

雑談対話における話題継続願望判定の検討

Estimation of users' desire to continue conversation on current topics in chat-oriented dialogue systems

別所克人¹ 東中竜一郎¹ 大塚淳史¹ 牧野俊朗¹ 松尾義博¹

Katsuji Bessho¹, Ryuichiro Higashinaka¹, Atsushi Otsuka¹, Toshiro Makino¹, and Yoshihiro Matsuo¹

¹NTT メディアインテリジェンス研究所

¹NTT Media Intelligence Laboratories

Abstract: In a chat-oriented dialogue system, if the topic of the system's utterance does not attract the interest of the user, it is desirable that the system switches the topic. In this paper, for the system to estimate the appropriate timing of switching the topic, we propose a method for estimating users' desire to continue conversation on current topics. Our proposed method improved the estimation accuracy by 13.7 point from the baseline that constantly estimates that the user has no desire to continue conversation on the current topic.

1 はじめに

非タスク指向型の対話システムは、特定の目的を遂行するタスク指向型システムと異なり、タスクを限定しない雑談対話を行うものである。コミュニケーション活性化のため、コンピュータが語り相手の役割を担ってくれることを期待するニーズは今後高まっていくものと考えられ、このような期待に応えるべく、NTT では雑談対話システムの構築を行っている[1]。

現状の NTT の雑談対話システムでは、ユーザとの対話のやり取りにおいて、現在の話題を認識し、その話題に沿った発話をシステムが行う。だが、システムからの発話は、必ずしもユーザの興味を引くとは限らない。システム発話を受けたユーザ発話がなされた後、直前のシステム発話の話題に対するユーザの反応を見ると、その反応は、2)その話題を継続したい、1)その話題を継続してもさしつかえない、0)その話題を継続したくない、というように分類できる。2)や 1)のケースでは、直近のシステム発話の話題を、システムはそのまま継続するか、あるいは、適宜、ユーザを飽きさせないように話題を変更することになる。0)のケースでは、直近のシステム発話の話題から必ず変更する必要がある。ユーザが話したい話題についての雑談を実現するためには、ユーザが直前のシステム発話の話題を継続したか否かの情報が必要となる。

しかしながら、現状、ユーザが直前のシステム発話の話題を継続したか否かを判定する機構がないため、0)のケースでも、システムが直近のシステム発話の話題を継続してしまうという問題が

ある。常に定期的に（例えば 2, 3 発話ごとに）話題を変更する処置を取ったとしても、今度は、2)のケースでも話題が変わってしまうという問題が出る。

本稿では、システムによる話題の適切な切替タイミングを推定するため、ユーザ発話の後、ユーザが該ユーザ発話の直前のシステム発話の話題を継続したか否かを判定する手法を提案する。提案手法では、対話のやり取りから素性を抽出し、機械学習による分類問題として解く。以下、2 節で、現状の話題遷移の課題を、現在の雑談対話システムの構成を踏まえて、より詳細に説明する。3 節で提案手法、4 節で評価実験、5 節で実験結果の考察、6 節で関連研究、7 節でまとめと今後の課題を述べる。

2 話題遷移の課題

図 1 は、NTT の雑談対話システムの構成である。発話解析部では、ユーザ発話に対し、対話行為推定、焦点抽出、照応解析などの解析を行う。対話行為推定は、発話文の対話行為を推定する。対話行為とはユーザの発話の意図（挨拶、共感・同意、自己開示、欲求など）を表すものであり、本システムでは 33 種類からなるタグセット[2]を用いる。

焦点抽出は、発話文から話題相当の名詞句である焦点を、系列ラベリングの手法によって抽出する。複数の焦点候補が存在する場合には複数保持する。これを焦点リストと呼ぶ。

照応解析で、発話文中のゼロ代名詞を、対話文脈及び当該発話文から抽出された焦点リストから選択する。ゼロ代名詞は話者間で特に共有されているものと考えられるため、ゼロ代名詞、当該発話文から

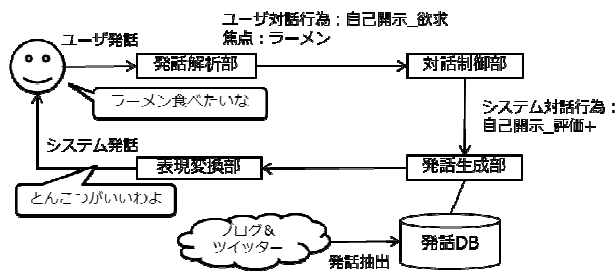


図 1：雑談対話システムの構成

抽出した焦点リスト、対話文脈からの焦点リストの順に優先順位をつけマージしたものを、当該発話文からの最終的な焦点リストとする。システム発話に対しても同様の処理にて焦点リストが決定される。

対話制御部で、過去の N 対話行為から、次にシステムが出力すべき対話行為を出力する。

発話生成部では、焦点と次のシステム発話の対話行為を満たす発話を生成する。あらかじめブログやツイッターといった大規模テキストデータから構築した発話生成知識や、既存の対話データから構築した発話リストから、条件を満たす発話を生成ないし選択する。

表現変換部では、システム発話の文末表現を変換し、システム管理者が望むようなシステムの個性の表出を行う。

以上が雑談対話システムの概要だが、把握した焦点に基づくシステム発話が、必ずしもユーザの興味を引くとは限らない。例えば、以下の対話例では、各発話文から焦点として「ランニング」が抽出され、システムは「ランニング」に関する発話をし続ける。だが、各ユーザ発話から推察するに、ユーザはシステム発話の話題を継続したいとは思っていない。

システム：ランニングはいいですね
 ユーザ：疲れるだけだよ
 システム：ランニングは面白いですね
 ユーザ：つまらないけど
 システム：ランニングは疲れますね
 ……以下ランニングの焦点が続く……

図 2：同じ焦点が続く対話例

しかし、同一焦点が続けば、ユーザが話題継続をしたくなるという訳ではない。後述する継続願望有無のアノテーションデータから、継続願望あり [2] の場合においても、同一焦点が続いているケースが一定の割合で存在することが分かっている。

また、同一焦点の連続出現回数がまだ 1 回のときに継続願望なし [0] となるケースも、全ユーザ発話の 13.2% の割合で存在した。これらのことは、単純に同一焦点の連続出現回数で、継続願望有無の判

断をすることはできないことを示唆している。

3 提案手法

3.1 問題の定式化

提案手法では、ユーザ発話の後、ユーザが該ユーザ発話の直前のシステム発話の話題を継続したか否かを、機械学習による分類問題として解く。学習及び評価データとして、複数の被験者に雑談対話システムと会話してもらい、図 3 のようなシステムとユーザとの間の会話データを 817 会話分、作成した。図 3 は、1 会話データの例であり、最初の行から順に各行において、システム発話・ユーザ発話の順で発話がなされたことを表す。この各行をターンと呼ぶ。2 人の作業者が、各会話データの各ターンに対し、ユーザ発話の後、ユーザが直前のシステム発話の話題を継続する願望を持っているか否かを示すフラグを以下の基準で付与した。

- 2)直前のシステム発話の話題を継続したい
- 1)直前のシステム発話の話題を継続してもさしつかえない
- 0)直前のシステム発話の話題を継続したくない

システム発話	ユーザ発話	フラグ
ラーメンは美味いよね	うん、好きだよ	2
ラーメンはスープが大事だね	本当にそう思う	2
ラーメンを食べたいな	まあね	1
ラーメンは麺が命	もういいよ	0
ラーメンを食べてダイエット することができるよ	本当に？	2

図 3：学習・評価のための会話データ例

2 人の作業者が付与したフラグの一致率は表 1 のようになり、作業者間で、かなりの揺れがあり、特にフラグ 1 で揺れ幅が大きいことが判明した。

表 1：2 人の作業者による付与結果の一致率

	再現率	適合率	F 値
フラグ 2	56.4%	51.1%	53.6%
フラグ 1	21.1%	42.4%	28.2%
フラグ 0	80.4%	60.8%	69.2%

両作業者がともにフラグ 2、あるいは、フラグ 0 と付与したターンは、両作業者が感じたことがフラグ 1 と異なり鮮明で、なおかつ一致している。それ以外のターン(フラグの組合せが (2,1), (2,0), (1,1), (1,0)) に対しては、システムが話題を切り替えても切り替えなくても、少なくとも一方の作業者は許容しうるので、一般に許容されると考えられる。このため、両作業者がともにフラグ 2、あるいは、フラグ 0 と付与したターン群を対象として、フラグ 2

か0かを決定する2値分類問題として解くこととした。

3.2 分類素性

分類対象のターンまでの発話列から、分類に寄与すると推察される素性として、4節記載の各素性単体の精度値を記述した表3中の素性をリストした。以下に各素性について述べる。

a)対象ターンのユーザ発話またはシステム発話の内容

対象ターンのユーザ発話は、フラグ2の場合は、システムと共感している内容が多い傾向があり、フラグ0の場合は、倦怠感やシステムに対する不満表明等の内容が多い傾向がある。これに対応する素性として、ユーザ発話やシステム発話の単語頻度ベクトルや、単語 bigram (または trigram) 頻度ベクトル、SVD に基づく概念ベクトル[3]、word2vec ベクトル[4][5]、意味カテゴリ頻度ベクトル[6]を採用した。

b)発話間の内容の類似度

対象ターンのシステム発話とユーザ発話の内容の類似度は、フラグ2の場合は高く、フラグ0の場合は低い傾向がある。また、過去のターンまでさかのぼると、フラグ0の場合、システム発話とユーザ発話とで互いに異なる話題を平行して継続している事例もある。これに対応する素性として、過去N発話における任意の2発話間の類似度を採用した。類似度としては、a)で挙げた各素性ベクトルの種類ごとに、素性ベクトル間のコサイン類似度を取った。

c)対話行為の列

対象ターンのユーザ発話の対話行為は、フラグ2の場合は共感・同意、自己開示_評価+、質問_事実が多く、フラグ0の場合は自己開示_評価-が多い等、フラグによって特徴がある。このことから、過去N発話の対話行為の列を素性として採用した。

d)同一焦点の連続出現回数

2節で、対象ターンのシステム発話（またはユーザ発話）に関する焦点が、当該発話に至るまでのシステム発話（またはユーザ発話）に何回、連続出現したかだけでは継続願望有無の判断はできないと述べたが、分類にいくばくか寄与する可能性があるので、同一焦点の連続出現回数も素性として採用する。表3のNo.13~No.15、No.17~No.20において、焦点リストの連続出現回数とは、焦点リスト間に共通する焦点があれば、同一焦点リストとみなしてカウントしたものである。No.12の3種類の焦点リストとは、発話ごとの、当該発話から抽出した焦点リスト、ゼロ代名詞、最終的な焦点リストを意味する。No.13のある焦点リストとは、システム発話の基となった焦点リスト、システム発話から抽出した焦点リスト、

ユーザ発話から抽出した焦点リストを意味する。

e)対象ターンのユーザ発話またはシステム発話の文末表現

対象ターンのユーザ発話の文末表現（文末の付属語等を連結した文字列）は、フラグ2の場合は共感表現が多く、フラグ0の場合は否定やシステムに対する不満表明等が多い傾向がある。これに対応する素性として、ユーザ発話やシステム発話の文末表現を採用した。

f)対象ターンのユーザ発話の文字情報

対象ターンのユーザ発話の文字数は、フラグごとの1~3文字の占める割合は、フラグ0の方が高く、フラグごとの5文字以上の占める割合は、フラグ2の方が高い。また、フラグごとのカタカナ文字の占める割合は、フラグ0の方が高い。これらのことから、対象ターンのユーザ発話の文字数や、各文字種の字数（及びその割合）を素性として採用した。

g)対象ターンのユーザ発話からのその他各種情報

対象ターンのユーザ発話は、フラグ2の場合はポジティブ表現の方が多く、フラグ0の場合はネガティブ表現の方が多く傾向がある。また、経験表現の出現数や要望文としての度合いは、フラグ2の方が多く傾向がある。これらのことから、ユーザ発話のポジティブ表現数とネガティブ表現数、経験表現の有無、要望文度を素性として採用した。これらの情報の抽出は機械学習に基づく手法[7]を使用した。

4 評価実験

3.1節で述べた817会話（ターン数/会話:29, 分類対象ターン数:11099）に対し、会話集合の8分割交差検定による精度検証を行った。1つの会話内で、学習データとテストデータに分割することはせず、会話単位で、学習データとテストデータに分け、学習データとテストデータとが完全に独立しているようにした。学習データとテストデータに対する各種値の平均値は表2のようになった。

表2：学習データとテストデータの各種値

	会話数	対象 ターン数	フラグ2の ターン数	フラグ0の ターン数
学習データ	714.875	9711.625	2448.25	7263.375
テストデータ	102.125	1387.375	349.75	1037.625

以降で示す精度値は、精度値の種類（正解率等）ごとに、マクロ平均をとったものを表す。

ベースライン手法として、全対象ターンをフラグ2とみなす手法、フラグ0とみなす手法、及び、システム発話の基となった焦点リストの連続出現回数がN回に達したらフラグ0とみなし、それ以外はフ

ラグ2 とみなす手法 (表3の No.3)、No.3 で対象ターンのユーザ発話の対話行為が質問_*ならば無条件

にラグ2 とみなす手法を取った。

表3：各素性単体の精度

No	手法 or 素性	正解率	フラグ2のF値	フラグ0のF値	※1
1	ベースライン (全部2)	0.251926	0.401828	0	—
2	ベースライン (全部0)	0.748074	0	0.855651	—
3	ベースライン (Sの基となった焦点連続回数) 1回	0.765988	0.250485	0.860994	—
4	ベースライン (Sの基となった焦点連続回数+質問) 1回	0.738138	0.332555	0.836706	—
5	単語頻度ベクトル (U)	0.692201	0.55271	0.76326	○
6	単語頻度ベクトル (S)	0.7215	0.312883	0.824898	○
7	単語頻度ベクトルの集まり中の任意のペア間の類似度 (過去14発話)	0.739493	0.308246	0.839171	○
8	単語頻度ベクトル (U) (内容語のみ)	0.55212	0.471793	0.593217	○
9	単語頻度ベクトル (S) (内容語のみ)	0.574064	0.325359	0.683585	○
10	単語頻度ベクトル (内容語のみ)の集まり中の任意のペア間の類似度 (過去21発話)	0.730803	0.38652	0.826753	○
11	対話行為 (過去9発話)	0.78792	0.475452	0.866793	○
12	3種類の焦点リストの集まり中の任意のペア間の交わり (過去4発話)	0.768957	0.339031	0.859498	○
13	システム発話とユーザ発話のある焦点リストの連続出現回数	0.753529	0.114693	0.856603	○
14	Sの基となった焦点リストの連続出現回数	0.764103	0.219017	0.860464	○
15	Sの基となった焦点リストの連続出現回数 (焦点なしを1回とカウント)	0.748074	0	0.855651	○
16	Sの基となった先頭焦点の連続出現回数	0.764103	0.219017	0.860464	○
17	Sの基となった焦点リストの連続出現回数 (Uの後のものからカウント)	0.755381	0.123407	0.857445	○
18	Sから抽出した焦点リストの連続出現回数	0.511256	0.248577	0.620477	○
19	Uから抽出した焦点リストの連続出現回数	0.519798	0.381766	0.567564	○
20	Uから抽出した空の焦点リストの連続出現回数	0.546429	0.159758	0.680956	○
21	単語表記 bigram 頻度ベクトル(U)	0.584199	0.483221	0.639665	○
22	単語表記 bigram 頻度ベクトル(S)	0.678879	0.332103	0.787143	○
23	単語表記 bigram 頻度ベクトルの集まり中の任意のペア間の類似度 (過去22発話)	0.728565	0.29864	0.831071	○
24	単語表記 trigram 頻度ベクトル (U)	0.52669	0.45646	0.548494	○
25	単語表記 trigram 頻度ベクトル(S)	0.614931	0.357316	0.718155	○
26	単語表記 trigram 頻度ベクトルの集まり中の任意のペア間の類似度 (過去3発話)	0.403375	0.363204	0.344829	○
27	文末表現 (U)	0.528578	0.437572	0.568043	○
28	文末表現 (S)	0.648056	0.2664	0.7648	○
29	カテゴリ頻度ベクトル (U)	0.679164	0.497617	0.762162	○
30	カテゴリ頻度ベクトル (S)	0.711609	0.346159	0.814669	○
31	カテゴリ頻度ベクトルの集まり中の任意のペア間の類似度 (過去19発話)	0.749392	0.364807	0.843497	○
32	概念ベクトル (U)	0.811962	0.60718	0.87617	○
33	概念ベクトル (S)	0.74842	0.310974	0.845691	○
34	概念ベクトルの集まり中の任意のペア間の類似度 (過去10発話)	0.747478	0.339554	0.843556	○
35	概念ベクトル (U) (内容語のみ)	0.776752	0.464943	0.858617	○
36	概念ベクトル (S) (内容語のみ)	0.738234	0.293117	0.838965	○
37	概念ベクトル (内容語のみ)の集まり中の任意のペア間の類似度 (過去4発話)	0.604461	0.353342	0.714175	○
38	word2vec (U)	0.752349	0.426454	0.841783	○
39	word2vec (S)	0.737246	0.057552	0.847144	○
40	word2vec ベクトルの集まり中の任意のペア間の内積 (過去10発話)	0.756313	0.308262	0.851925	○
41	word2vec (U) (長さ1に正規化)	0.748165	0.409014	0.83965	○
42	word2vec (S) (長さ1に正規化)	0.736704	0.044464	0.8471	○
43	word2vec ベクトル (長さ1に正規化)の集まり中の任意のペア間の内積 (過去10発話)	0.758719	0.20414	0.857597	○
44	文字数 (U)	0.748074	0	0.855651	○
45	各文字種の字数 (U)	0.748074	0	0.855651	○
46	各文字種の字数割合 (U)	0.748074	0	0.855651	○
47	positive 表現数と negative 表現数 (U)	0.373398	0.392254	0.277479	○
48	経験表現の有無 (U)	0.359643	0.332832	0.282425	○
49	要望文度 (U)	0.318153	0.354242	0.10835	○

表4：複数素性での精度

No	手法	正解率	フラグ2の評価値			フラグ0の評価値		
			再現率	適合率	F値	再現率	適合率	F値
1	ベースライン (全部2)	0.251926	1	0.251926	0.401828	0	0	0
2	ベースライン (全部0)	0.748074	0	0	0	1	0.748074	0.855651
3	ベースライン (Sの基となった焦点連続回数) 1回	0.765988	0.15589	0.65515	0.250485	0.9714	0.773407	0.860994
4	ベースライン (Sの基となった焦点連続回数+質問) 1回	0.738138	0.259662	0.466085	0.332555	0.899555	0.782373	0.836706
50	素性組合せ (※1)	0.884869	0.724042	0.797757	0.758248	0.938951	0.910115	0.924197

4.1 素性単体の精度

3.2 節で述べた各素性単体で、LIBLINEAR¹による学習とテストを行った。その精度値を表 3 に示す。表 3 の素性欄で、U はユーザ発話に関するもの、S はシステム発話に関するものであることを表す。また、表 3 において、過去発話数 N のパラメータをもつもの (No.3 等) は、検証した N の範囲で、最大の正解率をとったもののみ記載している。

ベースライン手法の中では、No.3 が最高の正解率となった。提案素性の中では、No.32 のユーザ発話の概念ベクトルが最高の正解率となり、いずれの精度値の種類でも、ベースラインの最高値を上回った。対象ターンのユーザ発話の内容は、フラグ 2 と 0 とで、ある程度特徴があり、学習データに出現する単語をさらに汎化させることにより、高精度が得られたものと考えられる。また、内容語のみを考慮する素性 (No.8 等) より、全単語を考慮する素性 (No.5 等) の方が高精度の傾向があり、付属語等が重要な役割を果たしていることが伺える。

4.2 複数素性での精度

表 3 で挙げた提案素性の中のいくつかを組合せ、各素性のベクトルを結合して得られる素性ベクトルに対し、LIBLINEAR による学習とテストを行った。

まず、表 3 の提案素性全てからなる組合せに対し精度値を算出した。その後、1 つの素性を除いた組合せの正解率が、直前の正解率より高精度となれば、該素性を除くという操作を、全ての素性に対し続けた。最終的に得られた素性の組合せは、表 3 の (※1) に○をつけたものとなった。この正解率最大となる素性組合せ (※1) の精度値は、表 4 のようになった。表 4 には、ベースライン手法の精度値も再掲している。素性組合せ (※1) は、No.2 の常に継続願望無しとみなすベースライン手法と比べ、正解率が 13.7 ポイント上回り、No.3 のベースライン手法と比べても、正解率が 11.9 ポイント上回り、符号検定により有意水準 1% で有意性を確認した。

素性組合せ (※1) から、一素性を除いた組合せを、その正解率の昇順に並べたときの上位 10 個の組合せは、表 5 のようになった。除いた各素性は、対応する正解率が低い程、分類への寄与度が高い。

過去 9 発話の対話行為の列が最も寄与度が高く、他に、対象ターンのシステム発話やユーザ発話の内容、過去の 2 発話間の類似度、ポジティブ表現数とネガティブ表現数の寄与度が高い。これらの素性が、特に、話題継続願望の有無により異なる特徴を持つ

表 5: 素性組合せ (※1) から一素性を除いた組合せの正解率 (昇順で上位 10 件)

No	素性組合せ	正解率	(※1) の正解率との差分
50	素性組合せ (※1)	0.884869352	0
11	(※1) - 対話行為 (過去 9 発話)	0.879114187	0.005755165
32	(※1) - 概念ベクトル (U)	0.88015998	0.004709372
28	(※1) - 単語表記 trigram 頻度ベクトルの集まり中の任意のペア間の類似度 (過去 3 発話)	0.880826925	0.004042427
6	(※1) - 単語頻度ベクトル (S)	0.881176215	0.003693137
9	(※1) - 単語頻度ベクトル (S) (内容語のみ)	0.881198677	0.003670675
37	(※1) - 概念ベクトル (内容語のみ) の集まり中の任意のペア間の類似度 (過去 4 発話)	0.881207921	0.003661431
47	(※1) - positive 表現数と negative 表現数 (U)	0.881211422	0.00365793
5	(※1) - 単語頻度ベクトル (U)	0.881343942	0.00352541
22	(※1) - 単語表記 bigram 頻度ベクトル (S)	0.881426307	0.003443045
36	(※1) - 概念ベクトル (S) (内容語のみ)	0.881543159	0.003326193

と言える。

5 考察

フラグ 2 のターンを、フラグ 0 と判定したケースの中には、対象ターンのシステム発話とユーザ発話との類似度が低く、フラグ 0 と判定されたものが散見された (例: S: 「そうですね」→U: 「寒いのは好きですか?」)。また、学習データのフラグ 0 のターンのユーザ発話において、高い割合を占める対話行為が、対象ターンのユーザ発話で推定されることによって、フラグ 0 と判定されたケースも散見される (例: S: 「担当者の名前が未だに分からない」→U: 「グズグズして取り置き解除になったら大変ですよ!」 (対話行為: 自己開示_評価一))。

フラグ 0 のターンを、フラグ 2 と判定したケースの中には、以下のようなものが散見された。

・システムからの質問をユーザが拒絶している

例) S: 「どんな音楽を聴かれるのですか」→U: 「おしえない」

・システム発話の内容が理解不可能であるとユーザが言っている

例) S: 「意味は恋ができてます??」→U: 「意味が分かりません」

・システムが理解できていないとユーザが言っている

例) S: 「結婚記念日は年数が経っているようです」→U: 「うんと、理解できてないみたいだけ」

・話題の変更をユーザが提案した

例) S: 「趣味の話以外しないのが悪いのはわかっている」→U: 「じゃあ、違う話をしよう」

・ユーザが同じことも何度も言われている

例) S: 「ペット飼ってますかー!」→U: 「飼っていないって!!!!!!」

・ユーザが呆れてそっけない返事を続けている

例) S: 「美味しいもの食べて楽しい音楽聴いて元気

¹ <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>

だそうねえ」→U:「はいはい」

これらの誤り事例を解決するには、学習データを増やす、フラグ 0 に相当する典型的なフレーズを登録する、ユーザが過去に発話した内容を記憶して活用する、等の方策が考えられる。

6 関連研究

本研究のように、対話に対する話者の感性を対象とした研究として、長谷川らの研究[8]や徳久らの研究[9]がある。[8]では、相手の発話を受けた後の聞き手の感情を予測する。感情としては、Plutchik[10]が定めた基本感情である 8 カテゴリを採用しており、本稿で扱う現在の話題に対する継続願望とは異なる。[9]では、発話行為・修辭構造と対話の盛り上がりとの関連を分析している。対話の盛り上がりと話題継続願望有無とは相関する可能性があるが、本稿の手法では、発話列から抽出した素性に基づき、話題継続願望有無の判定まで行っている点が異なる。

また、システムがどのような話題に遷移すべきかについては、中野ら[11]が、発話中の単語を単語分散表現を用いて関連語に展開する研究を行っている。

7 まとめと今後の課題

本稿では、ユーザ発話の後、ユーザが直前のシステム発話の話題の継続願望を持っているか否かを、これまでの発話列から抽出した素性ベクトルに対する 2 値分類問題として定式化し、88.5%の正解率を達成した。今後は、5 節で述べた誤り分析を基に、さらなる精度向上を図っていく。

また、話題継続願望の判定結果をもとに、次の適切な話題を求める課題を検討していく。

継続願望有りの場合、ユーザ発話から焦点が抽出されたとしても、基本的にユーザ自身は直前のシステム発話の話題に対する継続願望があるため、直前のシステム発話からの最終的な焦点と、現ユーザ発話からの最終的な焦点の双方を考慮の上、次発話のための最適な焦点を導出するのが良いと考えられる。それに加え、これまでの文脈における話題の遷移を踏まえ、関連する焦点への遷移が必要となることも考慮する必要がある。

継続願望無しの場合、本稿で対象とした 817 会話データにおいては、フラグ 0 のターンの内、ユーザ発話から焦点が抽出されたターン (A とする) は、37.8%を占めた。ターン A の次のターン B のシステム発話は、抽出した焦点に基づくものである。ターン B の内、フラグが両作業で 2 となったものが 2.7%、両作業で 0 となったものが 67.2%、両作業

者で一致しなかったものが 30.1%という割合となり、ユーザ発話から焦点が抽出されても継続願望がない場合、同じ焦点を用いてシステム発話を行っても、過半数は、やはり継続願望無しのままとなることが分かった。これは、継続願望無しの場合のユーザ発話から抽出された焦点は、必ずしもユーザがしたいと思っている話題に相当するとは限らず、安易にユーザ発話から抽出された焦点を用いてはならないことを意味している。継続願望無しの場合、ユーザ発話から焦点が抽出されても、それが以降の話題に適切なものかを吟味した上で、直前のシステム発話の焦点とは異なる次の焦点を導出する必要がある。

参考文献

- [1] 東中竜一郎: 雑談対話システムに向けた取り組み, 人工知能学会研究会資料 (SIG-SLUD-70), pp. 65-70, 2014.
- [2] Toyomi Meguro, Yasuhiro Minami, Ryuichiro Higashinaka, and Kohji Dohsaka: Learning to control listening-oriented dialogue using partially observable Markov decision processes, ACM TSLP, Vol. 10, No. 4, p. 15, 2013.
- [3] 別所克人, 内山俊郎, 内山匡, 片岡良治, 奥雅博: 単語・意味属性間共起に基づくコーパス概念ベースの生成方式, 情報処理学会論文誌, Vol. 49, No. 12, pp. 3997-4006, 2008.
- [4] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean: Efficient Estimation of Word Representations in Vector Space, Proc. of Workshop at ICLR, 2013.
- [5] Richard Socher, Eric H. Huang, Jeffrey Pennington, Andrew Y. Ng, and Christopher D. Manning: Dynamic Pooling and Unfolding Recursive Autoencoders for Paraphrase Detection, Proc. NIPS, pp. 801-809, 2011.
- [6] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦: 日本語語彙大系, 岩波書店, 1997.
- [7] 菊井玄一郎, 松尾義博: テキストからの知識抽出による新しい Web 情報アクセスに向けて, NTT 技術ジャーナル, Vol. 20, No. 6, pp. 8-11, 2008.
- [8] 長谷川貴之, 鍛冶伸裕, 吉永直樹, 豊田正史: オンライン上の対話における聞き手の感情の予測と喚起, 人工知能学会論文誌, Vol. 29, No. 1, pp. 90-99, 2014.
- [9] 徳久良子, 寺寫立太: 雑談における発話のやりとりと盛り上げの関連, 人工知能学会論文誌, Vol. 21, No. 2, pp. 133-142, 2006.
- [10] Robert Plutchik: A General Psychoevolutionary Theory of Emotion, Emotion: Theory, research, and experience: Vol.1, Theories of emotion, pp. 3-33, New York: Academic, 1980.
- [11] 中野哲寛, 荒木雅弘: 雑談対話システムにおける単語分散表現を用いた話題展開手法, 言語処理学会第 21 回年次大会, pp. 269-272, 2015.