単語の分散表現を用いた相槌生成タイミングの予測

Prediction of backchannel generation timing using distributed representation of words

黒田 和矢[1]　狩野 芳伸[2]

Kazuya Kuroda[1], Yoshinobu Kano[2]

[1]静岡大学大学院総合科学技術研究科情報学専攻
[1]Department of Informatics, Graduate School of Integrated Science and Technology, Shizuoka University
[2]静岡大学情報学部行動情報学科
[2]Department of Behavior Informatics, Faculty of Informatics, Shizuoka University

Abstract: In order to develop a spoken dialogue system that performs active listening in the same way as humans, we detect timing of backchannel generation using SVM. There are already many previous studies that predict generation timing of backchannels. However, most previous studies mainly use voice information with some linguistic information, such as part of speeches and ends of phrases. Meanwhile, distributed representation, such as word2vec, could express semantic relations between words. We assume that words used around backchannel generation have a close sematic relationship. In addition to features in the previous research, we use a similarity feature of words co-occurring with backchannels. We performed training and prediction of backchannel generation timing by our own backchannel corpus. Our experimental result shows that prediction performance was better when our semantic feature was used, than the features of previous study.

1. はじめに

近年, 音声対話システムの発展がめざましく, 人間とシステムとのコミュニケーションの実用化が進められている. しかし, システムは人間同士の会話でみられる相槌や頷きといった応答を行わず, 人間とシステムが交互に発話を行うことが想定されることが多い. 発話中に行われる相槌には,「続けてというシグナル」「内容理解を示す表現」「相手の意見, 考え方に賛成の意思表示をする表現」といった機能が含まれている[1]. そのため, 人間とシステムとの対話においてもシステムが人間に対して相槌を打つことで円滑なコミュニケーションを行うことができるようになると考えられる. 実際, 傾聴を行う音声対話システムに関する研究[2, 3, 4]が既に存在する. また相槌に関しても, 韻律情報[5, 6]や言語情報[7, 8, 9, 10], 視聴覚情報[11]を用いて相槌生成タイミングの検出を行う研究が存在する.

*連絡先：静岡大学大学院総合科学技術研究科
〒 432-8011 静岡県浜松市中区城北３-５-１
E-mail: kkuroda@kanolab.net

言語情報を用いて相槌生成タイミングの検出を行う際には, 文節や節の区切り, 単語の品詞, 単語の品詞が助詞であった場合その表層, 係り受けの数といった情報が用いられている. 一方, word2vec と呼ばれる単語の分散表現を用いることで単語間の意味関係を表現することができるようになった[12]. 単語の分散表現を利用した研究として, 評判分析[13], 対話破綻検出[14], 語義曖昧正解消[15]等が挙げられる. 本研究では, 相槌が打たれやすい単語間に意味的な距離があるのではないかと考え, 相槌生成タイミングの予測に単語の分散表現を用いる. また, 分散表現との比較のため単語の表層(書字形)を予測に用いる.

2. コーパスと相づち

2.1. コーパス

本研究では,「千葉大学 3 人会話コーパス」[16]を用いた. 千葉大学 3 人会話コーパスは, 千葉大学で収録された, 大学生・院生・OB を含む同性 3 人からなる友人同士 12 組の雑談が収められたものである. 各雑談について, 9 分 26 秒の音声データ, 転記テキスト, 形態論情報が収められている.

2.2. 予測単位

本研究では, 千葉大学 3 人会話コーパスの形態論情報に含まれる形態素, 及び形態素間の無音区間を相槌生成タイミングの予測を行うための基本区間として扱う. また, 本研究では先行研究[9]に倣い, 無音区間が 200 ミリ秒を超える際には, はじめの 200 ミリ秒の無音区間(sp)とそれ以降の無音区間(pause)に分けて, それぞれの区間について予測を行う.

2.3. 相槌

人間同士の相槌の頻度を分析した研究がいくつか存在しており[17, 18], 発話速度や発話内容に対する理解度によって相槌の頻度や反応時間に個人差がでるということが明らかとなっている. そのため千葉大学 3 人会話コーパス内で話者が生成した相槌のみを相槌生成可能なタイミングとして学習してしまうと, 別の人間であれば相槌生成可能だと感じる区間について相槌生成不可能として学習が行われてしまうことが考えられる. そこで, 本研究では筆者 1 人が千葉大学 3 人会話コーパスのすべての基本区間について, その区間に対して相槌を打つことができるか否かのタグ付けを行った. タグ付けを行う際には, 音声を 1 回以上聞いた上で, 形態論情報に対してタグを付与した. 表 1 にタグ付けを行った千葉大学 3 人会話コーパスの規模を示す.

表 1: タグ付けしたコーパスの規模

発話数	7,599
形態素	29,139
無音区間	5,997
相槌タグ	3,447
単語の種類	2,652

3. 素性

本研究では, 千葉大学 3 人会話コーパス中の連続する基本区間(コーパス中の形態素及び無音区間) m_1, …, m_n に対して, SVM を用いて, 基本区間 m_i の直後に相槌を生成できるか否かの予測を行う.

3.1. 基本素性

本研究において, 予測の基本となる素性を表 2 に示す. これらの素性は先行研究[9]で行われた SVM を用いた相槌生成タイミングの検出に使用された素性に基づいて決定した. ここで, 表 2 中の m_j は基本区間 m_i の直前に現れた形態素区間を表す. また, X_1 には UniDic の品詞分類のうち大分類に属する品詞[19]のいずれかが入る. X_2 には, 「名詞」「動詞」「形容詞」「その他の品詞」のいずれかが入る. なお, 本研究では, 文節の区切りを KNP[1], その他の素性を形態論情報から求めた.

表 2: SVM の学習に用いる基本素性

1.	m_j が文節の最終形態素であるか否か
2.	1 が真のとき, m_j の品詞が X_1 であるか否か
3.	1 が真のとき, m_j が属する文節内に X_2 が存在するか否か
4.	m_i が sp であるか否か
5.	4 が真のとき, mi のポーズ長(秒)
6.	m_i が pause であるか否か
7.	m_j を構成する最後の 1 モーラの時間長(秒)
8.	m_j の発話速度が平均発話速度より遅いか否か
9.	8 が真のとき, m_j の発話速度と平均発話速度との差(秒)

3.2. 追加素性

本研究では, 表 **3** に示す素性を基本素性とともに用いて相槌生成タイミングの予測を行う. 表 **3** 中の X_3 には千葉大学 3 人会話コーパスに含まれる単語のいずれかが入る. X_4 には「じゃん」「から」「ない」「けど」「ね」のいずれかが入る.

表 3: SVM の学習に用いる追加素性

1.	m_j の表層が X_3 であるか否か
2.	m_j と X_4 とのコサイン類似度

本研究では, 単語の分散表現を求めるために, Twitter から収集したツイートデータ約 500 万件を用いて word2vec の Skip-Gram モデルを構築した. 分散表現の次元数は 30 次元, 実装には TensorFlow を用いた. X_4 に入る単語を選択する際には, 相槌タグがつけられた形態素のうち, 出現回数が高いものから順に 5 つ選んだ. 順に選ぶ際, 既に選ばれた単語とのコサイン類似度が 0.95 以上の単語は選ばないようにした.

4. 評価実験

本研究では, タグ付けを行った千葉大学 3 人会話コーパスに対して, 基本素性及び追加素性を用いて相槌生成タイミングの学習及び評価を行う.

[1]http://nlp.ist.i.kyoto-u.ac.jp/?KNP

表 4: 素性の組み合わせと評価結果

	素性の組み合わせ	Accuracy	Precision	Recall	F1
1.	基本素性のみ	83.90	83.83	83.99	83.90
2.	基本素性 + 単語の表層(3 回以上)	83.54	83.48	83.63	83.55
3.	基本素性 + 単語の表層(50 回以上)	84.86	84.26	85.74	84.99
4.	基本素性 + コサイン類似度	86.09	86.61	85.39	85.99
5.	基本素性 + 単語の表層(3 回以上) + コサイン類似度	83.89	83.2	84.86	84.05
6.	基本素性 + 単語の表層(50 回以上) + コサイン類似度	84.51	83.00	86.80	84.85

基本区間	まっとう	な	意見	な	ん	だ	よ
正解							
予測結果		○					

図 1: 誤った相槌生成タイミングの予測の例

4.1. 実験概要

コーパス内に収められた 12 ファイルのうち 10 ファイルを学習データ, 残りの 2 ファイルを評価データとして使用した. 使用したデータの数は表 5 に示すとおりであり, 学習データ, 評価データそれぞれに対してアンダーサンプリングを行い, 正例及び負例の数を揃えた. SVM の学習には RBF カーネル使用し, 表 6 に示す範囲でグリッドサーチを行いコストパラメータ c, RBF カーネルのパラメータ γ の値を決定した. なお, 学習及び評価には LIBSVM[2] を用いた.

表 5: 学習及び評価に用いたデータ数

	正例	負例
学習データ	2876	2876
評価データ	568	568

表 6: グリッドサーチの範囲

c	2 ^ n (n = -5, -3, …, 13, 15)
γ	2 ^ n (n = -15, -13, …, 1, 3)

千葉大学 3 人会話コーパスに含まれる単語の中には, 出現回数が 1 回のみである単語が複数存在する. これらの単語を表 **3** の X_3 の素性として加えることで, 学習に偏りが生じる可能性がある. そのため本研究では表 **3** の X_3 に用いる単語について, コーパス内での出現回数が 3 回以上, 又は 50 回以上である単語のみを用いて実験を行った. X_3 に用いる単語数は, 単語の出現回数が 3 回以上の時に 912 個, 50 個以上のときは 70 個であった. 本研究で行った基本素性と追加素性の組み合わせを表 4 に示す.

[2]https://www.csie.ntu.edu.tw/~cjlin/libsvm

4.2. 実験結果と考察

基本素性と追加素性の組み合わせそれぞれについて評価を行った結果を表 4 に示す. 表 4 より, 基本素性に加え, コサイン類似度のみを使用した場合に, 最も高い精度が出ていることが見て取れる. 本研究では予測を行う際に基本区間とコサイン類似度を比較する単語として助詞, 助動詞である「じゃん」「から」「ない」「ね」「けど」を用いた. コーパス内の相槌タグが付与された形態素の多くが助詞, 助動詞であったため, 比較用の単語のいずれかとコサイン類似度の高い場合に相槌を生成しやすいといった学習が行われ, 相槌タグが付与された形態素に対して適切に相槌を生成することができたのではないかと考えられる. 一方単語の表層を素性に用いた場合, 素性に用いる単語の数によって予測精度に与える影響が異なることが見て取れる. コーパス内に含まれる単語のうち出現頻度の高い単語には助詞, 助動詞が多いため, 頻度の高い単語の表層を素性に用いた際にはタイミング予測の精度が若干向上したのではないかと考えられる.

最も精度の高かった基本素性と単語間のコサイン類似度を用いた評価を行った際に, 相槌タグの付与されていない箇所で予測時に相槌生成可能であると誤った例を図 1 に示す. 図 1 中の誤り箇所の形態素「な」はコサイン類似度の比較用の単語「ね」と高い類似度を示す. 先程述べたようにコサイン類似度が高い場合に相槌を生成するよう学習が行われたことにより, 誤って相槌生成可能であると判断したと考えられる. 図 1 中の誤って予測した箇所にお

いて相槌生成可能であるかどうかを発話内容のみを用いて判断することは難しい．このような誤りを避けるためには言語情報のみならず，韻律情報を用いる必要があるだろう．

いずれにせよ，実験に使用したコーパスの傾向に依存する部分が大きいと考えられるため，他のコーパスでも実験を行う必要があるだろう．また，正解データとなる相槌タグについても，今回は一人がタグ付けを行っている．そのため，アノテーターによる偏りも考慮する必要があるだろう．複数人でタグ付けを行ってその一致率をはかると同時に一定数以上のアノテーターが付与したタグを正解とする，といった対応が考えられる．

5. おわりに

本研究では，単語の表層，単語間のコサイン類似度を素性とした SVM を用いた相槌生成タイミングの予測を行った．実験の結果，単語の表層を素性とした際には素性に用いる単語の個数によって相槌生成タイミングの予測精度に与える影響が異なることが明らかとなった．一方，コサイン類似度を素性とした際には基本素性のみを用いた相槌生成タイミングの予測結果と比較して精度が向上することが明らかとなった．しかし，コサイン類似度を素性とした際には比較用の単語とのコサイン類似度が高い単語が出現した際に誤った相槌の生成が見受けられた．今後はタイミング予測の精度をより向上させるため，素性として使用する単語の変更や，素性の追加を行うことを検討している．また同時に使用するコーパスの変更，コーパスへの相槌タグの安定性の向上を図ることでより様々なデータに対して高い精度で予測を行うことができるモデルの構築を試みる．

参考文献

[1] 泉子・K・メイナード: 会話分析，くろしお出版 (1993)

[2] 下岡和也, 徳久良子, 吉村貴克, 星野博之, 渡部生聖: 音声対話ロボットのための傾聴システムの開発，自然言語処理, Vol. 24, No. 1, pp. 3–47 (2017)

[3] 青木裕哉, 伊島翔大, 中野大樹, 藤江真也: 質問への応答に依存しない傾聴対話システムのリアクション生成，情報処理学会第 80 回全国大会講演論文集, Vol. 2018, No. 1, pp. 397–398 (2018)

[4] 石田真也, 井上昂治, 高梨克也, 河原達也: 共感・発話促進のための多様な聞き手応答を生成する傾聴対話システム，情報処理学会第 80 回全国大会講演論文集, Vol. 2018, No. 1, pp. 403–404 (2018)

[5] 岡登洋平, 加藤佳司, 山本幹雄, 板橋秀一: 韻律情報を用いた相槌の挿入，情報処理学会論文誌, Vol. 40, No. 2, pp. 469–478 (1999)

[6] 竹内真士, 北岡教英, 中川聖一: 韻律情報を用いた相槌生成システムとその評価，情報処理学会全国大会講演論文集, Vol. 64, No. 2, pp. 101–102 (2002)

[7] 西村良太, 中川聖一: 応答タイミングを考慮した音声対話システムとその評価，情報処理学会研究報告, Vol. 2009, No.22, pp. 1–6 (2009)

[8] 山口貴史, 井上昂治, 吉野幸一郎, 高梨克也, Nigel G. Ward, 河原達也: 傾聴対話システムのための言語情報と韻律情報に基づく多様な形態の相槌の生成，人工知能学会論文誌, Vol. 31, No. 4, pp. 1–10 (2016)

[9] 神谷優貴, 大野誠寛, 松原茂樹: 音声対話コーパスに基づくあいづち生成タイミングの検出とその評価，言語処理学会年次大会発表論文集, Vol. 17, pp. 103–106 (2011)

[10] 竹内真士, 北岡教英, 中川聖一: 韻律・表層的言語情報を発話タイミング制御に用いた雑談対話システム，情報処理学会研究報告, Vol. 15, pp. 87–92 (2004)

[11] 佐野正太郎, 日下航, 尾形哲也, 高橋徹, 奥乃博: 神経力学モデルの引き込みによる相槌タイミングの予測，情報処理学会全国大会講演論文集, Vol. 73, No. 2, pp. 89–90 (2011)

[12] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean: Efficient estimation of word representations in vector space, ICLR Workshop (2013)

[13] 加藤和平, 大島考範, 二宮崇: Word2Vec と深層学習を用いた大規模評判分析，言語処理学会 第 21 回年次大会 発表論文集, pp. 525–528 (2015)

[14] 稲葉通将，高橋健一：Long Short-Term Memory Recurrent Neural Network を 用いた対話破綻検出，人工知能学会研究会資料 SIG-SLUD-B502, p. 57–60 (2015)

[15] 菅原拓夢, 笹野遼平, 高村大也, 奥村学: 単語の分散表現を用いた語義曖昧性解消，言語処理学会 第 21 回年次大会 発表論文集, p. 648–651 (2015)

[16] Den, Y., Enomoto, M: A scientific approach to conversational informatics: Description, analysis, and modeling of human conversation, Conversational informatics: An engineering approach, pp. 307–330 (2007)

[17] 羅希, 定延利之: 日本語の相づちの頻度とタイミングに関する総合的考察，日本認知科学会第 32 回大会, pp. 171–180 (2015)

[18] 上野洋, 井上雅史: 相槌に個性を用いたテキスト対話システム，情報処理学会研究報告 音声情報処理, Vol. 15, No. 10, pp. 1–9 (2015)

[19] 小木曽智信, 中村壮範:『現代日本語書き言葉均衡コーパス』形態論情報データベースの設計と実装 改訂版，国立国語研究所内部報告 (2011)