2C4-4

非タスク指向型対話における盛り上がり箇所の抽出

Automatic Extraction of "Enthusiasm" in Non-task-oriented Dialogues

稲葉 通将*1　鳥海 不二夫*1　石井 健一郎*1
Michimasa INABA　Fujio TORIUMI　Kenichiro ISHII

*1名古屋大学大学院情報科学研究科
Graduate School of Information Science, Nagoya University

Recently, computerized dialogue agents are studied actively. However, most of the studies of the dialogue agents are related to the task-oriented dialogue agents. That is, there are few studies of non-task-oriented dialogue agents. The analysis of dialogues between humans is necessary to make a practical system that can present enthusiastic dialogues. In this paper, we propose an automatic method that evaluates "enthusiastic" utterances in a dialogue using Conditional Random Fields(CRF). As a result, our method indicates as much performance as human determinations.

1. まえがき

経済産業省は 2009 年 3 月，介護・福祉の現場で役立つロボットの実用化に向けた支援策を講じるという方針を打ち出した．また，ASIMO に代表されるようなヒト型ロボットの開発も活発に進められている．今後はこのような流れが加速し、ロボットの社会進出は急速に進むと考えられる．ロボットが人間社会に溶け込むためには，人間と円滑なコミュニケーションが行えることが必要不可欠である．人間の命令を正確に理解し，それを実行するということはもちろん重要である．しかし，真に人間とロボットが共存する社会の実現のためには，挨拶や日常会話も人間と同じように行い，人間と良好な関係を築くことが可能なロボットである必要がある．

しかしながら，人間と対話を行うコンピュータ（対話エージェント）の研究は道案内など，特定のタスク達成を目的としたタスク指向型対話に関するものがほとんどである．一方，人間同士の対話において大部分を占める日常会話・雑談のような非タスク指向型対話に関する研究は少なく，活発に研究が行われているとは言い難い．

タスク指向型対話エージェントは，特定のタスクをいかに迅速かつ確実に達成するかが求められる．一方で，特定のタスクが存在しない非タスク指向型エージェントは，話を盛り上げ，相手を楽しませることにより，人間が対話を続けたいと思えるような対話を展開することが求められる．しかし，どのようにすれば盛り上がる対話が展開できるのかは自明ではない．それを明らかにするためには，人間同士の対話を分析し，対話の盛り上がりに関する何らかの知見を得ることが必要である．その知見は対話エージェントの設計に役立つだけでなく，人間同士のコミュニケーションの支援にも有効であると思われる．

しかし，対話の盛り上がりを人手で判定するのは極めてコストが高い．また，盛り上がりの判定基準は個人差が大きいと思われるため，首尾一貫した判定を行うためにも，自動判定法が必要である．

そこで本研究では対話中における，盛り上がりの度合い (盛り上がり度) の高い発話を Conditional Random Fields(CRF) を用いて自動判定する手法を提案する．本手法では，出現する語に関する素性及び，大規模コーパスから獲得した共起情報に基づく，発話と発話の間の意味的な関係の強さを用いる．これにより，対話中の各部分発話系列において，どの程度の密なやりとりがなされたのかや，話題ごとのまとまりを把握することが可能となる．

本手法を適用することより，大量の対話データを，盛り上がり度の高い箇所とそれ以外の箇所に弁別することが可能となる．その 2 つを比較分析することにより，人間同士の対話の盛り上がりのメカニズム等に関する定量的な解析が期待される．また，人間同士の対話を学習する対話エージェントに，盛り上がり度の高い対話のみを学習させる，といった手法も適用可能となる．

過去に対話の盛り上がりの自動抽出行った研究としては，Gatica-Perez らによる研究がある．Gatica-Perez らは，会議中の動画と音声を用いて，議論の盛り上がった部分の抽出を行った [1]．ただし，素性として用いられたのは，体の動きや声のトーン・大きさなどの表層的な情報であり，対話の内容には踏み込んでいない．そこで本研究では，対話の内容に踏み込んだ評価法を提案する．

なお，本論文では，問題を言語処理に特化するため，テキスト対話のみを対象とする．

2. 盛り上がり箇所抽出法

2.1 概要

本章では人間同士のテキスト対話を取り上げ，盛り上がり度の高い箇所の自動抽出法を提案する．なお，ここでいう対話とは，一回の発言を発話と定義したときに，会話の始まりから終わりまでの発話の系列である．

本研究では，1 発話ごとに対話の盛り上がり度を判定する．ただし，対話の盛り上がりは話者同士のインタラクションの結果生じるものであることから，発話単体だけでなく，その発話に至るまでの文脈も考慮する．

対話の盛り上がり度は「話者がある発話を行った時点での、対話を継続することに対する積極性」と定義する．

例えば，話者がある事柄について非常に話したいことがある場合や，その対話を楽しんでおり，対話を続けたいと考えている場合は，盛り上がり度の高い状態である．

本手法では，発話ごとの盛り上がり度の判定を，盛り上がり度の強さを示す複数のラベルを用いて，それらを発話系列

連絡先: 名古屋大学大学院情報科学研究科
〒 464-8603 愛知県名古屋市千種区不老町
E-mail: inaba@kishii.ss.is.nagoya-u.ac.jp

表 1: 語彙的結束性が強い 2 発話

発話者	発話
A	サントリーニ島全体が本当にすてきなんですよ。 でもニュージーランドも 大自然でいいとこそうですよね？
B	エーゲ海って響きがすでになんか良い感じですね。 ニュージーランドも自然が多くて良い所ですよ！ 人より羊が多かったり、民家にペンギンがいたり、 ホントに良いところでした。

に付与する系列ラベリング問題として定式化する．そして，その系列ラベリング問題を解く手法として CRF (Conditional Random Fields) を用いる [2]．学習に用いる素性は，各発話中の情報だけでなく，発話間の語彙的結束性を用いる．語彙的結束性は，結束性を持つ語の出現により，テキストの意味的なつながりを明示する表層的な情報である．発話間の語彙的結束性を素性として用いることにより，対話中で，どの程度の密なやりとりがなされたのかや，話題ごとのまとまりを把握することが可能となる．

表 1 に語彙的結束性が強い対話例を示す．この例では，海外旅行の話題で話が進んでおり，2 発話の間に「サントリーニ島 ↔ エーゲ海」や「ニュージーランド ↔ 羊」など，意味的に関連する語 (結束性を持つ語) のペアが複数確認でき，密なやり取りがなされていることが分かる．また，対話が盛り上がっており，双方の話者が対話を楽しんでいることが読み取れる．

このように，語彙的結束性は対話の盛り上がりと密接に関連していると考えられることから，本研究では語彙的結束性を，盛り上がりを判定するための素性として用いる．

2.2 共起情報の獲得

語彙的結束性の判定には，類義語辞典やシソーラスが用いられることが多いが [3]，それらは収録語彙数が数千〜数万語程度と十分な数であるとはいえず，ドメインが限定されない非タスク指向型対話には適さない．そこで本研究では，大規模コーパスから共起語を獲得する．その語彙数は非常に多く（今回使用したコーパスでは約 256 万語）,「りんご」から「青森」「赤」のような連想語や,「地震」から「火災」「津波」のような原因と結果の関係にある語も獲得できるという利点がある．

本研究では共起情報の獲得のための大規模コーパスとして Web 日本語 N グラム [4] を使用した．Web 日本語 N グラムはグーグル株式会社から提供されている約 200 億文の日本語データから作成された n-gram データであり，長さ 1〜7 の単語 n-gram データが収録されている．

本研究では，Web 日本語 N グラムに収録されている単語 n-gram のうち最短の 1-gram と，最長の 7-gram を使用する．1-gram は単独での語の出現頻度を獲得するために，7-gram は 2 語の共起頻度を獲得するためにそれぞれ使用した．なお，共起頻度を獲得する語の品詞は名詞，動詞，形容詞とした．共起関係の強さを測る指標には対数尤度比 [5] を使用した．対数尤度比は 2 語が従属している度合が強いほど大きい値を取る．実験では，対数尤度比の値が 10000 以上となった 2 語を共起語対として獲得した．

2.3 CRF による盛り上がりの自動判定

2.3.1 Conditional Random Fields

盛り上がり度の自動判定には Conditional Random Fields(CRF) を用いる [2]．CRF は系列ラベリング問題を解くために設計された識別モデルであり，長さ n の入力発話系列 $\mathbf{x} = x_1 x_2 \cdots x_n$ と盛り上がり度を表すラベルの系列 $\mathbf{y} = y_1 y_2 \cdots y_n$ の対応関係を条件付き確率 $P(\mathbf{y}|\mathbf{x})$ で表現する．ここで，x_i は 1 つの発話を表し，y_i はその発話の盛り上がり度を表す．CRF では素性を素性関数 $\mathbf{f}$，それに対応する重みパラメータを $\boldsymbol{\lambda}$ で表し，k 番目の素性に対する素性関数を f_k，重みパラメータを λ_k とする．素性関数は，0，1 の 2 値を返す関数である．このとき，$P(\mathbf{y}|\mathbf{x})$ は次式で表される．

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z_x} \exp\left(\sum_{i=1}^{n} \sum_{k} \lambda_k f_k (\mathbf{x}, \mathbf{y}, i) \right) \tag{1}$$

ただし，Z_x は全系列を考慮したとき，確率の和が 1 になるようにするための正規化項であり，

$$Z_x = \sum_{y} \exp\left(\sum_{i=1}^{n} \sum_{k} \lambda_k f_k (\mathbf{x}, \mathbf{y}, i) \right) \tag{2}$$

となる．パラメータを $\boldsymbol{\lambda}$ は最尤推定で求めることができる [6]．

入力発話系列 $\mathbf{x}$ に対する，最適な盛り上がり度の系列 $\hat{\mathbf{y}}$ は，

$$\hat{\mathbf{y}} = \underset{\mathbf{y}}{\operatorname{argmax}} P(\mathbf{y}|\mathbf{x}) \tag{3}$$

により求めることが出来る．ここで，$\hat{\mathbf{y}}$ の計算には Viterbi アルゴリズムを用いる．

2.3.2 使用した素性関数

本手法で用いた素性は大まかに分けると以下の 4 種類である．

(1) 各発話中に含まれる語彙に関する素性

(2) 発話の長さ

(3) 発話間の語彙的結束性

(4) 1 つ前の発話の盛り上がり度

この中で (4) の素性については特殊であり，それ以外の 1 つの素性と組み合わせても用いる．つまり，これらの素性から，(1),(2),(3),(4),(1)+(4),(2)+(4),(3)+(4) の計 7 種類の素性関数を定義する．

以下では，各素性の詳細と，定義される素性関数について述べる．

(1) 各発話中に含まれる語彙に関する素性

各発話中に含まれる語彙に関する素性は，

- 学習データに 3 回以上出現した 1-gram
- 学習データに 3 回以上出現した 2-gram

の 2 つである．ここから定義する素性関数は，発話中にある語彙が含まれれば 1，含まれなければ 0 を返す関数である．つまり，学習データに 3 回以上出現した 1-gram および 2-gram の種類数だけ素性関数が定義されることになる．例えば,「大学」という 1-gram が学習データ中に 3 回以上出現した場合，定義される素性関数は，発話 x_i の中に「大学」が含まれる場合には 1 を，含まれない場合には 0 を返す関数である．

表 2: 語彙的結束性に関する素性関数

番号	素性関数の種類
1	<(a) x_i>
2	<(b) x_i>
3	<(c) x_i>
4	<(a) x_{i+1}>
5	<(b) x_{i+1}>
6	<(c) x_{i+1}>
7	<(a) x_i>,<(b) x_i>
8	<(a) x_i>,<(c) x_i>
9	<(b) x_i>,<(c) x_i>
10	<(a) x_{i-1}>,<(a) x_i>
11	<(a) x_i>,<(a) x_{i+1}>
12	<(a) x_{i+1}>,<(a) x_{i+2}>
13	<(a) x_i>,<(b) x_i>,<(c) x_i>
14	<(a) x_{i-2}>,<(a) x_{i-1}>,<(a) x_i>
15	<(a) x_{i-1}>,<(a) x_i>,<(a) x_{i+1}>
16	<(a) x_i>,<(a) x_{i+1}>,<(a) x_{i+2}>
17	<(a) x_{i+1}>,<(a) x_{i+2}>,<(a) x_{i+3}>

(2) 発話の長さ

発話の長さでは，学習データ中の全発話と評価発話を形態素数順に並べたとき，評価発話が 10% 刻みで上位何% に含まれるかを素性とする．すなわち，発話の長さに関する素性関数は計 10 個定義される．

(3) 発話間の語彙的結束性

発話間の語彙的結束性には，

(a) ある発話とその 1 つ前の発話との語彙的結束性

(b) ある発話とその 2 つ前の発話との語彙的結束性

(c) ある発話とその 3 つ前の発話との語彙的結束性

の 3 つがある．各素性がとる値は {強い結束性, 弱い結束性, 結束性無し, 該当発話無し} の 4 つである．値は，2 発話間に含まれる共起語対の数によって決定する．共起語対の数が 5 より大きい場合“ 強い結束性 ”，1〜5 の場合“ 弱い結束性 ”，0 の場合“ 結束性無し ”とした．“ 該当発話無し ”は，前の発話が存在しない場合に付与される．

語彙的結束性に関する素性関数の種類を**表 2** に示す．表中の各<>は 1 つの値を指す．例えば，番号 1 の<(a) x_i>は，評価発話が x_i のときの (a)，つまり発話 x_i と，その 1 つ前の発話である x_{i-1} の語彙的結束性の値となる．取り得る値は 4 種類なので，ここからは 4 つの素性関数が定義されることになる．

<>が 2 つ以上並んでいる場合は，その組み合わせである．例えば，表中の番号 10 では，発話 x_{i-1} と x_{i-2} の語彙的結束性と，発話 x_i と x_{i-1} の語彙的結束性の組み合わせを意味する．これは，x_{i-2}, x_{i-1}, x_i という，評価発話を含む 3 発話からなる部分発話系列において，隣接する発話間の語彙的結束性の強さ，すなわち，発話間の意味的な繋がりの強さを素性関数とすることを示す．番号 10 からは，全ての値の組み合わせである $4 * 4 = 16$ 個の素性関数が定義される．

これらの複数の素性関数により，発話間の繋がりの強さの分布と盛り上がりの関係も学習可能となる．例えば，**図 1** の評価発話は，話題の切り替えを行っている発話である．図では評

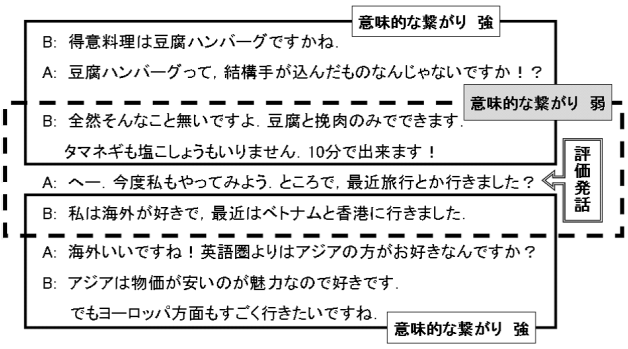

図 1: 意味的な繋がりの強さの分布

価発話を挟む形で，前後に発話間の意味的な繋がりが強い発話系列 (実線内) が存在しており，評価発話をまたぐ発話系列 (破線内) では，意味的な繋がりが弱い．この上から順に「強→弱→強」という発話間の繋がりの強さの分布は，話題を切り替えを行っている発話が存在するときに生じる分布であると考えられる．話題の切り替えを行っている発話は，盛り上がり度が高くなりにくいと考えられることから，分布と盛り上がり度の関係は非常に大きい．また，発話間の繋がりの強さの分布を使うことにより，話題が切り替わっている箇所と盛り上がり度の関係を,「ところで」のように話題を切り替える明示的な語が無かったとしても，CRF に学習させることが可能となる．

ただし，以上で説明したのは非常に単純な例である．これまで説明した語彙的結束性に関する素性関数により，さらに複雑な分布と盛り上がり度の関係も網羅的に学習可能である．

(4) 1 つ前の発話の盛り上がり度

評価発話の 1 つ前の発話の盛り上がりに関する素性では，素性関数は盛り上がり度を表すラベルの種類数ぶん定義し，加えて前述した素性 (1)〜(3) と組み合わせた素性関数を定義する．

3. 盛り上がり箇所抽出実験

3.1 実験に用いた対話

提案手法の有効性を示すために，人間同士のテキスト対話データを用いて実験を行った．

実験で使用した対話は合計 40 対話である．1 対話は 30 分とし，コンピュータを介したチャットにより収集した．また，対話は互いに面識のない被験者により行われた．

全ての発話には，評価情報が人手によって付与されている．評価は大学生 12 人によって個別に行われた．評価は前述した「話者がある発話を行った時点で，対話を継続することに対してどの程度積極的であるか」という盛り上がり度の定義に従い，評価者が最も適切だと考えるものを以下の 3 つから選択し，得点を付与した．

- 対話が盛り上がっている (1 点)
- 対話が盛り上がっていない (-1 点)
- どちらともいえない (0 点)

なお，評価の際には，評価発話に至るまでの対話も考慮する．

最終的に各発話に付与され，学習に使用されるラベルは，12 人の平均得点に応じて決定する．実験では，付与するラベルの種類数を変化させてそれぞれ実験を行う．

付与するラベルは以下のそれぞれとした

表 3: 判定実験結果

ラベルの種類数	精度	再現率	F 値
2	0.77	0.86	0.81
3	0.75	0.85	0.80
4	0.75	0.88	0.81
人手 (参考)	0.72	0.80	0.76

- 平均得点が 0.7 以上のとき「盛り上がり度：高」，それ以外は「その他」とする計 2 種類
- 上の条件における「その他」を 0.3 で 2 分割し,「盛り上がり度：中」「盛り上がり度：低」とする計 3 種類
- 上の条件における「盛り上がり度：低」を-0.1 で 2 分割し，低い方を「盛り上がり度：最低」とした計 4 種類

3.2 実験設定

提案手法の性能評価には 40 分割交差検証法を用いた．すなわち，全 40 対話のうち，39 対話をパラメータ推定のためのデータとし，残り 1 対話の発話系列に対して盛り上がり度の判定を行う．盛り上がり度の判定を行う対話を順次変更し，40 通りの実験を行う．

提案手法の性能評価は F 値で行う．F 値は精度と再現率の調和平均である．精度，再現率は，盛り上がり度の高い発話を抽出するという観点から，それぞれ以下の式で計算される．

$$\text{精度} = \frac{|A \cap B|}{|A|} \quad \text{再現率} = \frac{|A \cap B|}{|B|}$$

式中の A,B は，

A = 識別結果が「盛り上がり度：高」であった発話

B = 人手で「盛り上がり度：高」と判定された発話

である．

3.3 実験結果

表 3 に結果を示す．同表に，人手による結果を参考として示した．これは，評価者 12 人による評価結果のうち，11 人の評価の平均値を正解データ，残りの 1 人のものを評価法による判定結果と見なし，各値の計算を行ったものである．表には，判定結果と見なす 1 人を順次変更した，全 12 人の平均を示した．この結果より，提案手法はラベルの数によらず，人間が 1 人で盛り上がり度の判定を行った場合と同程度以上の性能を示すことが確認された．

盛り上がり度の高い発話を抽出するという観点からは，ラベルの種類数によらずほぼ同等の性能を示したが，盛り上がり度をより詳細に判定するという観点からは，ラベル数 4 が最も優れているといえる．

図 2 は 1 対話中の人手による評価の平均得点 (点線) と，ラベル数 4 における判定結果 (実線) をプロットした図である．横軸の番号は各発話に対応した発話番号である (つまり，この例で用いた対話は 20 発話からなる)．

ここから，提案手法の判定は，人手による評価と同様の傾向を示していることが読み取れる．このように，本手法を用いることで，自動判定した盛り上がり度の高い箇所を視覚的に捉えることも可能となる．

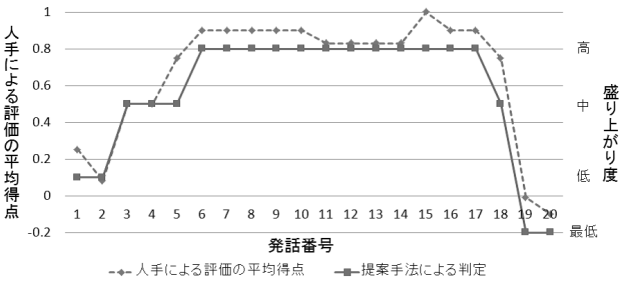

図 2: 人手による評価と提案手法による判定の比較

4. むすび

本論文では CRF を用いて，人間同士の対話における，盛り上がり度の高い発話を自動抽出する手法を提案した．本手法では，発話間の語彙的結束性を素性として用いることにより，対話中の各部分発話系列において，どの程度の密なやりとりがなされたのかや，対話中の話題ごとのまとまりの分布を把握することを目指した．実験の結果，提案手法は人手による評価と同等の性能を示すことを確認した．

今後の課題としては，本手法の音声対話への拡張が挙げられる．本手法では，様々な素性を柔軟に用いることができる CRF を用いたため，声の大きさ・トーンなどを素性とすることは容易である．ただし，音声対話の場合には,「はい」「そうです」などの単純な頷きや肯定を示す発話がテキスト対話と比べて多くなるため，語彙的結束性を確認する発話間の距離の範囲を，より大きくとる必要があると思われる．

参考文献

[1] D. Gatica-Perez, I. McCowan, D. Zhang, and S. Bengio. Detecting Group Interest-Level in Meetings. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05)*, Vol. 1, 2005.

[2] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. Eighteenth International Conference on Machine Learning Table of Contents*, pp. 282–289, 2001.

[3] G. Hirst and A. Budanitsky. Correcting real-word spelling errors by restoring lexical cohesion. *Natural Language Engineering*, Vol. 11, No. 01, pp. 87–111, 2005.

[4] 工藤拓, 賀沢秀人. Web 日本語 N グラム第 1 版. 言語資源協会発行.

[5] T. Dunning. Accurate methods for the statistics of surprise and coincidence. *Computational linguistics*, Vol. 19, No. 1, pp. 61–74, 1993.

[6] F. Sha and F. Pereira. Shallow parsing with conditional random fields. In *Proceedings of HLT-NAACL*, pp. 213–220, 2003.