2H2-4

雑談対話システムにおける前後の発話の繋がりを用いた最適発話選択手法の検討

A Study for Suitable Utterance Selection Using Adjacency in Chat Dialogue Systems

成松宏美 *1 Hiromi Narimatsu　杉山弘晃 *1 Hiroaki Sugiyama　堂坂浩二 *2 Kohji Dohsaka　東中竜一郎 *1 Ryuichiro Higashinaka

*1日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
Nippon Telegraph and Telephone Corporation, NTT Communication Science Laboratories

*2秋田県立大学システム科学技術学部電子情報システム学科
Akita Prefectural University, Faculty of Systems Science and Technology, Department of Electronics and Information Systems

For chat with people, it is required to natural utterances as people do. Although the conventional chat dialogue systems prepare a database of question and answers, it is difficult to build enough database for chat because of the breadth of the topic. To make the problem simple, we tackle the problem of finding the unnatural utterances from the pair of sentences. To find or detect unnatural pair of utterances accurately will lead to the realization of choosing a suitable utterances. We will use the problems in a dialogue to evaluate our utterance selection performance.

1. はじめに

近年，人と対話するエージェントへの期待が高まり，決められたタスクに関して質問応答を行うタスク対話だけではなく，雑談対話においても人と同じように自然な応答ができることが求められている．ドメインが限定されない雑談対話において，人の発話を正確に解釈し，更に文脈に合わせて自然な応答を返すことは非常に難しい課題である．雑談対話における応答生成手法として，従来のタスク対話同様に予め手作業により作成したパタンをルールとして記載しておくものと [Wallace 04]，ウェブ上のあらゆる大規模データを用いて，統計的に適切な発話を選択して応答するもの [Murao 03] とがある．これらはいずれも，候補群の中から適切な発話を選択するものであるため，最適発話選択の精度向上が重要な課題である．

最適発話の選択を目的として，人の発話文から発話行為や発話意図を推定し，自然な発話の流れを考慮する手法が提案されている [Higashinaka 14, 東中 15]．発話文そのままではなく抽象化された情報を用いることで，ドメインが限定されない雑談対話においても，ある程度の対話を実現しているが，抽象化のために，細かい情報を考慮することが難しい．ユーザの発話文に現れる単語をそのまま入力として，システムの次の発話を自動で生成する手法も提案されているが [Inaba 16]，本手法もニューラルネットワーク内で抽象化が行なわれている．

一方，堂坂らは，文からの特徴量をできる限り抽象化せずに，そのまま用いる手法を提案している [堂坂 16]．堂坂らは，「ロボットは東大に入れるか」プロジェクト [Arai 14] の英語問題のうち，2 話者間の対話中に含まれる空所を埋める会話文完成問題に対して取り組み，前後の発話の繋がりを各発話の n-gram の特徴量のペアを用いて SVM で学習し，全体の感情や発話意図の流れの自然さを用いる手法を提案している．提案手法は，従来手法の発話意図や感情極性などの抽象化された特徴量のみを用いるよりも良い精度が得ている．SVM は，特徴量をそのまま分類に用いるため，予め抽出した n-gram の特徴量のペアが有効である場合には精度良く分類できる一方で，抽出していない特徴量の組み合わせは考慮されない．従って，特徴量ペアの可能性を考慮しきれていない可能性がある．

我々は，抽象化を行う手法と，発話文そのままから得られる特徴量を用いる手法の良さを生かすために，ニューラルネットワークと SVM のハイブリット手法を提案する，本稿では，堂坂らと同様に会話文完成問題 [堂坂 16] に取り組み，その正答率で提案手法を評価する．会話文完成問題は，日常に起こりうる雑談対話中に空所がある問題の一つと捉えることができ，本問題の精度向上は，エージェントが自身の対話の評価できること，更に対話中に自身の対話の誤りに気づきリカバリできることに寄与すると考える．

連絡先: 成松宏美，NTT コミュニケーション科学基礎研究所，京都府相楽郡精華町光台 2-4，0774-93-5311，Fax:0774-93-5245，narimatsu.hiromi@lab.ntt.co.jp

Greg　: Are you doing anything Friday night?
Michio : Nothing special. Why?
Greg　: My wife and I would like to invite you and Ayako to our house for dinner.
Michio : [18]
Greg　: OK, maybe some other time.

18 の選択肢 :
1. I'll bring the bottle of Japanese sake her father sent us.
2. I'm sorry but we're having dinner at Arthur's then.
3. Thanks. I'm free but I'm afraid Ayako has other plans.
4. That's very kind of you. What time shall we come?

図 1: 会話文完成問題の例.

2. 会話文完成問題とそのアプローチ

会話文完成問題の 1 例を図 1 に示す．空所 [18] に当てはまる発話を選択肢 1〜4 から回答する問題である．本例では，金曜日の夜に Michio が特に予定がないのに対して，Greg が夕飯の誘いをしているが，最後の Greg の発言より，Michio は Greg の誘いに対して何らかネガティブな発言をしたことが読み取れる．従って，3 が正解であるとわかる．このように，空所の次を見ることで初めて空所に最適な発話が選択できる問題に取り組む．

会話文完成問題に対して，堂坂らは，前後の発話のつながりの自然さ，すなわち隣接発話らしさ，と感情極性を用いる解法を提

案し，従来の発話意図の流れの自然さを用いた手法よりもスコアが上がることを示している [堂坂 16]．隣接発話らしさを用いた手法では，図 2 に示すように，隣接する文の 1-gram，2-gram，3-gram のペアを入力とし SVM で，学習している．しかし，学習データが大量になると，予めペア作成する手間と，ペア数すなわち入力ベクトルの次元が大きくなる懸念が生じる．ある程度の出現頻度で切り落とすことも可能であるが，切り落としにより必要な情報量が落ちることが考えられる．そこで，本稿では，できる限り情報量を落とさずに入力ベクトルの次元を抑えて学習する手法として，n-gram ベクトルの連結をニューラルネットワークで学習する手法に着目する．

ニューラルネットワークを用いた対話システムにおける最適発話の選択では，対象となる文をそのまま入力層に与え，隠れ層にて特徴量間の関係性をうまく学習できることが期待されている．例えば，Vinyals らは，チャットボットに対して，ユーザの発話を文として入力し，次のシステムの発話を生成する *seq2seq* モデルを用いる手法を提案している [Vinyals 15]．本手法では，応答にあまり揺らぎがない対話においては，適切な応答を生成するが，ドメイン限定しない対話では，適切な応答を生成することが難しい．従って，我々の扱う雑談のような対話においては，効果を得ることが難しい．

本稿では，特徴量作成に依存しない，SVM とニューラルネットワークのハイブリット手法を提案する．

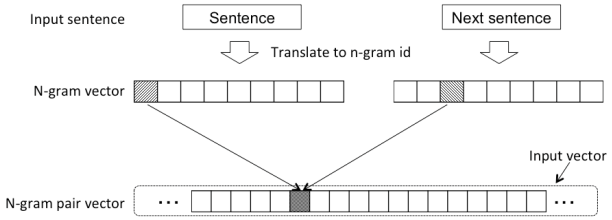

図 2: 従来手法の入力ベクトルの生成法.

3. 提案手法

前後の発話のつながり，すなわち隣接発話らしさにおいて，文そのままの入力から得られる特徴量と，n-gram ペアや抽象化された特徴量などの両方を考慮するために， 特性の異なる複数の判別器を組み合わせることを提案する．入力ベクトルの作成方法と学習方法についてそれぞれ述べる．

入力ベクトルは，従来手法と異なり，図 3 に示すように，隣接する発話からそれぞれ作成した n-gram ベクトルを，前後の順で保持しながらそのまま連結することで作成する．従って，2 文の隣接確率を表現するベクトルのサイズは，n-gram ベクトルの 2 倍となる．

判別器には，従来手法 [堂坂 16] の隣接発話らしさの判定に用いられていたサポートベクタマシン（SVM）と，多層パーセプトロンを表現するニューラルネットワーク [LeCun 98] とを組み合わせる．ニューラルネットワークは，隠れ層にて入力ベクトルの共起関係を学習することが可能である．また，各ベクトルを適切に抽象化し，細かな表記ゆれなどに頑健な学習ができる可能性がある．逆に特徴量を抽象化せず，詳細な違いを考慮できる SVM と組み合わせることで，相互の長所を活かした学習器を構築できると考えられる．

図 4 にニューラルネットワーク内で想定される n-gram ペアの表現について示す．最上列は入力ベクトル，2，3 列目の丸は

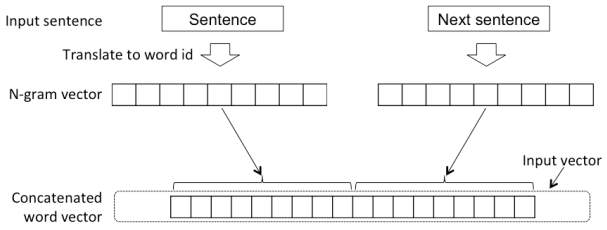

図 3: 提案手法の入力ベクトルの生成法.

隠れ層の各ユニットを表し，最下列は出力のクラス，すなわち隣接発話らしさの 2 値分類であることを示す．入力は各文から生成した n-gram のベクトルの連結であるが，隠れ層にて，グレーに塗られたユニットのように 2 つの入力ベクトル同士の結合を表現できることが期待できる．更に，層を増やすことで，ベクトル同士が結合されたユニット同士の結合も考慮できると考えられる．

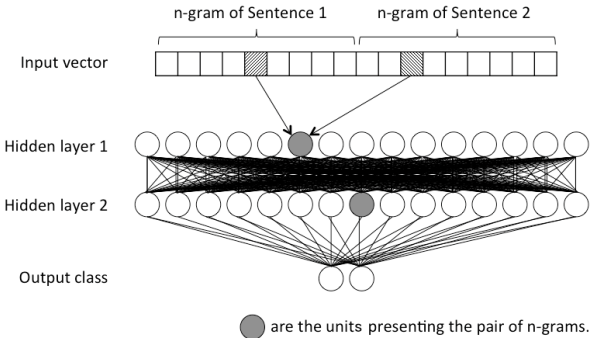

図 4: ニューラルネットワーク内での n-gram ペアの表現.

本稿では，2 つの異なる判別器の結果を，隣接発話らしさのスコアに基づき選択する．S 個の選択肢のうち 1 つの選択肢 s に対して，各判別器 X_1，X_2 が算出した隣接発話らしさのスコアを，$p_s(X_1)$，$p_s(X_2)$ であるとする．このとき，各判定器内のスコアの和を判別器の最適発話の選択スコアであるとして，そのスコアが最大となる発話 s^* を選択する．

$$s^* = \underset{s \in S}{\operatorname{argmax}} \sum_{m=1}^{M} (p_s(X_m) \cdot w_m)$$

$m \in M$ は用いる判別器の識別子，w_m は重み係数を表す

学習するデータは，従来手法同様に．NTT のシチュエーション対話データを用いる．シチュエーション対話は，1 つのシチュエーションに対して起こりうる 2 話者の対話を想定して，10 発話からなる 1 対話を人手により作成したものである．場所と，属性，例えば，天気，スポーツ，ショッピング等や，声明，要求，質問等，複数指定し，試験問題に出現しやすい日常的なシチュエーションを想定した．作成した対話中で，誤字を含むものや対話の繋がりが不適切であった対話を除いた 6802 対話を学習に用いる．シチュエーション対話の例を，図 5 に示す．本対話例は，場所に対してホテルを指定して作成されたものである．A/B は，話者，添字は各話者の n 回目の発話であることを示す．本対話データから正例と負例を作成する．正例と負例が

1:1 となるように，次のように作成する．

正例の作成：正例は，$\{A_1,B_1\}$，$\{B_1,A_2\}$，$\cdots$，のように，隣接する話者の発話を前後につながる発話として，各対話から 9 個の正例を作成する．

負例の作成：負例は，同一のシチュエーションに対する全く異なる対話 K，L の組みを用いる．対話 KA_n 番目の発話を，A_n^K，と表現するとき，$\{A_1^K,B_1^L\}$，$\{B_1^K,A_2^L\}$，$\cdots$，のように，同様に各対話の組みから $2 \times 9 = 18$ 個の負例を作成する．尚，発話は必ず話者 A から始まるものとする．

作成した正例と負例の文の対を用いて，上記で示した n-gram のベクトルに変換し，前後関係を保持しながら結合したベクトルを入力として，2 つの判別器でそれぞれ学習する．

A_1: Is this your first time staying here?
B_1: Yes, it is. And you?
A_2: It is my first time as well, how are you finding your stay?
B_2: It's been disappointing.
A_3: Why is that?
B_3: The heat is off in my room. It is very cold.
A_4: That's a shame.
B_4: How is your room?
A_5: Now that you mention it, my room is cold, too.
B_5: They need to fix their heating.

図 5: 学習データの対話例.

4. 評価実験

4.1 実験環境

提案手法を用いて隣接発話らしさを算出したときの，会話文完成問題の正答率を評価する．評価には，大学入試センター試験の本試験過去問 32 問と追試過去問 29 問 (1991〜2015 年度の奇数年度および 2014 年度)，代ゼミセンター模試 15 問 (2013 年第 1〜3 回および 2014 年第 1 回)，ベネッセ模試 9 問 (2014 年 6 月，9 月，11 月) と，独自に収集したその他の問題 73 問の合計 158 問を用いる．Section 3. で述べた学習データより，頻出上位 σ 個の n-gram を抽出し前後の文で連結した，$\sigma \times 2$ 次元の n-gram ベクトルを用いる．SVM には，LIBLINEAR [Fan 08] を用い，コスト関数は，e^3 を用いる．ニューラルネットワークは隠れ層 2 層の，4 層多層パーセプトロンネットワークを用いる．隠れ層のユニット数は 5000，最適化関数には，Adam[Kingma 14] を用いた．

評価では，各判別器 NN（ニューラルネットワーク），SVM による正答率の比較と，提案手法であるハイブリット手法と従来手法との正答率を比較する．比較手法は，選択した答えに対する隣接発話らしさのスコアが高い方の判別器が選択した回答を最適であるとする Hybrid Conventional，選択肢毎の分散の大きい判別器が選択した回答を最適であるとする Hybrid Variance，判別器のスコアの和が最大の選択肢を最適であるとする提案手法;Hybrid Summation，どちらかの判別器で正解を選択できた場合に正解とする Hybrid Reference，n-gram の隣接ペアを用いる Conventional[堂坂 16] を比較する Hybrid Variance については，最大スコアを持つ選択肢と他の選択肢との差分が大きい程，システムが迷いなく選択しているものと考えられる可能性を想定し，比較対象とした．分散は，

$$V(X) \;=\; \frac{1}{2}\sum_{m=1}^{M}(p_s(X_m)-\mu)^2$$

(μ は $p_s(X)$ の平均値を表す）により算出し，分散 $V(X)$ の大きい方の判別器が選択した最適発話を，システムが選択する最適発話であるとする．提案手法の重み係数は判別器毎に均一の $w_m = 0.5$ を用いた．

4.2 評価結果

表 1 に，会話文完成問題の正答率の結果を示す．列は用いた手法であり，それぞれ NN（ニューラルネットワーク），SVM であることを示す．手法のカッコ内は，n-gram ベクトルの次元数 σ の値を表す．NN と SVM のハイブリット手法は，センター追試以外は，全て正答率が向上していることがわかる．また，NN の $\sigma = 5000$ と $\sigma = 7000$ を比較すると，センター以外は全て向上している．このことから，特徴量が精度向上に寄与することがわかる．また，SVM，NN いずれにおいても，n-gram 特徴量をそのまま用いるだけでは，チャンスレートであった．

表 1: 各判別器の正答率.

	NN (5000)	NN (7000)	SVM (7000)
センター	0.21 (7/32)	0.13 (4/32)	0.21 (7/32)
センター 追試	0.17 (5/29)	0.20 (6/29)	0.17 (5/29)
代ゼミ	0.26 (4/15)	0.33 (5/15)	0.2 (3/15)
ベネッセ	0.33 (3/9)	0.55 (5/9)	0.33 (3/9)
その他	0.26 (19/73)	0.30 (22/73)	0.28 (21/73)
合計	0.24 (38/158)	0.26 (42/158)	0.24 (39/158)

次に表 2 に，従来手法と提案手法の正答率の比較結果を示す．各列は，比較手法，Hybrid Conventional，Hybrid Variance，Hybrid Summation，Conventional を示し，各行は評価対象問題を示す．尚，本実験では，Conventional で使用しているデータセットの一部を用いているため，問題数が異なるため，割合で比較する．

結果より，スコアの和をシステムの最適発話選択スコアとすることで，各判別器単体のスコアよりも正答率が向上することがわかった．また，どちらかの判別器の結果を用いる場合には，最大スコアを示す判別器の結果を選択するより，分散の大きい判別器を回答を選択する方が正答率が向上することがわかった．また，Hybrid Reference では，判別器毎の平均正答率 0.25 よりも約 1.7 倍正答率が向上することから，各判別器 SVM 及び NN の両者が正答の選択に対して補完関係にあることがわかる．尚．正解だったものも含め，全問題に対する各判別器が選択した回答の一致率は，0.64 であった．

5. おわりに

本稿では，前後の発話のつながりを特徴量として，異なる特性を持つ複数の学習手法を組み合わせることで，隣接発話らしさを求める手法を提案した．各発話から得られる特徴量をそのまま結合し，サポートベクタマシンとニューラルネットワークを用いてそれぞれスコアを算出することで，予め前後の発話のペアの特徴量を作成しなくても，同等程度の精度が得られることがわかった．これにより，前後の 2 文の隣接だけでなく 3 文，4 文の隣接についても，組み合わせを考慮することによる特徴

表 2: 提案手法及び比較手法の正答率.

	Hybrid Conventional (large value)	Hybrid Variance (large variance)	Hybrid Proposed (large summation)	Hybrid Reference (oracle)	Conventional [堂坂 16]
センター	0.15 (5/32)	0.25 (8/32)	0.21 (7/32)	0.31 (10/32)	0.37 (19/51)
センター 追試	0.20 (6/29)	0.13 (4/29)	0.34 (10/29)	0.31 (9/29)	0.30 (14/47)
代ゼミ	0.33 (5/15)	0.33 (5/15)	0.27 (4/15)	0.46 (7/15)	0.50 (9/18)
ベネッセ	0.55 (5/9)	0.77 (7/9)	0.56 (5/9)	0.77 (7/9)	0.67 (6/9)
その他	0.30 (22/73)	0.34 (25/73)	0.37 (27/73)	0.50 (36/73)	0.38 (47/114)
合計	0.27 (43/158)	0.31 (49/158)	0.34 (53/158)	0.43 (69/158)	0.38 (91/239)

数の爆発を抑えて学習できる可能性を示した．本稿では，判別器の算出したスコアの和を最適発話の選択に用いた場合に，最も良い正答率となることがわかった．判別器のどちらかで正答しているものを含めると，更に正答率が向上することから，重み付けを考慮することにより更なる精度向上も期待できる．

今後は，複数文の隣接を用いた手法を検討する．また，従来手法に用いられていたような発話意図推定の結果や感情極性のスコアを用いることで，更なる精度向上を目指す．

謝辞

本研究を推進するにあたって，大学入試センター試験問題のデータをご提供くださった独立行政法人大学入試センター及び株式会社ジェイシー教育研究所に感謝いたします．本研究を推進するにあたって，実験データをご提供くださいました学校法人高宮学園，株式会社ベネッセコーポレーションに感謝いたします．

参考文献

[Arai 14] Arai, N. H. and Matsuzaki, T.: The impact of A.I. on education—Can a robot get into the University of Tokyo?, *Proc. of international conference on consumer electronics (ICCE)*, pp. 1034–1042 (2014)

[Fan 08] Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., and Lin, C.-J.: LIBLINEAR: A library for large linear classification, *Journal of Machine Learning Research*, pp. 1871–1874 (2008)

[Higashinaka 14] Higashinaka, R., Imamura, K., Meguro, T., Miyazaki, C., Kobayashi, N., Sugiyama, H., Hirano, T., Makino, T., and Matsuo, Y.: Towards an open domain conversational system fully based on natural language processing, *Proc. of the 25th International Conference on Computational Linguistics*, pp. 928–939 (2014)

[Inaba 16] Inaba, M. and Takahashi, K.: Neural Utterance Ranking Model for Conversational Dialogue Systems, *Proc. of the 17th Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL48)*, pp. 393–403 (2016)

[Kingma 14] Kingma, D. and Adam, J. B.: Adam: A Method for Stochastic Optimization., *Proc. of International Conference on Learning Representation (ICLR)*, pp. 2278–2324 (2014)

[LeCun 98] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P.: Gradient-based learning applied to document recognition., *Proc. of the IEEE*, No. 11, pp. 2278–2324 (1998)

[Murao 03] Murao, H., Kawaguchi, N., Matsubara, S., Yamaguchi, Y., and Inagaki, Y.: Example-based spoken dialogue system using woz system log, *Proc. of the 4th SIGdial Workshop on Discourse and Dialogue*, pp. 140–148 (2003)

[Vinyals 15] Vinyals, O. and Le, Q. V.: A Neural Conversational Model, *arXiv preprint* (2015)

[Wallace 04] Wallace, R. S.: The Anatomy of A.L.I.C.E., *A.L.I.C.E. Artificial Intelligence Foundation, Inc.* (2004)

[東中 15] 東中竜一郎, 杉山弘晃, 磯崎秀樹, 菊井玄一郎, 堂坂浩二, 平博順, 南泰浩：センター試験における英語問題の回答手法, 言語処理学会第 21 回年次大会発表論文集, pp. 187–190 (2015)

[堂坂 16] 堂坂浩二, 坂本祐磨, 高瀬惇：隣接発話らしさを利用した英語会話文完成問題の回答手法, *Proc. of the 30th Annual Conference of the Japanese Society for Artificial Intelligence*, pp. 1–4 (2016)