1B2-OS-25b-3

雑談対話ログを用いた話者の潜在的興味対象の推定

Latent interest estimation using open-domain conversational dialogues

縄手 優矢 *1 Yuya NAWATE　稲葉 通将 *1 Michimasa INABA　高谷 智哉 *2 Tomoya TAKATANI　山田 整 *2 Hitoshi YAMADA　高橋 健一 *1 Kenichi TAKAHASHI

*1 広島市立大学 Hiroshima City University　*2 トヨタ自動車 Toyota Motor Corporation

If a chat-oriented dialogue system can estimate user's latent interest, it can communicate him/her more smoothly. In this paper, we propose a method to estimate speaker's latent topic of interest using dialogue data. The model receives a dialogue and nouns that do not appear in the dialogue, and estimates speaker's interest to the nouns The model consists of multiple neural networks and they are updated by one loss function. Experimental result indicated that the performance of the proposed model could estimate latent interest appropriately in comparison with baseline models.

1. まえがき

近年，エンターテイメントなどを目的とした雑談対話システムの研究が増加している．こういったシステムが人間の生活に浸透するためには円滑なコミュニケーションが不可欠であり，そのための心理的な働きを表す概念として社会的スキルがある [大坊 06]．その構成要因として，人をどう見るかという対人認知や感情の察知・推測をはじめとしたメタ・コミュニケーションなどが挙げられる．システムがこの社会的スキルを駆使してユーザの情報を良く知ることができれば，システムと人間とが円滑な関係を築けるだろう．

人間どうしの雑談では，会話に出現した情報やそこからの推測によって，出身地や性格など相手の情報を抽出している．こういった対話相手の情報の中でも，特に「興味をもっていること」についての話題はコミュニケーションをより円滑にしやすいと考えられる．そこで本研究では，対話システムによるユーザの興味対象の推定に焦点を当てる．

人の対話から話者の興味を推定した研究は Schuller らによるものが知られている [Schuller 06]．この研究では，ある製品に関して紹介する話者と，その聞き手の 2 者間の対話で，聞き手側の製品に対する興味を推定するというタスクを対象としている．音響的特徴と言語的特徴を用い，SVM により 3 段階で興味を推定した．また，同じタスクを対象とした研究として，Wang らの研究もある [Wang 13]．ただし，これらの研究では興味を推定する対象はあらかじめ決められており，任意の話題に関する興味推定は対象としていない．

平野らは対話から任意のユーザに関する情報を < 述語項構造，エンティティ，人物属性，トピック > の構造化された形で機械学習を用いて抽出する手法を提案している [平野 16]．また，Bang らは対話を通じて < エンティティ1，エンティティ2，エンティティ1 と 2 の関係 > の形式でユーザ情報を獲得し，獲得した情報から応答を生成する対話システムを提案した [Bang 15]．しかし，これらの研究において興味の有無は考慮しておらず，また，抽出可能なのは対話中に出現したエンティティについてのみである．一方，対話中に登場しない話題に対しても興味推定ができれば，より円滑なコミュニケーションにつながると考えられる．そこで，本論文では特に，ユーザが興味をもっているものの直接会話には現れていない潜在的興味対象を推定することを目的として，ニューラルネットワークを用いたモデルを提案する．

なお，本研究では対話データとして，2 名の話者による人間同士のテキスト対話ログを用いる．

連絡先: 広島市立大学情報科学部
〒 731-3194　広島市安佐南区大塚東 3-4-1
E-mail: nawate@cm.info.hiroshima-cu.ac.jp

2. 潜在的興味対象推定手法

2.1 概要

話者 u_1 に対応する潜在的興味推定対象の単語集合を $w_{u_1} = (w_1^{u_1}, w_2^{u_1}, ..., w_n^{u_1})$，および u_2 に対応する単語集合を $w_{u_2} = (w_1^{u_2}, w_2^{u_2}, ..., w_m^{u_2})$ とする．n と m はそれぞれ単語数であり，話者ごとに異なる．本研究における潜在的興味推定とは，話者 u_1 と u_2 による対話ログ c_{u_1,u_2} が与えられ，潜在的興味推定対象単語集合 w_{u_1} と w_{u_2} におけるそれぞれの単語に対し，各話者が興味を持つか否かを 2 クラスで正しく分類することと定義する．ただし，今回は w_{u_1} と w_{u_2} はあらかじめ与えられるものとし，潜在的興味推定対象の生成については扱わない．

提案モデルの概観を図 1 に示す．モデルはまず，対話ログ中の全名詞に対して話者が興味を持っているか否かを推定する．そのために，対象名詞を含む発話を LSTM(Long Short-Term Memory) を中間層に用いた Recurrent Neural Network(RNN) によってベクトル化し (図の①)，この文ベクトルを対話ログ中の名詞のための Neuwal Network(NN)(図の②) に入力して興味推定する．ここで「興味あり」と判定された単語は対話外名詞の興味推定の手掛かりとするため，当該文ベクトルを保存する．最後に，保存した文ベクトルの平均ベクトルと対話外名詞を潜在的興味対象推定のための NN(図の③) に入力して興味推定を行う．

以下で，提案モデルについて詳しく説明する．

2.2 対話ログ中の単語の興味推定

対話ログに現れていない潜在的興味対象を推定するためには，対話ログ中の話題に対する話者の興味が手がかりとなると考えられる．そこで，モデルはまず対話ログ中の全名詞に対して話者が興味を持っているか否かを推定する．話者がある単語に興味があるか否かは，その単語の前後の文脈が重要であると考えられる．

そこでまず，興味推定対象の名詞を含む発話を形態素解析し，

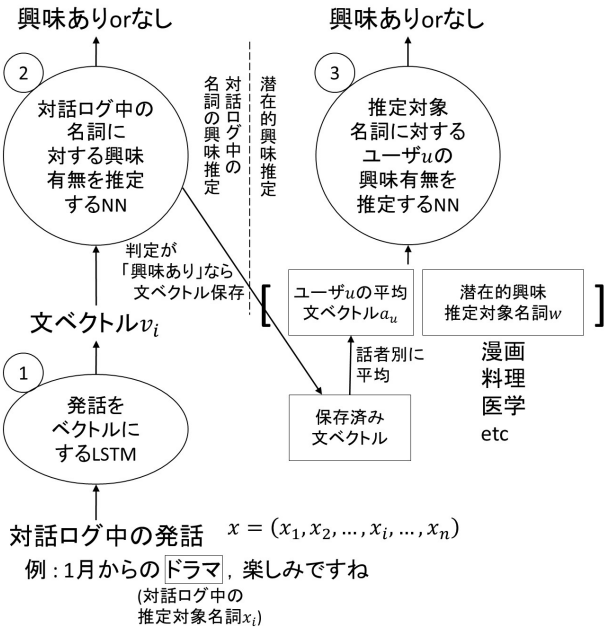

図 1: 潜在的興味推定モデル

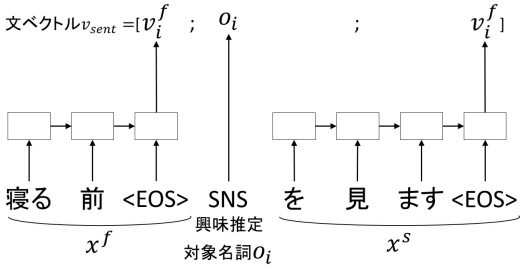

図 2: RNN による文章のベクトル化

各単語を Word2Vec[Mikolov 13] により d 次元のベクトルに変換することで，単語ベクトル系列 $x=(x_1,x_2,...,x_t,...,x_k)$ を得る．ここで，興味推定対象の名詞を x_t とする．形態素解析には Mecab[*1] を使用した．興味推定に有用なのは名詞 x_t の周辺の単語であると考えられることから，単語ベクトル x_t の前後 l 個の単語ベクトルを文脈として用いる．すなわち，提案モデルは，x における推定対象名詞より前の単語ベクトル系列 $x^f=(x_{t-l},x_{t-(l+1)},...,x_{t-1})$ と後の単語ベクトル系列 $x^s=(x_{t+1},x_{t+2},...,x_{t+l})$ をそれぞれ RNN により文脈ベクトルに変換する．ここで，RNN への入力時には，各系列の終端を意味する <EOS> を挿入する．この <EOS> もベクトルであり，その要素は 1 次元目から d 次元目までが 0，$d+1$ 次元目が 1 のベクトルである．各系列は逐次 RNN に入力していき，<EOS> ベクトルを入力した時点の RNN の出力を文脈ベクトルとする．

この x^f と x^s をそれぞれ RNN によって文ベクトル化する (図 1 の①)．例として,「寝る前 SNS を見ます」というユーザ発話から文ベクトルを作成する例を図 2 に示す．この発話における興味推定対象名詞は「SNS」である．なお，図中の単語は日本語で表記しているが，実際には word2vec で変換済みのベクトルである．

*1 http://taku910.github.io/mecab

モデルは最初に単語ベクトル系列 x^f(「寝る」,「前」,「<EOS>」) を文頭から 1 単語ずつ RNN に入力していき，<EOS> が入力された時点での RNN の出力を前文脈ベクトル v^f とする．同様に，単語ベクトル系列 x^s(「を」,「見」,「ます」,「<EOS>」) から後文脈ベクトル v_i^s を得る．推定対象名詞 (「SNS」) は RNN に入力せず，単語ベクトルをそのまま x_t として利用する．

次に v^f，x_t，v^s を連結した文ベクトル $v_{sent}=[v_f;x_t;v^s]$ を用いて，名詞 x_t に対する興味有無の推定を行う (セミコロンはベクトルの連結を表す)．そこで，v_{sent} を全結合ニューラルネットワークに入力する (図 1 の②)．この全結合ニューラルネットワークの出力 2 次元の確率分布であり，それぞれ名詞 x_t に興味がある確率と興味がない確率を意味する．この確率分布をもとに，対象名詞に関するユーザの興味有無を判定する．このとき「興味あり」と判定された v_{sent} は，主目的である潜在的興味推定のための手掛かりと見なせる．そこで，推定結果が「興味あり」であった場合，v_{sent} を保存し，次節で述べる潜在的興味推定に用いる．

2.3 平均文ベクトルを用いた潜在的興味推定

対話ログに出現していない名詞に対してそれぞれの話者が興味を持っているか否かを推定するため，前節で述べた方法により保存した v_{sent} を話者別に平均する．ここで，ユーザ $u_j(j=0,1)$ に関する平均文ベクトルを a_{u_j} とする．

対話ログに出現していない潜在的興味対象の名詞 $w_{u_j}=(w_1^{u_j},w_2^{u_j},...,w_n^{u_j})$ に対してそれぞれの話者が興味を持っているか否かを推定するため，a_{u_j} と興味推定したい名詞のベクトル $w_i^{u_j}$ を連結した $[a_{u_j};w_i^{u_j}]$ を全結合ニューラルネットワークに入力する (図 1 の③)．このニューラルネットワークの出力は対話ログ中の名詞の興味推定時と同様，興味の有無に関する 2 次元の確率分布を出力する．こうして得た確率をもとに，名詞 $w_i^{u_j}$ に対する潜在的興味有無を推定する．

2.4 モデルの学習

本モデルでは 3 つのニューラルネットワークを内包しており，各ニューラルネットワークは互いに独立ではなく，発話を文ベクトルに変換する RNN，および対話ログ中の名詞に対する興味推定を行うニューラルネットワークの出力に応じて，潜在的興味推定を行うニューラルネットワークの入力が変化する関係となっている．このモデルを効率良く学習させるため，対話ログ中の名詞に対する興味推定における正解との誤差と潜在的興味推定における正解との誤差の和を損失関数とする．提案モデルは損失関数を用いて全てのニューラルネットワークを同時に最適化することで，対話ログからの適切な文ベクトルの抽出，および潜在的興味対象の推定が実現できる．

3. 評価実験

提案手法を評価するため，2 人の話者による対話データを用いて性能評価実験を行った．対話データ中の全ての名詞には興味の有無を意味するラベルを付与した．実験では，対話データを 1 対話に含まれる発話数で 2 等分し，前半部をモデルに与える対話ログ，後半部の名詞を潜在的興味の推定対象とした (後半部の対話データはモデルには与えない)．潜在的興味推定対象の名詞は，後半部の対話データにおいて，どの話者の発話に含まれていたかという観点から話者別に与えられ，その名詞を含む発話を行った話者が興味を持つか否かを推定することでモデルを評価する．

表 1: 使用データ

対話数	100
1 対話あたりの発話数	53.87
1 発話あたりの単語数	40.06
対話前半部の名詞数	6279
対話前半部の興味あり名詞数	1530
対話後半部の名詞数	6257
対話後半部の興味あり名詞数	1084
Fleiss' Kappa	0.50

3.1 使用データ

対話データはクラウドソーシングサービスの Crowdworks[*2] で被験者を募集し，1 対話 1 時間のデータを 100 件収集した．収集には Skype インスタントメッセンジャーを用いた．

学習および評価のため，収集した対話データに含まれる全ての名詞に対して話者が興味を持っているか否かのアノテーションを付与した．名詞の抽出は形態素解析器により自動で行った．

アノテータは対話データと対話中で使用された名詞が与えられ，各名詞に対し,「興味あり」もしくは「判断不能 or 興味なし」の 2 種類のラベルのいずれかを付与した．付与基準は「アノテーション対象の名詞を含む発話をした話者に「あなたは「○○ (当該名詞が入る)」に興味がありますか？」もしくは「あなたは「○○ (当該名詞が入る)」に関する話題に興味がありますか？」という質問をした際の，その話者の予想回答」とした．名詞の抽出は形態素解析器により自動で行ったため，抽出ミスも含まれている．そのため，質問の「○○」に名詞を入れた際，日本語として意味が通らない場合は「判断不能・興味なし」のラベルを付与することとした．

以上の基準により，Crowdworks で募集したアノテータ 10 名が個別にアノテーションを行い，多数決により正解ラベルを決定した.「興味あり」と「判断不能 or 興味なし」がそれぞれ同数 (5 名ずつ) だった場合は「判断不能 or 興味なし」とした．

使用したデータの統計情報を表 1 に示す．アノテータ間のアノテーションの一致度を示す Fleiss' Kappa は 0.50 であり，中程度の一致を示した．

3.2 実験設定

Word2Vec はウインドウサイズは 5，最小出現頻度は 10，ベクトルの次元数は 1000 とし，約 100GB の Twitter データを用いて SkipGram で学習を行った．

文脈ベクトルへのエンコードには，興味推定対象名詞の前後 5 個ずつを用いた ($l = 5$)．文脈ベクトルへのエンコードを行う RNN の入力層は 1001 次元，対話中の名詞の興味推定のためのネットワークの入力層は 3000 次元，潜在的興味推定のためのネットワークの入力層は 4000 次元，その他の中間層の次元は全て 1000 次元とした．また，全ての中間層に dropout を適用し，dropout 率は 30%とした．学習エポック数は 30 とし，10 分割交差検定で「興味あり」判定の精度，再現率，F 値により評価した．

3.3 比較手法

提案手法の評価のため，2.4 節で述べた 2 つのニューラルネットワークを同時に更新する unite モデルと，潜在的興味推定のためのニューラルネットワークを分離した separate モデル，およびニューラルネットワークの層の数を全て 1 層とした shallow モデルと全て 8 層の ResNet[He 16] とした deep モデルを用意した．separate モデルは，文脈ベクトルへのエンコードのための RNN，および対話データ中の名詞の興味推定のためのネットワークと，潜在的興味推定のためのネットワークを分離し，損失関数も別とした．また，ネットワークを分離するため，潜在的興味推定のために RNN の出力 (文脈ベクトル) を使用せず，平均文ベクトルの代わりに平均単語ベクトルを用いる．平均する単語ベクトルの選択には，unite モデルと同じく，対話データ中の名詞の興味推定のためのネットワークの結果を用いる．

実験では，上記のそれぞれを組み合せた separate shallow, separate deep, unite shallow, unite deep の 4 つの設定で実験を行った．またベースライン手法として，以下の 2 つの手法による実験を行った．

3.3.1 ベースライン 1：単語類似度

ベースラインとして，対話ログ中の名詞との類似度により，潜在的興味の推定を行う．本手法は，対話ログに含まれる全名詞を Word2Vec により単語ベクトルに変換し，その単語ベクトルの平均と，潜在的興味推定対象名詞のベクトルとのコサイン類似度により「興味あり」か「判断不能 or 興味なし」かを判定する．類似度のしきい値は 0.0 から 1.0 まで 0.1 刻みで変更させて評価を行い，F 値が最大となった結果をこの手法の結果とする．

3.3.2 ベースライン 2：SVM

もう一つのベースラインとして，Support Vector Machine(SVM) による推定を行う本手法では，比較手法 1 の単語類似度で用いた平均ベクトルと，潜在的興味推定対象名詞のベクトルを連結したベクトルを入力ベクトルとし，潜在的興味の有無を SVM により 2 値分類する．

表 2: 実験結果

手法	精度	再現率	F 値
単語類似度	0.294	0.393	0.336
SVM	0.479	0.797	0.598
separate shallow	0.517	0.732	0.606
separate deep	0.523	0.709	0.602
unite shallow	0.539	0.716	0.615
unite deep	0.546	0.712	0.618

3.4 実験結果

実験の結果を表 2 に示す．表より，unite deep の F 値が最も高く，続いて unite shallow，separate shallow, separate deep の順となり，提案手法のすべての設定においてベースライン手法の単語類似度と SVM よりも優れた性能を示した．

単語類似度の手法では，対話ログに含まれる全名詞を用いたため，平均単語ベクトルの計算に興味の有無とは無関係の単語も多く使用されることになったため，低い性能に留まったと考えられる．

提案手法が SVM よりも優れていた理由としては，ニューラルネットベースの 4 つのモデルが RNN を用いて発話の文脈情報を利用しているのに対し，SVM モデルが発話中の名詞のみに注目して推定していることが考えられる．

ニューラルネットベースのモデル同士で比較すると，separate shallow モデルと saparate deep モデルに対して，unite shallow モデルおよび unite deep モデルが優れた推定結果を示している．このことから，2 つの累積誤差を統合してモデル全体を学習させるという提案手法が有効であったことが分かる．simple

*2 https://crowdworks.jp/

表 3: 対話ログ前半の一部

ユーザ 1	最近漬物にハマっています。
ユーザ 2	漬物ですか。漬物なら、 私はたくあんが好きです。
ユーザ 1	いいですね。お酒にもよく合います。
ユーザ 2	お酒飲まれるんですね。何が好きですか？
ユーザ 1	ウイスキーが一番好きです。
ユーザ 2	私もたまにウィスキー飲みますよ。 よく飲むのが赤ワインを少々です。
ユーザ 1	ワインなら白より赤ですか？

表 4: 潜在的興味推定結果 (unite deep)

ユーザ 1			ユーザ 2		
	正解	推定		正解	推定
チーズ	○	○	チーズ	○	○
燻製	○	○	生ハム	○	○
楽器	○	○	楽器	×	×
ギター	○	○	ギター	×	×
アコーディオン	×	○	鉄琴	×	×
用事	×	×	ワイン	○	○
経験	×	×	大人	×	×
ヘビーメタル	○	○	実感	×	×

表 5: 潜在的興味推定結果 (SVM)

ユーザ 1			ユーザ 2		
	正解	推定		正解	推定
チーズ	○	○	チーズ	○	○
燻製	○	○	生ハム	○	○
楽器	○	○	楽器	×	○
ギター	○	○	ギター	×	○
アコーディオン	×	○	鉄琴	×	○
用事	×	×	ワイン	○	○
経験	×	×	大人	×	×
ヘビーメタル	○	○	実感	×	×

と deep の違いに注目すると，separate モデルは simple の方が deep より F 値が大きくなっているが，一方で unite モデルでは shallow より deep の方がわずかに良い結果になっている．特に，unite deep モデルは全ての手法の中でも最良の結果となっている．このことから，多くの層を持つニューラルネットと unite モデルの組み合わせが有効であったことが示唆された．

3.5 考察

表 3 は実験で用いた 1 対話の前半から一部を抜き出したものである．この対話中の後半から抽出した潜在的興味の推定対象名詞に対して，unite モデルで潜在的興味の推定を行った結果の一部を表 4 に，SVM で推定を行った結果の一部を 5 に示す．表では「興味あり」を○,「不明・興味なし」を×で表す．

対話例より，ユーザ 1 とユーザ 2 はともにお酒に興味を持っていることがわかる．ここから，酒のおつまみに対して一定の興味があると考えられるが，提案モデル，SVM ともに「チーズ」,「燻製」,「生ハム」に対して正しく興味推定ができていることがわかる．また，提案モデルではユーザ 1 は「楽器」や「ギター」など音楽に関するものに興味があり，ユーザ 2 はあまり興味が無いことも正しく推定できている．一方，SVM では話者別の推定に失敗しており，ユーザ 1 とユーザ 2 で共通している単語「チーズ」,「楽器」,「ギター」は両者とも全て興味ありと推定されていることがわかる．以上の結果から，提案モデルは話者別に正しく潜在的興味が推定できていることが確認できた．

4. まとめ

本研究では，ユーザが興味をもっているものの直接会話には現れていない潜在的興味対象を推定することを目的に，ニューラルネットベースの手法を提案した．提案手法では，まず対話ログ中の発話を RNN を用いて文ベクトル化し，対話ログ中の名詞に対してユーザが興味を持っているか否かを推定する．その結果を平均して対話外名詞とともにニューラルネットに入力することで，対話に登場しない名詞に対してもユーザの興味の有無を推定した．また学習においては，対話ログ中の名詞に対して興味推定したときの累積誤差と，対話外名詞に対して潜在的興味を推定したときに累積された誤差を足し合わせて統合してモデル全体の重みを同時に更新する手法を用いた．実験によって，提案手法による推定性能が，対話ログ中の単語のみに注目した SVM によるモデルを上回ることが示された．

今回の研究では，潜在的興味の推定対象となる名詞は与えられており，これら対象名詞に対してユーザが興味を持っているか否かを推定した．今後の課題として，こういった対話外名詞が与えられていない場合で潜在的興味対象を提案することなどが挙げられる．

参考文献

[Bang 15] Bang, J., Noh, H., Kim, Y., and Lee, G. G.: Example-based Chat-oriented Dialogue System (2015)

[He 16] He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)

[Mikolov 13] Mikolov, T., Chen, K., Corrado, G., and Dean, J.: Efficient estimation of word representations in vector space, *Computation and Language* (2013)

[Schuller 06] Schuller, B. W., Köhler, N., Müller, R., and Rigoll, G.: Recognition of interest in human conversational speech., in *INTERSPEECH* (2006)

[Wang 13] Wang, W. Y., Biadsy, F., Rosenberg, A., and Hirschberg, J.: Automatic detection of speaker state: Lexical, prosodic, and phonetic approaches to level-of-interest and intoxication classification, *Computer Speech & Language*, Vol. 27, No. 1, pp. 168–189 (2013)

[大坊 06] 大坊郁夫: コミュニケーション・スキルの重要性, 日本労働研究雑誌, Vol. 48, No. 1, pp. 13–22 (2006)

[平野 16] 平野 徹, 小林 のぞみ, 東中 竜一郎, 牧野 俊朗, 松尾 義博: パーソナライズ可能な対話システムのためのユーザ情報抽出, 人工知能学会論文誌, Vol. 31, No. 1, pp. 1–10 (2016)