作成した全ての組で SVM による 2 クラス分類を行う
4. 最も多く選ばれた分類を現在の階層における分類として決定
5. 決定された分類の直下にて、2 以降を繰り返す
6. 最下位階層の分類が決定すると終了

上記の分類手法では、全タグの組み合わせの数だけ分類を行うのはタグ数の少ない最上位階層のみである。以後は上位階層で選ばれたタグの直下に属するタグ間のみで分類が行われる。そのため、上位階層で誤った分類が行われた場合は全て誤った分類を行ってしまうが、処理時間は大幅に軽減される。提案手法はユーザとの会話中に分類を行う必要があるため、処理時間の軽減を優先している。

5.2 階層型 DAMSL を利用する対話エージェントの評価

改良後の提案手法では、上記の通り階層型 DAMSL タグ項目とそれに基づく分類手法を導入した。改良した提案手法によって適切に雑談が行えるようになったかどうか評価実験を行う。評価実験では 4.3 節と同様に提案手法と既存手法の比較を行う。ここで提案手法は階層型 DAMSL を用いる対話エージェントである。既存手法については 4.3 節と同じものを用いている。また訓練コーパスは人間同士のテキスト会話文から 618 発言を用いている。

評価実験では、提案手法と既存手法を 17 人の被験者に利用してもらい、以下の 5 つの質問に答えてもらった。

1. こちらの話を聞いているか
2. 話題を継続できているか
3. 快適に会話できるか
4. エージェントが人間らしく感じられるか

表 3 階層型 DAMSL タグ一覧

大分類 (略称)	中分類 (略称)	小分類 (略称)	訓練コーパス内出現率 (%)		
能動的通常発言 (N)	通常発言 (NS)	意見を含む通常発言 (NSO)	19.3		
		意見を含まない通常発言 (NSN)	12.6		
	指示 (ND)	命令・依頼 (NDO)	1.8		
		提案 (NDS)	1.6		
		疑問文 (Q)	肯定・否定可能な発言 (QY)	肯定・否定を問う疑問文 (QYY)	9.7
	5WIH(QW)	同意を求める疑問文 (QYA)	4.5		
		理由を問う疑問文 (QWY)	2.3		
		選択を問う疑問文 (QWH)	0.0		
		人物を問う疑問文 (QWO)	1.0		
		時を問う疑問文 (QWN)	0.2		
		量を問う疑問文 (QWW)	0.8		
		説明を求める疑問文 (QWD)	6.1		
		返答発言 (R)	ポジティブ返答 (RP)	肯定 (RPY)	6.1
		同意 (RPA)		8.4	
理解・納得 (RPU)	1.9				
	ネガティブ返答 (RN)	否定 (RNN)	7.4		
		曖昧な答え (RNF)	1.6		
		通常形相槌 (RBS)	4.9		
	相槌 (RB)	疑問形相槌 (RBQ)	3.1		
		5WIH 返答 (RW)	理由に関する返答 (RWY)	0.6	
			選択に関する返答 (RWH)	0.0	
	人物に関する返答 (RWO)		0.3		
	時に関する返答 (RWN)		0.0		
	量に関する返答 (RWW)		0.2		
	求められた説明を行う (RWD)		1.0		
儀礼的定型句 (G)	あいさつ (GF)		時間に依存するあいさつ (GFT)	1.1	
その他あいさつ (GFO)		0.8			
	お礼 (GS)	感謝 (GST)	0.6		
		謝罪 (GSA)	0.8		
		その他 (OOO)		0.2	

表 4 階層型 DAMSL 利用エージェントの評価結果

質問項目	提案手法	既存手法	同じ
1 話を聞く	5	9	3
2 話題の継続	6	6	5
3 快適に会話	3	7	7
4 人間らしさ	6	8	3
5 総合評価	6	10	1

5. 総合的にどちらが優れているか

なお本実験では話題の継続に関する質問を行っているが、これは提案手法が DAMSL タグを重視するため、話題変更の頻度が上昇することが懸念されたためである。話題変更が頻繁に行われることはユーザのストレスにつながるため、話題を継続できる方が望ましいと言える。

評価実験の結果を表 4 に示す。

表を見ると、まず「こちらの話を聞いているか」という質問に対して、多くの被験者が既存手法の方が優れていると答えたことが分かる。提案手法はユーザに話を合わせるような受け答えを行うため、ユーザの話を聞いているように感じられるはずである。この結果から提案手法が有効に働いていなかったことが分かる。実際の会話を確認すると、ほとんどの発言がごく一

部のタグに分類されていた。特に多く見られたのは「Yes/No の疑問文」、「同意を求める発言」、「意見を含む能動発言」である。一方、5WIH に関するタグなどには全く分類されなかった。上記のタグはいずれも「肯定」、「否定」、「相槌」などのタグに遷移しやすいタグである。そのため、対話エージェントがユーザの発言に対して単調な答えばかり返してしまい、まじめに話を聞いていないという印象を与えてしまったと思われる。

話題の継続については、提案手法と既存手法の間に有意の差は見られなかった。そのため、提案手法により話題変更の頻度が増えることは無かったと思われる。また「快適に会話できるか」と「エージェントが人間らしく感じられるか」の 2 つの質問についても既存手法のほうが優れているという答えが多かった。これは提案手法が単調な答えばかりを行った結果、提案手法の対話エージェントへの印象が悪くなったためであると思われる。

上記の問題点に対しては、十分な量の訓練コーパスを用意することが効果的であると思われる。これは訓練コーパスの量が増えることで、出現頻度の低いタグについても十分な特徴を求めることができるためである。また多くのタグで現れる単語 n-gram が出現頻度の高いタグの特徴として扱われるという問題も防ぐことができる。

ただし訓練コーパスは人手で発言を分類する必要があるため、大量の訓練コーパスを作成するのは困難である。そこで訓練

表 5 分類辞書利用エージェントの評価結果

質問項目	提案手法	既存手法	同じ
1 話を聞く	8	1	1
2 話題の継続	4	0	6
3 快適に会話	9	1	0
4 人間らしさ	8	1	1
5 総合評価	9	1	0

コーパスが必要となる SVM を用いず、分類辞書によって分類を行った場合、提案手法に有効性が見られるかどうか調べた。ここで用いる分類辞書は一般的な応答用の辞書と同じくキーワードの項目を持つが、発言の項目は持たず、分類タグの項目を持つ。ユーザの発言中にキーワードが現れた場合、ユーザの発言を分類タグに指定された DAMSL タグとして判定する。また分類辞書はキーワードの重みを表す項目を持つ。ユーザの発言中に複数のキーワードが現れた場合は、キーワードの出現位置と重みからキーワード得点を求める。キーワードの出現位置については、より文末に近い位置に存在するキーワードに大きな得点を与えている。これは 1 つの発言内に複数のタグの要素が含まれる場合、より後ろにある要素ほどその発言に対する返答に影響しやすいと考えられるためである。これらによりキーワード得点を求め、最も得点が高いキーワードに対応する発言を対話エージェントの発言とする。

本手法を用いた対話エージェントについても評価実験を行った。本評価実験も、提案手法と既存手法の比較実験である。提案手法は階層型 DAMSL を用いる対話エージェントだが、分類に SVM ではなく分類用の辞書を用いている。既存手法については 4.3 節と同様である。10 人の被験者に提案手法と既存手法を利用してもらい、SVM を用いた場合と同様の 5 つの質問に答えてもらった。

評価実験の結果を表 5 に示す。

この実験では多くの被験者が、提案手法の方が話をよく聞いていたと答えている。このことから提案手法によりユーザの発言に応じた返答が行えたことが分かる。また、話題の継続を除く他の質問に対しても、提案手法が有効に働いたことが読み取れる。

なお、話題の継続に関しては提案手法による変化は小さいと考えられる。この質問には、提案手法と既存手法に差異が無いと答えた被験者が多かったが、提案手法の方が優れていると答えた被験者もいた。これは提案手法が話題の継続に関しても有効であるためではなく、他の要素から作られた印象によるものと思われる。

本評価実験より提案手法が対話エージェントの雑談に有効であることが分かった。

6 まとめ

本研究では、対話エージェントが雑談を行うために、DAMSL の遷移に注目する手法を提案した。提案手法は応答に辞書を利用する対話エージェントを対象としている。提案手法では、ユーザの発言を分類して DAMSL タグを付与し、遷移元タグとする。その後人間同士の会話において、遷移元タグの直後に現れやすいタグを求め、遷移先タグとする。この遷移先タグを持つ発言を辞書から探して返答とする。これにより人間が相手に話を合わせるかのように雑談を行うことができる。

提案手法では、通常の DAMSL ではなく、日本語対話エージェント用に調整した DAMSL を用いる。本稿ではタグ数の少ない DAMSL と、それを改良して 3 層の階層構造とした DAMSL の 2 つを用いた。

本稿ではさらに実験によって提案手法の評価を行った。評価

実験は辞書利用型のエージェントと、これに提案手法を組み込んだエージェントとの比較実験とした。タグ数の少ない DAMSL を用いた場合については、評価実験により提案手法の有効性を確認することはできなかったが、階層構造を持つ DAMSL を用いることで、提案手法の有効性を確認できた。

ただし提案手法では、ユーザの発言を分類して DAMSL タグを付与するため、SVM を用いることを前提としていた。しかし、SVM を用いた場合については、提案手法は有効に働かなかった。

これは SVM に用いた訓練コーパスのサイズが小さかったことが大きな原因と思われる。今回の訓練コーパスでは出現回数の極端に低いタグが複数存在し、これらの特徴がほとんど得られなかった。また多くのタグが共通して持つはずの特徴が、出現率の高いタグの特徴として扱われてしまっていた。これらの理由により、SVM が誤ったタグ付与を行ってしまったと思われる。

SVM ではなく分類用の辞書を用いてタグ付与を行うことで、提案手法により適切な雑談を行うことができた。そのため、DAMSL の遷移に注目して返答を決定する提案手法の考え方は、雑談を行う対話エージェントの設計に有効であると言える。

今後は、十分な訓練コーパスが存在する場合の SVM によるタグ付与の評価を行う必要がある。また、少ない訓練コーパスでも正しいタグを付与するための分類手法やタグ項目の構築も今後の課題である。

参考文献

- [1] Joseph Weizenbaum, ELIZA - A Computer Program for the Study of Natural Language Communication between Man and Machine, *communications of the Association for Computing Machinery* 9, pp.36-45, 1966 年.
- [2] D.Jurafsky, E.Shriberg, D.Biasca, Switchboard SWBD-DAMSL Shallow-DiscourseFunction Annotation Coders Manual, Institute of Cognitive Science Technical Report, pp.97-02, 1997 年.
- [3] 磯村直樹, 島海不二夫, 石井健一郎, HMM による非タスク指向型対話システムの評価, 電子情報通信学会論文誌 D Vol. J92-D No.4 pp.542-551, 2009 年.
- [4] C.-C. Chang, C.-J. Lin, LIBSVM : a library for support vector machines, *ACM Transactions on Intelligent Systems and Technology* Volume 2 Issue 3, April 2011 Article No. 27 , 2011 年.