The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

文章中の潜在要素を考慮した対話システム

A Dialogue System Implemented with Latent Parameters

李為達 *1 Edward Li　日永田智絵 *2 Chie Hieida　長井隆行 *2*3 Takayuki Nagai

*1聖光学院中学校高等学校 Seiko Gakuin High School
*2電気通信大学 The University of Electro-Communications
*3大阪大学 Osaka University

When given a conversation, traditional dialogue systems mainly focus on the context that can be observed on the surface of sentences; concretely, they process and determine the output based on the grammar, visible keywords and structure of the sentence. However, the content we convey to others is affected by a multitude of latent parameters, such as emotional state, personal knowledge and personality. Therefore, we have attempted to validate the integrity of a dialogue system which takes these latent parameters into measure, and have sucessfully developed a dialogue system which utilizes latent parameters as input.

1. はじめに

近年，深層学習の発展に伴い，様々な状況における人間との対話を目的とした対話システムの開発が進んでいる．従来のシステムでは，話し手が与える入力文の表層的なコンテクストに着目し，文章の構造やキーワードに重点を置いて出力を生成するようなモデルが多い．一例として，二つの LSTM（Long short-term memory network）[Hochreiter97] をエンコーダーとデコーダーとしてつなげた Sequence to Sequence モデル [Sutskever14] を使用した対話システムが挙げられる [Csaky17]．Sequence to Sequence モデルは，シーケンスのペアを大量に学習することで片方のシーケンスからもう一方を生成できる性質を利用して，文章の表層的なコンテキストに着目した対話を実現することができる（図 1）．さらに，こうした対話システムを応用して返答の質を上げた事例として，2017 年にアマゾンが開催した「Alexa Prize Competition」においてモントリオール大学の研究室が開発した対話システム「MILABOT」が挙げられる．これは，今までに開発されたモデルを複数集め，各モデルの出力の中で一番質が良い返答を出力として使用するものである [Serban17].

これらの対話システムは，入力文の表面的な要素のみに注目しており，人同士の会話において重要な役割を果たす「感情」などの潜在的な要素を陽に考慮していない．そのため，それらを無意識に感じる人間のように会話を行うことが難しい．そこで本稿では，入力文から抽出できる潜在要素を陽に算出し，それらをモデルの入力として使用した対話システムを構築し，その有用性を検証することを目標とする．

2. 会話中の潜在要素

2.1 人間の意思決定プロセス

対話システムは人間と会話を行うためのシステムであり，人と会話を行ったときに相手から人との区別がつかないような会話を行うことが究極的な開発目標であると言える．対話システムが人間と区別がつかないようにするために必要とされる要素は，人間が持つ要素を分析し，それらを陽にモデル化するのが一つの方向性である．ここで，人同士で会話を行う場合に会話

連絡先: 李 為達，聖光学院中学校高等学校，神奈川県横浜市中区滝之上 100，liweida9@gmail.com

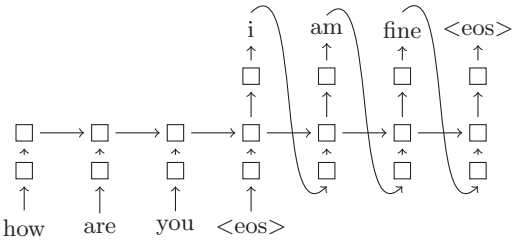

図 1: Seq2Seq を使用した対話システム [Csaky17]

の進行を左右する要素として以下の 4 つの特徴を考える：

- 問われている返答の種類の理解

  会話中に話し手が聞き手へ伝達している文章が聞き手に何を求めているかを理解することで，聞き手はどのような返答を行えば良いのかが理解できる．例えば，質問を受けたときにはそれに対する答えを返すのが一般的であるが，質問に答えずに挨拶を行うと会話としては不適切となるため，聞き手に対する文章の，要求の分類と検出が必要となる．

- 会話文中の感情の認識

  話し手から同じ文を与えられても，聞き手は受けた時の感情により返答が変化することがしばしばある．従って，相手と自分の感情は出力に影響を与える重要な要素であると言うことができる．

- 会話中のキーワードに関する前提知識

  人は会話内で出現するキーワードについて連想をし，例えば話題を変えようとする時などに，それが相手への返事に影響を与えることがある．対話システム内でもこのプロセスを含めることで，会話を円滑に進める可能性を残すことができる．

- 過去の会話文の内容の把握

  人同士の会話は一文一文を単体でかけ合うというよりも複数のやりとりから成り立つものであり，会話中の

ある時点に至るまでのタイムステップより以前の会話の吟味が可能でなければ，人間のような会話は実現が難しい．会話の内容を始終まで維持するためには，この要素が必要不可欠である．

これらの要素は単に考慮するだけでなく，総合的に考慮して出力を考える必要がある．また，会話の経験に基づいて得た要素に適した返答を生成する必要がある．つまり，要素をすべて踏まえた上で，今までその人が経験してきた会話文の構造やパターンに基づいて返答を生成する．これらの特徴を再現するために，対話システムには以下の機能が必要とされると考えられる．

- 会話文分類

  人間が問われている返答を認識できるという要素を模擬するために，入力文から問われていることを計算する機能が必要である．

- 会話文中の感情検出

  人間が自分の感情を認知すると共に，発現に含まれる感情を認識しながら会話を行っているように，対話システムでも同様の機能を果たすモデルが必要である．

- 会話に必要な前提知識

  人は他人と会話を行う時に何も知識がない状態から始まるのではなく，世間一般において常識と定義づけられる，例えば，果物についての会話の時,「果物は食べ物である」や,「果物は植物である」といった，ある程度の前提知識を持っている．よって，これをシミュレートする機能も対話システムには必要となると考えられる．

- 会話文貯蓄

  人間が過去の会話の内容を覚えているのと同様に，会話中のあるタイムステップに至るまでの会話の趣旨を大まかに保存する必要がある．

- 会話文のベクトル表現

  機械学習モデルが自然言語の処理を可能にするためには，文章をベクトルとして扱う必要がある．

本稿では，文章の潜在的な要素を認識するモジュールとして，上記の 5 つの機能に着目し，システムを構築することとした．

3. 対話システムの構築

図 2 に，本稿で構築した対話システムのダイアグラムを示す．タイムステップごとに入力文から返答の出力文を生成するプロセスを繰り返し，同じ会話内である限りタイムステップごとの会話文を潜在空間で抽象化してできたベクトル列を保存し，次のタイムステップで利用する．

話し手から受け取った文章を単語に分割し，GloVe[Pennington14] を用いて各単語をベクトルに変換する．それらを集めたベクトル列を会話文のベクトル表現として扱い，それぞれのモジュールでこのベクトル列を入力して処理を行う．処理後に会話文から潜在要素を抽出した各モジュールの出力値を集め，それらを会話文を抽象化した数値として自己組織化モジュールに入れ，出力の計算を行う．以下，各モジュールについて説明する．

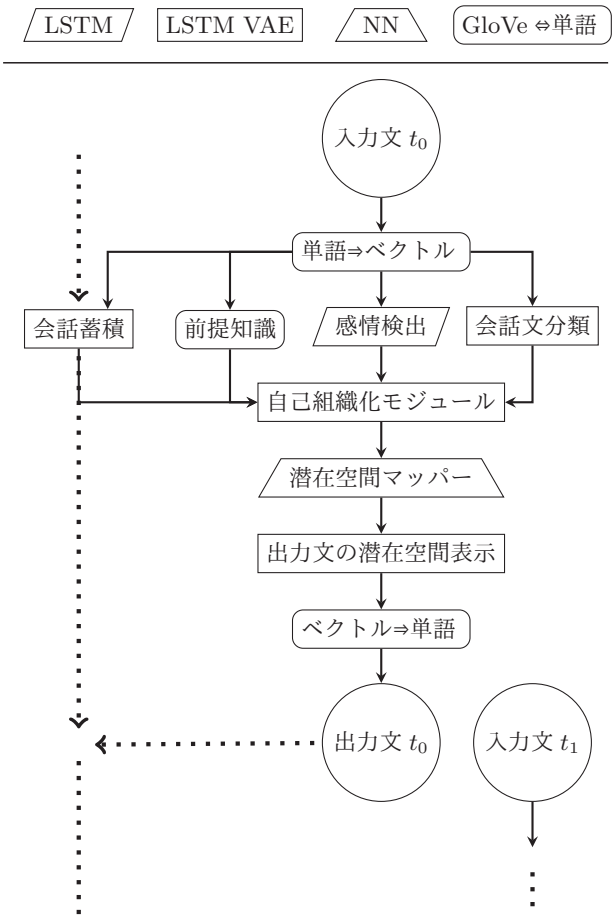

図 2: 提案するシステム

3.1 単語ベクトル変換モジュール

文章のままでは処理が行えないため，GloVe でワードエンベディングを行い，ベクトル化した．本稿では GloVe のモデルとして，Stanford NLP Group のウェブサイトから入手可能な “Wikipedia 2014 + Gigaword 5” の 200 次元及び 300 次元のバージョンを使用した．入力で受けた文章は単語と記号をスペースで分割し，それぞれに対応したワードエンベディングを文中の出現順にクトル化したベクトル列を構築する．構築したベクトル列は，感情処理モジュール，文種類処理モジュール，会話蓄積モジュールの 3 つのモジュールに引き渡される．

3.2 会話文分類モジュール

ベクトル列として表現された文章の会話内における役割を，可変長である文章を入力として受け，教師なしで潜在空間における分類が可能な LSTM VAE を用いて算出する．

3.3 感情検出モジュール

ベクトル列で表現された文章から算出できる相手の感情と自分の感情を計算し，8 種類の感情の中から最も近い感情を話し手と聞き手それぞれについて出力する．モデルとしては，可変長である文章を入力として受けることができる LSTM を使用する．

3.4 前提知識モジュール

元の英文の中でキーワードとなる単語を，Rapid Automatic Keyword Extraction algorithm [Rose10] で抽出し，それらと

関連するとされる単語の GloVe におけるベクトル表現を，最大 3 つまで出力する．

3.5 会話蓄積モジュール

話し手と聞き手の一回のやりとりをペアとして，次のタイムステップで使用するために保存する．これは人が会話中にキーワードと呼ぶことができる単語を聞いた時に，無意識に連想をしてしまい会話に影響を与えるような傾向があることに基づいたモジュールである．「会話の流れ」を会話蓄積モジュールの潜在変数のベクトルを通して数値化することで，会話の進み方を可視化するなどといった応用も可能となる．

3.6 自己組織化モジュール

上記の 4 つのモジュールの出力を総合し，一つ一つのベクトルが GloVe の辞書内である一つの単語に対応するようにベクトル列の形で表現された文章を出力する．モデルには，LSTM VAE を使用した．

3.7 出力文潜在空間モジュール

このモジュールは，2 章の中で定義される「会話の経験に基づいて，得た要素に適した返答生成」である．LSTM VAE を用いて，あらかじめ返答文の文の構造を教師なしで学習を行う．これによって，自己組織化モジュールの出力で潜在空間が定まれば出力文の趣旨も定まることになる．

3.8 潜在空間マッパー

自己組織化モジュールで会話文中の要素をまとめたものの潜在空間においての表現と出力文潜在空間モジュールの潜在空間的表現の対応を学習し，新規の文の入力を受けたときに抽出された要素に基づいて出力文の潜在空間的表示の生成を行う．

3.9 ベクトル単語変換モジュール

単語ベクトル変換モジュールと同じように，ワードエンベディングのベクトルをそれらに対応した単語に変換し，組み合わせて生成文として出力を行う．

各モジュールの学習誤差，Optimizer，ロス関数及び入力・出力の次元数を表 1 に示す．

4. 使用したモデルの詳細

4.1 LSTM

感情検出モデルに使用した LSTM は基本的に中間層を持たず，入力は GloVe によってワードエンベディングでベクトル化した 200 次元の単語ベクトルを可変数回受けることができる．感情検出モデルは，感情に対応する 1 つの数値を出力する．

4.2 LSTM VAE

LSTM VAE は，一般的に使われる VAE のエンコーダーとデコーダーを LSTM で置き換えたものである．可変サイズの入力を受けることができるエンコーダー側から出力されるベクトル列は，入力の長さに依存せず一定の大きさの行列を出力し，これを一般的な VAE と同様にデコーダーに入力し元の可変長の入力を復元する．LSTM VAE の誤差は，元の潜在要素の値の復元を行なった時の元の値と復元した値の誤差を意味する．本稿でこれを使用した理由は，殆どの場合で可変長である会話文を汎用的に潜在空間に変換するモデルであるためであり，会話文分類モジュールなどに使用することができる．

5. 検証

5.1 検証方法

モデルの会話の質を把握するための手段として，モデルの本来の目的である人との会話を実行した． 今回は人が話し手，モデルが聞き手として人から話をかけられた時に返答を行うという設定のもとで検証を行なった．会話の終了の判断は，話し手の人が不自然な会話となりつつあると感じる，または会話が終わったことを人が自然に感じ取ることができたときとする．

5.2 検証結果・考察

検証の 2 つの例を表 2 と 3 に示す．

一つ目の結果について，会話のやりとりを一行ずつ，2 章で触れた会話中の要素を考えながら見ていきたい．

1 行目では人間は感情なしの挨拶を行ったといえる．この時，一般的な人間は殆どの場合は同じく挨拶で返すと予想される．人が予想する通りに，対話システムは人間の挨拶への返答をし，同じ主旨のことを聞く文を返している．2 行目と 3 行目では，すべて人間が返しうる返答といえるので，それぞれ適切な出力を行ったといえる．2，3，4 行目の人間が与えた文章はそれぞれコメント，質問と賛同と概ね分類することができる．4 行目において，会話が終わってしまうような返答を人間がしている．しかし，対話システムは話題を変えて会話を続けようとする意向が見える返答を行っている．ここでは，コメントとともに質問を返しており，人間が会話を続けようとするときと非常に似た形式の会話を行っているとも言える．5 行目では，人間は感情を含まない挨拶を行ってから，それに対するコメントを返しておりこれは適切な返答であると考えられる．

次に，例 2 について見ると，概ね例 1 と似た形式を取っていることがわかる．しかし，5 行目において “Know you not reading?” という文法的に誤った出力を返している．この原因として考えられるのは，学習不足である．この実験から，今回得た対話システムはある程度の会話を実践することが可能であり，かつ会話を継続しようとするテクニックも所持しているように見受けられる．

6. まとめ

本稿では，人が会話中に感じると推測される潜在的要素を陽にパラメータとした対話システムのモデルを提案した．モデルの検証については，人が実際にインタラクションを複数回行った時の対話システムの返答の質を特定の基準でなく，人の感覚によって判断しているため，安定して使用できる正確な検証方法とは言い難い．よって，今後は Conversation-turns Per Session(CPS) [Zhou18] という対話システムの会話の続行性を測る尺度で検証を行いたいと考えている．

今回の検証から，潜在要素を意識した対話システムはある程度人間に近いような会話を行うことができることが分かった．しかし下記の様にい，くつか疑問点が残るところがある．

- 潜在要素の影響はどれくらいあったか?

  今回の対話システムでは，潜在要素をまとめる自己組織化モジュールへの入力は受け取った入力文自体は使用しておらず，抽出した潜在感情やキーワードなどを使用している．しかし，使用した LSTM VAE 自体，与えられたデータへの学習の正答率は極めて高いモデルである．よって，今回の対話システムでも潜在要素自体が大きな意味を持ったのではなく，ただ単に使用したモデルが与えられたデータに適応してしまった可能性がある．今後，この点を検証する必要がある．

- 潜在要素はこれらで十分か?

  今回の対話システムはある程度成功したと言うことができる．しかし，今回注目した潜在要素は人が意識を

表 1: 各モジュールの詳細

Module name(Model type)	Loss	Loss function	Optimizer	Dimentions
会話文分類 (LSTM VAE)	1.26x10−2	KL Loss	RMSProp	300 → 200
感情検出 (LSTM)	1.03x10−2	sparce_categorical_crossentropy	AdamOptimizer	200 → 1
自己組織化 (LSTM VAE)	8.7x10−3	KL Loss	RMSProp	200 → 60
出力文潜在空間 (LSTM VAE)	4.6x10−3	KL Loss	RMSProp	300 → 200
潜在空間マッパー (NN)	1.5918x10−9	mean_squared_error	Adam	60 → 200

表 2: 人間と対話システムの会話検証: パターン 1

Timestep	Human	Dialogue System
1	Hi, how are you today?	Doing great, you?
2	I'm doing great.	That's nice. What do you do for work?
3	I code. How about you?	Just office work, but it is a lot.
4	Oh, ok. Do you enjoy your work?	Yeah, it's quite rewarding.
5	Yeah, it does pay the bills.	When I'm not working, I play basketball. What do you do for fun?
6	I hike at the park.	That sounds relaxing!
7	It keeps my mind off things. Do you travel?	Not much.

表 3: 人間と対話システムの会話検証: パターン 2

Timestep	Human	Dialogue System
1	Hi! How are you?	Hi good and you?
2	Great! Just ran outside.	What do you do for a living?
3	I am a student. How about you?	I play basketball.
4	Nice! Do you love books?	I read all the time. How about you?
5	I always liked movies more than books.	Know you not like reading?
6	I do, but prefer the movies.	Ok, that makes sense.

しうる要素の中の一部分であるということは容易に分かる．よって，これらの要素のみである程度の会話が可能であるということは，人間も簡単な対話ではこの程度の要素しか考慮していないという可能性も考えられる．この真偽は定かでないため，今後検証の余地がある．提案モデルは，対話文の要素を陽にモデル化しているため，対話データの要素を解析することもできる．

- 対話システムの返答内容の統一性

  一般的にあるデータセットを使って対話システムの学習を行うと，対話システムの返答が統一性に欠けることが頻繁にある．これは対話システムにおける難点の一つであるが，今後対話システム特有の「プロフィール」を作り，プロフィールの内容が必要とされるパターンの検出とプロフィールに基づく返答文の生成を行う．

今後はこれらを改善する方向に研究を進めたい．また，現時点のモデルへ会話に影響を与えうるモジュールを加えるとともに，モデル自身のパーソナリティの固定など，より人間が持つ特徴をモジュールの形で付け加え，さらなる検証を行うことが今後の課題である．

参考文献

[Hochreiter97] S.Hochreiter, and J.Schmidhuber, "Long Short-term Memory," Neural Computation 9(8):1735-80 (1997)

[Sutskever14] I.Sutskever, O.Vinyals, and Q.V.Le, "Sequence to Sequence Learning with Neural Networks," NIPS 2014 (2014)

[Csaky17] R.Csaky, "Deep Learning Based Chatbot Models," Technical Report (2017)

[Serban17] I.Serban, C.Sankar, M.Germain, S.Zhang, Z.Lin, S.Subramanian, T.Kim, M.Pieper, A.Chandar, N.Ke, S.Mudumba, A.Brebisson, J.Sotelo, D.Suhubdy, V.Michalski, A.Nguyen, J.Pineau, and Y.Bengio, "A Deep Reinforcement Learning Chatbot," CoRR2017 (2017)

[Pennington14] J.Pennington, R.Socher, C.D.Manning, "Glove: Global Vectors for Word Representation," <https://nlp.stanford.edu/projects/glove/>, (2014)

[Rose10] S.Rose, D.Engel, N.Cramer, and W.Cowley, "Automatic Keyword Extraction from Individual Documents," Text Mining: Applications and Theory (2010)

[Zhou18] L.Zhou, J.Gao, D.Li, and H.Shum, "The Design and Implementation of XiaoIce, an Empathetic Social Chatbot," arXiv:1812.08989 (2018)