

複数言語複数タスクを扱う発話意図推定モデリング のための敵対的学習の検討

増村 亮^{1,a)} 篠原 雄介¹ 東中 竜一郎¹ 青野 裕司¹

概要：本研究では、音声対話システムを構築する上で必要となる発話意図推定モジュールについて、複数の言語のデータ、および、複数の意図推定タスクのデータを同時に用いてモデル化することで、各言語、各タスクでの性能向上を狙う。複数のデータを用いて同時にモデル化する場合、各言語、各タスクで集められるデータ量がそれぞれ異なるため、しばしばデータ量が多いタスクや言語にモデル全体が依存してしまうことが課題である。この課題を解決するために、本研究では特定の言語に依存しない結合モデリング、および、特定のタスクに依存しない結合モデリングを同時に実現するための敵対的学習手法を提案する。提案法では、異なる言語間で共有可能なタスク依存ネットワークと異なるタスク間で共有可能な言語依存ネットワークから構成される結合ネットワークに対して、言語種別についての敵対的ネットワークとタスク種別についての敵対的ネットワークを導入する。これにより、言語やタスクの異なりに変な特徴抽出と、発話意図推定の性能向上を両立させるモデル化が可能となる。評価実験では、日本語と英語の2言語について、対話行為推定、話題種別推定、質問種別推定の3つの意図推定タスクのデータを用いて、提案手法の有効性を示す。

キーワード：発話意図推定、ニューラルネットワーク、結合モデル、敵対的学習

1. はじめに

音声対話システムでは、入力されたユーザの発話の意味を精緻に理解するために、発話意図推定が広く用いられている。音声対話システムを対象とする場合、対話行為 [1], [2], ドメイン [3], 質問種別 [4] 等の様々な意図推定タスクが検討されており、複数の発話意図推定を同時に用いて音声対話システムを構成することも一般的である。発話意図推定における課題の1つとして、意図推定タスクごとに、また、扱う言語ごとに大量のデータを準備しなければいけないことが挙げられる。この課題に対して、本研究では複数の言語のデータセットや複数のタスクのデータセットを効率的に利用することで、各言語、各タスクにおける性能を改善させる手法の確立を目指す。

機械学習分野において、ネットワーク共有に基づく結合モデリングが近年広く注目されている。自然言語処理分野における結合モデリングは、複数の言語のデータセットや、複数のタスクのデータセットを同時に用いてモデル化することで実現でき、異なる言語間で共有可能な知識や、異なる

るタスク間で共有可能な知識を効率的に獲得することが可能である。この結合モデリングを用いることで、個々の言語や個々のタスクにおける性能を改善可能であることが報告されている [5]。

そこで本研究では、複数言語、複数タスクのデータセットを同時に用いた発話意図推定の結合モデリングを試みる。発話意図推定のモデル化には、ニューラルネットワークに基づく手法 [6], [7], [8] が数多く検討されており、結合モデリングとも相性が良いことが期待される。複数言語、複数タスクでの結合モデリングは、異なるタスク間で共有可能な言語依存ネットワークと、異なる言語間で共有可能なタスク依存ネットワークを組み合わせることで実現できる [9], [10]。しかしながら、このような複雑な結合モデリングでは、各言語、各タスクで集められるデータ量がそれぞれ異なるため、学習データが比較的少ない言語やタスクに対しては、個別でモデル化する場合よりもしばしば性能劣化してしてしまう問題が知られている。この問題は、異なるタスク間で共有可能な言語依存ネットワークがデータ量の多いタスクに依存してしまうこと、および、異なる言語間で共有可能なタスク依存ネットワークがデータ量の多い言語に依存してしまうことに起因する。

これらの問題を解決するために、本稿では複数言語複数

¹ 日本電信電話株式会社 NTT メディアインテリジェンス研究所
NTT Corporation, NTT Media Intelligence Laboratories, 1-1 Hikari-nooka Yokosuka-shi, Kanagawa 239-0847, Japan

^{a)} masumura.ryo@lab.ntt.co.jp

タスクを扱う結合モデリングのための敵対的学習手法を提案する。我々のアイデアは、ドメイン敵対的学習 [11] に着想を得たものであり、言語依存ネットワークが特定のタスクに特化しないようにするための学習基準、および、タスク依存ネットワークが特定の言語に特化しないようにするための学習基準を結合モデルに導入することである。これにより、言語やタスクの異なりに不変な特徴抽出と、発話意図推定の性能向上を両立させるモデル化が期待できる。

本稿では、長短期記憶に基づくリカレントニューラルネットワーク (LSTM-RNN) と注意機構 [12], [13] を組み合わせた発話意図推定に対して、複数言語複数タスクを扱う結合モデリング、およびその敵対的学習手法について詳細を述べる。日本語と英語の2言語、および対話行為、話題種別、質問種別の3種類の意図推定タスクを対象とした評価実験において、敵対的学習を導入した結合モデリングにより、各言語、各タスクにおける発話意図推定性能が改善できることを示す。

2. 関連研究

結合モデリング: 自然言語処理分野において、結合モデリングは、複数タスクを同時に扱うモデリングと、複数言語を同時に扱うモデリングの2種類に大きく分けられる。複数タスクを同時に扱うモデリングでは、同一言語の異なるタスク間でネットワークを共有することにより、各タスクの性能を改善させることを目的とする [5], [14], [15], [16], [17]。一方、複数言語を同時に扱うモデリングでは、異なる言語間で意味が同一の言語表現に対して同様のベクトルが割り当てられるようにモデル化を行うことで、同一のタスクを同一の識別器で扱うことを可能とする [18], [19], [20], [21]。さらに、両者の結合モデリングを同時に扱う手法も近年検討が始まっており、複数の言語のデータセットと複数のタスクデータセットを同時に用いることで、各言語、各タスクの性能改善を目的としている [9], [10]。これらの結合モデリングでは、複数のデータセット全体に対しての最尤基準の学習が通常導入されている。これに対して本研究では、複数言語複数タスクを扱う結合モデリングのために敵対的学習を導入する。

敵対的学習: 敵対的学習は、敵対的生成ネットワーク [22] の登場以来、機械学習の様々な技術領域で導入が試みられている。敵対的学習は、テキスト分類において様々な検討がなされている [11], [23], [24], [25], [26]。自然言語処理分野においては、言語の違いに不変となるような敵対的学習 [23], [27]、および、タスクの違いに不変となるような敵対的学習 [11], [24], [26] の2種類に分類できる。本研究は、この両者の敵対的学習を両立させるという立ち位置であり、複数の入力種別、複数の出力種別を想定したネットワークのための敵対的学習は、我々の知る限り本研究が初出である。

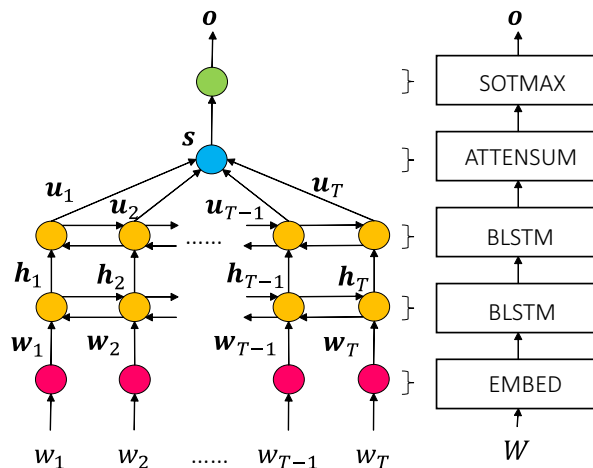


図1 2層 BLSTM-RNN と注意既往を用いたネットワーク構造。

3. 発話意図推定

本節では、結合モデリングを行わない場合の発話意図推定、つまり、単言語単タスクを扱う発話意図推定について、ニューラルネットワークに基づくモデル化手法と学習方法を説明する。

発話意図推定は、入力発話の単語系列 $W = \{w_1, \dots, w_T\}$ から、ある意図推定タスクのラベル $l \in \{l_1, \dots, l_K\}$ を決定する問題である。ここで意図推定タスクとは、対話行為、話題種別、質問種別などを表し、それぞれ異なるラベルセットを持つ。

ニューラル発話意図推定では、入力発話の単語系列が与えられた際のラベルに対する条件付き確率 $P(l|W, \Theta)$ を直接ニューラルネットワークを用いてモデル化する。ここで、 Θ はモデルパラメータを表す。ニューラル発話意図推定のモデル化には、様々なネットワーク構造が利用できるが、本稿では図1に示すような2層の双方向 LSTM-RNN (BLSTM-RNN) と注意機構を用いたネットワーク構造を採用する。

3.1 ネットワーク構造

ニューラル発話意図推定では、最初に入力発話 W の各単語を連続ベクトルに変換する。 w_t の連続ベクトル w_t は (1) 式で表される。

$$w_t = \text{EMBED}(w_t; \theta_w) \quad (1)$$

ここで、 $\text{EMBED}()$ は単語を連続ベクトルに変換するための線形変換関数、 θ_w はモデルパラメータを表す。次に、2層の BLSTM-RNN を用いて各単語を前後のコンテキストを考慮した連続ベクトルに変換する。 t 番目の単語の連続ベクトルに対する変換は、(2)-(3) 式に従う。

$$h_t = \text{BLSTM}(\{w_1, \dots, w_T\}, t; \theta_h) \quad (2)$$

$$\mathbf{u}_t = \text{BLSTM}(\{\mathbf{h}_1, \dots, \mathbf{h}_T\}, t; \boldsymbol{\theta}_u), \quad (3)$$

ここで、 $\text{BLSTM}()$ は BLSTM-RNN の機能を持つ関数であり、 $\boldsymbol{\theta}_h$, $\boldsymbol{\theta}_u$ はモデルパラメータを表す。さらに、これらの連続ベクトルを用いて、発話単位の連続ベクトルを構成する。その際、注意機構により個々のベクトルの重要度を考慮した埋め込みを行う。発話単位の連続ベクトル \mathbf{s} は (4) 式に従い構成される。

$$\mathbf{s} = \text{ATTENSUM}(\{\mathbf{u}_1, \dots, \mathbf{u}_T\}; \boldsymbol{\theta}_z) \quad (4)$$

ここで、 $\text{ATTENSUM}()$ は注意機構に基づき入力ベクトル系列の重み付き和をとるための関数であり、 $\boldsymbol{\theta}_z$ はそのモデルパラメータを表す。最後に発話単位の連続ベクトルから (5) 式に従い予測確率分布 $\mathbf{o} \in \mathbb{R}^K$ を算出する。

$$\mathbf{o} = \text{SOFTMAX}(\mathbf{s}; \boldsymbol{\theta}_o) \quad (5)$$

ここで、 $\text{SOFTMAX}()$ は線形変換後にソフトマックス関数を適用するものであり、 $\boldsymbol{\theta}_o$ はそのモデルパラメータを表す。 \mathcal{O} における k 次元目の値は、 $P(l_k|W, \Theta)$ に対応する。なお、 Θ は $\{\boldsymbol{\theta}_w, \boldsymbol{\theta}_h, \boldsymbol{\theta}_u, \boldsymbol{\theta}_z, \boldsymbol{\theta}_o\}$ に一致する。

3.2 学習

ニューラルネットワークに基づく発話意図推定のモデルパラメータ Θ は、発話とラベルのペアの集合であるデータセット \mathcal{D} から学習できる。学習では、正解の確率分布と予測確率分布の交差エントロピーに基づく損失関数を (6) 式として定義する。

$$L = - \sum_{n=1}^{|\mathcal{D}|} \sum_{k=1}^K \hat{\mathbf{o}}_{n,k} \log \mathbf{o}_{n,k} \quad (6)$$

ここで、 $\hat{\mathbf{o}}_{n,k}$, $\mathbf{o}_{n,k}$ は、学習データにおける n 番目の発話についての k 番目のラベルの正解確率と予測確率を表す。このモデルパラメータは、確率的勾配法に基づき (7) 式に従い更新される。

$$\Theta \leftarrow \Theta - \epsilon \frac{\partial L}{\partial \Theta} \quad (7)$$

ここで、 ϵ は学習率を表す。

4. 提案手法

本節では、ニューラルネットワークに基づく発話意図推定について、複数言語複数タスクを扱う結合モデリング、およびその敵対的学習手法について説明する。

ここでは、 $i \in \{1, \dots, I\}$ 番目の言語の入力発話を $W^{(i)} = \{w_1^{(i)}, \dots, w_T^{(i)}\}$, $j \in \{1, \dots, J\}$ 番目の意図推定タスクの意図ラベルを $l^{(j)} \in \{1, \dots, K^{(j)}\}$ と表す。なお、 I は結合モデリングで扱う言語の種類数、 J は結合モデリングで扱う意図推定タスクの種類数を表す。

この条件下において、複数言語複数タスクを扱う結合

モデリングでは、 i と j の組み合わせごとに、条件付き確率 $P(l^{(j)}|W^{(i)}, \Theta^{(i,j)})$ をそれぞれモデル化する。ここで、 $\Theta^{(i,j)}$ は、 i 番目の言語、 j 番目の意図推定タスクを扱う際のモデルパラメータを表し、 I 種類の言語、 J 種類の意図推定タスク内で部分的に共有化されている。複数言語複数タスクを扱う結合モデリングでは、 $\{\Theta^{(1,1)}, \dots, \Theta^{(I,J)}\}$ を、 I 種類の言語ごとの J 種類のタスクのデータを用いて同時に学習する。

敵対的学習を行う際の結合モデリングは、結合ネットワークと敵対的ネットワークを組み合わせにより構成される。以下ではそれぞれの詳細を述べる。

4.1 結合ネットワーク

結合ネットワークは、 I 個の言語依存ネットワーク、 J 個のタスク依存ネットワーク、および J 個の意図推定ネットワークにより構成される。ここでは3節で説明したニューラル発話意図推定を結合ネットワークとして扱う場合について説明する。

言語依存ネットワークは、特定の言語についての異なるタスク間で共有可能なネットワークであり、ある言語の単語系列を前後の文脈を考慮した連続ベクトル系列に変換する役割を担う。 i 番目の言語に依存した変換は、(8)-(9) 式に従う。

$$\mathbf{w}_t^{(i)} = \text{EMBED}(w_t^{(i)}; \boldsymbol{\theta}_w^{(i)}) \quad (8)$$

$$\mathbf{h}_t^{(i)} = \text{BLSTM}(\{\mathbf{w}_1^{(i)}, \dots, \mathbf{w}_T^{(i)}\}, t; \boldsymbol{\theta}_h^{(i)}) \quad (9)$$

ここで、 $\boldsymbol{\theta}_w^{(i)}$, $\boldsymbol{\theta}_h^{(i)}$ は、 i 番目の言語についての言語依存ネットワークのモデルパラメータを表す。

一方、タスク依存ネットワークは、特定の意図推定タスクについての複数の言語間で共有可能なネットワークであり、言語依存ネットワークの出力である連続ベクトル系列から対象の意図推定タスク向けの連続ベクトル系列に変換する役割を担う。 j 番目のタスクに依存した変換は、(10) 式に従う。

$$\mathbf{u}_t^{(j)} = \text{BLSTM}(\{\mathbf{h}_1^{(i)}, \dots, \mathbf{h}_T^{(i)}\}, t; \boldsymbol{\theta}_u^{(j)}) \quad (10)$$

ここで、 $\boldsymbol{\theta}_u^{(j)}$ は、 j 番目の意図推定タスクについてのタスク依存ネットワークのモデルパラメータを表す。

最後に意図推定ネットワークでは、タスク依存の連続ベクトル系列から意図推定タスクのラベルについての事後確率を出力する。 j 番目の意図推定タスクについての事後確率 $\mathbf{o}^{(j)} \in \mathbb{R}^{K^{(j)}}$ への変換は、(11)-(12) 式に従う。

$$\mathbf{s}^{(j)} = \text{ATTENSUM}(\{\mathbf{u}_1^{(i)}, \dots, \mathbf{u}_T^{(i)}\}; \boldsymbol{\theta}_z^{(j)}) \quad (11)$$

$$\mathbf{o}^{(j)} = \text{SOFTMAX}(\mathbf{s}^{(j)}; \boldsymbol{\theta}_o^{(j)}) \quad (12)$$

ここで、 $\boldsymbol{\theta}_z^{(j)}$, $\boldsymbol{\theta}_o^{(j)}$ は j 番目の意図推定ネットワークのモデルパラメータを表す。

以上をまとめると、結合ネットワークにおける i 番目の

言語, j 番目のタスクを扱うコンポーネントのモデルパラメータ $\Theta^{(i,j)}$ は, $\{\theta_h^{(i)}, \theta_h^{(j)}, \theta_u^{(j)}, \theta_z^{(j)}, \theta_o^{(j)}\}$ として表される。

4.2 敵対的ネットワーク

提案手法では, タスク種別についての敵対的ネットワークと言語種別についての敵対的ネットワークを組み合わせることにより, 結合モデリングの敵対的学習を実現する。タスク種別についての敵対的ネットワークは, タスク種別に関する情報を浄化するために導入される。一方, 言語種別についての敵対的ネットワークは, 言語種別に関する情報を浄化するために導入される。

前述の結合ネットワークと敵対的ネットワークの両者を確率的勾配法に基づく学習で同時に最適化するために, Gradient reversal layer (GRL) [11] を用いる。GRL はフォワードプロパゲーション時は恒等変換関数として動作し, バックプロパゲーション時は勾配の符号を反転する関数として動作するものであり, 敵対的学習を効率的に実現することができる。

タスク種別についての敵対的ネットワークは, 言語依存ネットワークごとに準備される。 i 番目の言語依存ネットワークに対応するタスク種別についての敵対的ネットワークでは, タスク種別についての事後確率 $\mathbf{x}^{(i)} \in \mathbb{R}^J$ を (13)-(15) 式に従い算出する。

$$\tilde{\mathbf{h}}_i^{(i)} = \text{GRL}(\mathbf{h}_i^{(i)}) \quad (13)$$

$$\tilde{\mathbf{h}}^{(i)} = \text{ATTENSUM}(\{\tilde{\mathbf{h}}_1^{(i)}, \dots, \tilde{\mathbf{h}}_T^{(i)}\}; \theta_x^{(i)}) \quad (14)$$

$$\mathbf{x}^{(i)} = \text{SOFTMAX}(\tilde{\mathbf{h}}^{(i)}, \theta_x^{(i)}) \quad (15)$$

ここで, $\text{GRL}()$ は GRL を表し, $\theta_x^{(i)}$ は i 番目の言語依存ネットワークに対応するタスク種別についての敵対的ネットワークのモデルパラメータを表す。

同様に, 言語種別についての敵対的ネットワークは, タスク依存ネットワークごとに準備される。 j 番目のタスク依存ネットワークに対応する言語種別についての敵対的ネットワークでは, 言語種別についての事後確率 $\mathbf{y}^{(j)} \in \mathbb{R}^I$ を (16)-(18) 式に従い算出する。

$$\tilde{\mathbf{u}}_i^{(j)} = \text{GRL}(\mathbf{u}_i^{(j)}) \quad (16)$$

$$\tilde{\mathbf{u}}^{(j)} = \text{ATTENSUM}(\{\tilde{\mathbf{u}}_1^{(j)}, \dots, \tilde{\mathbf{u}}_T^{(j)}\}; \theta_y^{(j)}) \quad (17)$$

$$\mathbf{y}^{(j)} = \text{SOFTMAX}(\tilde{\mathbf{u}}^{(j)}, \theta_y^{(j)}) \quad (18)$$

ここで, $\theta_y^{(j)}$ は j 番目のタスク依存ネットワークに対応する言語種別についての敵対的ネットワークのモデルパラメータを表す。

2 言語 2 タスクで全体をモデル化する場合の結合ネットワークと敵対的ネットワークの関係を図 2 に示す。

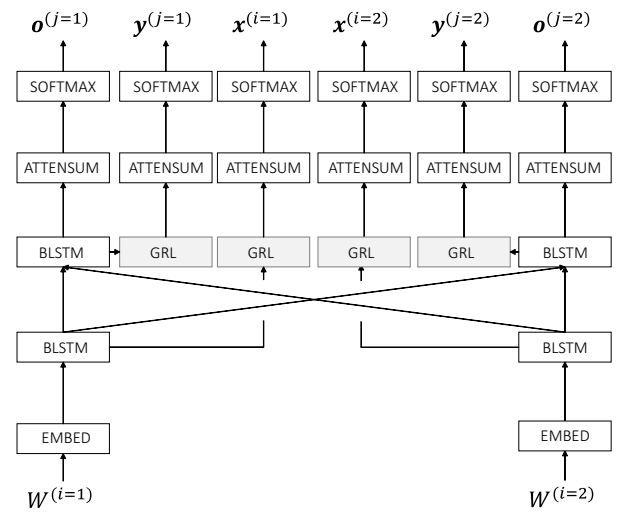


図 2 2 言語 2 タスクを扱う場合の結合ネットワークと敵対的ネットワーク。

4.3 敵対的学習

複数言語複数タスクを扱う結合モデリングのための敵対的学習について述べる。ここで, i 番目の言語の j 番目のタスクの学習データ (単語系列とラベルの組の集合) を $\mathcal{D}^{(i,j)}$ として表す場合, 全体の学習データは $\{\mathcal{D}^{(1,1)}, \dots, \mathcal{D}^{(I,J)}\}$ として表される。この時, 3 つの交差エントロピーに基づく損失関数を定義できる。

1 つ目は意図推定ネットワークにおける損失関数であり, (19) 式で表される。

$$L_o = - \sum_{i=1}^I \sum_{j=1}^J \sum_{n=1}^{|\mathcal{D}^{(i,j)}|} \sum_{k=1}^{K^{(j)}} \hat{o}_{n,k}^{(j)} \log o_{n,k}^{(j)} \quad (19)$$

ここで, $\hat{o}_{n,k}^{(j)}$, $o_{n,k}^{(j)}$ は, 学習データ $\mathcal{D}^{(i,j)}$ における n 番目のデータについての j 番目の意図タスクの識別器ネットワークにおける k 番目ラベルの正解確率と予測確率を表す。

2 つ目はタスク種別についての敵対的ネットワークの損失関数であり, (20) 式で表される。

$$L_x = - \sum_{i=1}^I \sum_{j=1}^J \sum_{n=1}^{|\mathcal{D}^{(i,j)}|} \sum_{j'=1}^J \hat{x}_{n,j'}^{(i)} \log x_{n,j'}^{(i)} \quad (20)$$

ここで, $\hat{x}_{n,j'}^{(i)}$, $x_{n,j'}^{(i)}$ は, 学習データ $\mathcal{D}^{(i,j)}$ における n 番目のデータについての i 番目の言語依存ネットワークに対応するタスク種別についての敵対的ネットワークにおける j' 番目ラベルの正解確率と予測確率を表す。

3 つ目は言語種別についての敵対的ネットワークの損失関数であり, (21) 式で表される。

$$L_y = - \sum_{i=1}^I \sum_{j=1}^J \sum_{n=1}^{|\mathcal{D}^{(i,j)}|} \sum_{i'=1}^I \hat{y}_{n,i'}^{(j)} \log y_{n,i'}^{(j)} \quad (21)$$

ここで, $\hat{y}_{n,i'}^{(j)}$, $y_{n,i'}^{(j)}$ は, 学習データ $\mathcal{D}^{(i,j)}$ における n 番目のデータについての j 番目のタスク依存ネットワークに

表2 各意図推定タスクごとの入力発話例と意図ラベル例.

意図推定タスク	入力発話	意図ラベル
対話行為	Hello, how are you today?	GREETING
	I am so sorry to hear of your son's accident.	SYMPATHY/AGREE
	Lets go to school an hour early today.	PROPOSAL
話題種別	What is the highest mountain in the world?	MOUNTAIN
	Who is president of the united states?	PERSON
	What is the name of the most recent Star Wars movie?	MOVIE
質問種別	Do you like egg salad?	TRUE/FALSE
	How do you correct a hook in a golf swing?	EXPLANATION:METHOD
	Why is blood red?	EXPLANATION:CAUSE

応する言語種別についての敵対的ネットワークにおける i' 番目ラベルの正解確率と予測確率を表す.

次に, 3つの損失関数に基づく各モデルパラメータの確率的勾配法による更新について述べる. 意図推定ネットワークのモデルパラメータは (22)-(23) 式に従い, 発話意図推定の性能を向上させるように更新される.

$$\theta_o^{(j)} \leftarrow \theta_o^{(j)} - \epsilon \frac{\partial L_o}{\partial \theta_o^{(j)}} \quad (22)$$

$$\theta_z^{(j)} \leftarrow \theta_z^{(j)} - \epsilon \frac{\partial L_o}{\partial \theta_z^{(j)}} \quad (23)$$

タスク依存ネットワークに対応する言語種別についての敵対的ネットワークでは, 言語種別の識別性能を向上させるようにパラメータが更新される. 一方, タスク依存ネットワークでは GRL を加味することで, 発話意図推定の性能を向上させるように, かつ, 言語種別に対して識別しづらくなるようにパラメータが更新される. j 番目のタスク依存ネットワーク対応する言語種別についての敵対的ネットワークのモデルパラメータ $\theta_y^{(j)}$, および j 番目のタスク依存ネットワークのモデルパラメータ $\theta_u^{(j)}$ のパラメータ更新は (24)-(25) 式に従う.

$$\theta_y^{(j)} \leftarrow \theta_y^{(j)} - \epsilon \beta \frac{\partial L_y}{\partial \theta_y^{(j)}} \quad (24)$$

$$\theta_u^{(j)} \leftarrow \theta_u^{(j)} - \epsilon \left(\frac{\partial L_o}{\partial \theta_u^{(j)}} - \beta \frac{\partial L_y}{\partial \theta_u^{(j)}} \right) \quad (25)$$

ここで, β は敵対的学習を調整するためのハイパーパラメータを表す.

同様に, 言語依存ネットワークに対応するタスク種別についての敵対的ネットワークでは, タスク種別の識別性能を向上させるようにパラメータが更新される. 一方, 言語依存ネットワークでは GRL を加味することで, 発話意図推定の性能を向上させるように, かつ, タスク種別や言語種別に対して識別しづらくなるようにパラメータが更新される. i 番目の言語依存ネットワークに対応するタスク種別についての敵対的ネットワークのモデルパラメータ $\theta_x^{(i)}$, i 番目の言語依存ネットワークのモデルパラメータ $\theta_h^{(i)}$, $\theta_w^{(i)}$ のパラメータ更新は (26)-(28) 式に従う.

表1 各データセットの発話数, およびラベル数.

言語	意図推定タスク	ラベル数	学習	開発	評価
日本語	対話行為	28	201 K	4 K	4 K
	話題種別	168	40 K	4 K	4 K
	質問種別	15	55 K	4 K	4 K
英語	対話行為	28	25 K	3 K	3 K
	話題種別	168	25 K	3 K	3 K
	質問種別	15	22 K	2 K	2 K

$$\theta_x^{(i)} \leftarrow \theta_x^{(i)} - \epsilon \alpha \frac{\partial L_x}{\partial \theta_x^{(i)}} \quad (26)$$

$$\theta_h^{(i)} \leftarrow \theta_h^{(i)} - \epsilon \left(\frac{\partial L_o}{\partial \theta_h^{(i)}} - \alpha \frac{\partial L_x}{\partial \theta_h^{(i)}} - \beta \frac{\partial L_y}{\partial \theta_h^{(i)}} \right) \quad (27)$$

$$\theta_w^{(i)} \leftarrow \theta_w^{(i)} - \epsilon \left(\frac{\partial L_o}{\partial \theta_w^{(i)}} - \alpha \frac{\partial L_x}{\partial \theta_w^{(i)}} - \beta \frac{\partial L_y}{\partial \theta_w^{(i)}} \right) \quad (28)$$

ここで, α は敵対的学習を調整するためのハイパーパラメータを表す.

なお, α , β を 0 と設定することにより, この学習は通常の結合ネットワークに対する最終基準の学習となる.

5. 評価実験

評価実験のために, 日本語と英語の2つの言語について, 3つの意図推定タスクのデータセットを準備した. 3つの意図推定タスクは, 対話行為推定, 和田種別推定, および質問種別推定の3種類であり, いずれのタスクにおいても, 日本語と英語で共通のラベル集合となっている. 我々は, 各言語の各タスクデータセットをそれぞれ学習データ, 開発データ, 評価データに分割した. 表1に各データセットの発話数, およびラベル数, 表2に各意図推定タスクについての英語の発話例とラベル例を示す.

5.1 実験条件

評価実験では, 言語と意図推定タスクの組ごとに独立なモデリングを行う場合, 同一タスクについて複数言語のデータを同時に用いた結合モデリングを行う場合, 同一言語について複数タスクのデータを同時に用いた結合モデリングを行う場合, そして複数言語複数タスクのデータを同時に用いた結合モデリングを行う場合の4種類のモデリン

表3 各評価データに対する発話意図推定の正解率(%)

	結合ネットワーク		敵対的ネットワーク		日本語			英語		
	複数タスク	複数言語	タスク種別	言語種別	対話行為	話題種別	質問種別	対話種別	話題種別	質問種別
(1).	-	-	-	-	66.6	79.1	87.7	61.8	64.5	83.4
(2).	✓	-	-	-	66.5	79.6	89.3	60.6	64.4	83.7
(3).	✓	-	✓	-	66.5	80.6	89.5	61.6	65.7	83.7
(4).	-	✓	-	-	66.7	78.7	87.2	61.4	64.3	83.0
(5).	-	✓	-	✓	66.9	79.8	88.2	61.8	64.8	83.3
(6).	✓	✓	-	-	66.6	79.7	89.3	60.5	65.4	82.6
(7).	✓	✓	✓	-	67.3	81.1	89.6	61.5	66.1	83.5
(8).	✓	✓	-	✓	66.7	80.7	89.5	60.9	66.7	83.0
(9).	✓	✓	✓	✓	67.6	81.3	90.0	61.9	66.7	83.7

グを準備した。そして、各結合モデリングについては、敵対的学習を行う場合と行わない場合の両者を準備した。

ネットワーク構造の詳細について述べる。単語単位の連続ベクトルの次元数は128、それぞれのBLSTM-RNNのユニットサイズは400、発話単位の連続ベクトルの次元数は400とした。学習時は、単語連続ベクトルに対して50%のドロップアウト、BLSTM中の再帰的な連続ベクトルに対して50%のドロップアウトを適用した。また、学習データ中の頻度1の単語に対して未知語トークンを与えることとした。最適化には、ミニバッチ確率的勾配法を用い、初期学習率を0.1として設定し、各開発セットに対する損失を各エポック後に観測することで調整した。敵対的学習を行わない場合は、 α , β の値を0に設定し、敵対的学習を行う場合は、0.001から1.0までの値を変更させながら、開発データで調整した。

5.2 実験結果

表3に評価データに対する発話意図推定の正解率の評価結果を示す。なお、各実験条件において学習時の初期値を変化させた上で、学習、評価の施行を5回実施し、その平均の正解率を評価した。

(1)が言語と意図推定タスクの組ごとに独立なモデリングを行った場合の結果であり、本稿におけるベースラインである。(2)-(3)が同一言語について複数タスクのデータを同時に用いた結合モデリングによる結果、(4)-(5)が同一タスクについて複数言語のデータを同時に用いた結合モデリングによる結果、そして(6)-(9)が複数言語複数タスクのデータを同時に用いた結合モデリングによる結果である。また、(3), (5), (7), (8), (9)が敵対的学習を導入した場合の結果であり、タスク種別にチェックが入っている場合はタスク種別についての敵対的ネットワークを用いた場合、言語種別にチェックが入っている場合は言語種別についての敵対的ネットワークを用いた場合を表す。

最初に、(2)や(4)において、結合モデリングを行うことにより、(1)と比較して性能が改善する場合と性能が悪化する場合の両者が観測できる。劣化する理由は、結合モデ

リングを行うことにより、異なる言語間や異なるタスク間で共有するネットワークが、特定の言語やタスクに依存することに起因すると考えられる。一方、(3)や(5)において敵対的学習を導入することにより、(2)や(4)よりも性能改善できることが見て取れる。このことから、敵対的学習が結合モデリングにおいて有効であることが示唆された。

次に、(6)において、複数言語複数タスクについて結合モデリングを行うことにより、日本語の話題種別、日本語の質問種別、英語の話題種別で改善効果を示した一方、英語の対話行為、英語の質問種別で性能劣化が見られた。このことから、複数言語複数タスクについての結合モデリングを様々な学習データを用いて最尤基準で学習するのみでは、不十分であることが示唆された。特に、学習データが比較的少ない場合において性能劣化が見取れ、異なる言語間や異なるタスク間で共有するネットワークが、特定の言語やタスクに依存してしまうことが理由であると考えられる。これに対して、(7), (8)において、言語種別について敵対的ネットワークやタスク種別について敵対的ネットワークを導入することにより、(6)と比較して改善効果を示した。さらに、(9)において両者の敵対的ネットワークを併用することにより、最高性能が得られ、いずれの評価データに対してもベースラインよりも高い性能を示した。この結果から、複数言語複数タスクを同時に扱う場合には、両者の敵対的ネットワークの併用が有効であることが示された。

6. まとめ

本稿では、発話意図推定の複数言語複数タスクを同時に扱う結合モデリングのための敵対的学習手法を提案した。提案法のポイントは、異なる言語間で共有可能なタスク依存ネットワークと異なるタスク間で共有可能な言語依存ネットワークに基づく複数言語複数タスクの結合ネットワークに、言語種別についての敵対的ネットワークとタスク種別についての敵対的ネットワークを組み合わせることである。これにより、特定の言語や特定のタスクに共有ネットワークが依存してしまう問題を緩和し、各言語、各

タスクにおける発話意図推定の性能向上を効率的に改善することを可能とした。評価実験により、複数言語複数タスクを扱う結合モデリングに対して提案手法の敵対的学習を導入することが有効であることが示された。

参考文献

- [1] Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Marton, R., Ess-Dykema, C. V. and Metter, M.: Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech, *Computational Linguistics*, Vol. 26, No. 3, pp. 339–373 (2000).
- [2] Khanpour, H., Guntakandla, N. and Nielsen, R.: Dialogue Act Classification in Domain-Independent Conversations Using a Deep Recurrent Neural Network, *In Proc. International Conference on Computational Linguistics (COLING)*, pp. 2012–2021 (2016).
- [3] Xu, P. and Sarikaya, R.: Contextual Domain Classification in Spoken Language Understanding Systems using Recurrent Neural Network, *In Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 136–140 (2014).
- [4] Wu, C.-H., Yeh, J.-F. and Chen, M.-J.: Domain-Specific FAQ Retrieval Using Independent Aspects, *ACM Transactions on Asian Language Information Processing*, Vol. 4, No. 1, pp. 1–17 (2005).
- [5] Collobert, R. and Weston, J.: A unified architecture for natural language processing: Deep neural networks with multi-task learning, *In Proc. International Conference on Machine Learning (ICML)* (2008).
- [6] Ravuri, S. and Stolcke, A.: A comparative study of neural network models for lexical intent classification, *In Proc. Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 368–374 (2015).
- [7] Ravuri, S. and Stolcke, A.: Recurrent Neural Network and LSTM Models for Lexical Utterance Classification, *In Proc. Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pp. 135–139 (2015).
- [8] Ravuri, S. and Stolcke, A.: A comparative study of recurrent neural network models for lexical domain classification, *In Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6075–6079 (2016).
- [9] Masumura, R., Tanaka, T., Higashinaka, R., Masataki, H. and Aono, Y.: Multi-task and Multi-lingual Joint Learning of Neural Lexical Utterance Classification based on Partially-shared Modeling, *In Proc. International Conference on Computational Linguistics (COLING)*, pp. 3586–3596 (2018).
- [10] Lin, Y., Yang, S., Stoyanov, V. and Ji, H.: A Multi-lingual Multi-task Architecture for Low-resource Sequence Labeling, *In Proc. Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 799–809 (2018).
- [11] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M. and Lempitsky, V.: Domain-Adversarial Training of Neural Networks, *Journal of Machine Learning Research*, Vol. 17, pp. 1–35 (2016).
- [12] Yang, Z., Yang, D., Dyer, C., He, X., Smola, A. J. and Hovy, E. H.: Hierarchical Attention Networks for Document Classification, *In Proc. Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pp. 1480–1489 (2016).
- [13] Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H. and Xu, B.: Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification, *In Proc. Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 207–212 (2016).
- [14] Liu, P., Qiu, X. and Huang, X.: Deep multi-task learning with shared memory, *In Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 118–127 (2016).
- [15] Liu, P., Qiu, X. and Huang, X.: Recurrent Neural Network for Text Classification with Multi-Task Learning, *In Proc. International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2873–2879 (2016).
- [16] Zhang, X. and Weng, H.: A Joint Model of Intent Determination and Slot Filling for Spoken Language Understanding, *In Proc. International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2993–2999 (2016).
- [17] Liu, Y., Li, S., Zhang, X. and Sui, Z.: Implicit Discourse Relation Classification via Multi-Task Neural Networks, *In Proc. AAAI Conference on Artificial Intelligence (AAAI)*, pp. 2750–2756 (2016).
- [18] Guo, J., Che, W., Yarowsky, D., Wang, H. and Liu, T.: A representation learning framework for multi-source transfer parsing, *In Proc. AAAI Conference on Artificial Intelligence (AAAI)*, pp. 2734–2740 (2016).
- [19] Duong, L., Kanayama, H., Ma, T., Bird, S. and Cohn, T.: Learning Crosslingual Word Embeddings without Bilingual Corpora, *In Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1285–1295 (2016).
- [20] Zhang, M., Liu, Y., Luan, H., Liu, Y. and Sun, M.: Inducing Bilingual Lexica From Non-Parallel Data with Earth Mover’s Distance Regularization, *In Proc. International Conference on Computational Linguistics (COLING)*, pp. 3188–3198 (2016).
- [21] Zhang, M., Peng, H., Liu, Y., Luan, H. and Sun, M.: Bilingual Lexicon Induction from Non-Parallel Data With Minimum Supervision, *In Proc. AAAI Conference on Artificial Intelligence (AAAI)*, pp. 3379–3384 (2017).
- [22] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y.: Generative Adversarial Nets, *In Proc. Advances in Neural Information Processing Systems (NIPS)*, pp. 2672–2680 (2014).
- [23] Chen, X., Sun, Y., Athiwaratkun, B., Cardie, C. and Weinberger, K.: Adversarial Deep Averaging Networks for Cross-Lingual Sentiment Classification, *arXiv preprint arXiv:1606.01614* (2016).
- [24] Liu, P., Qiu, X. and Huang, X.: Adversarial Multi-task Learning for Text Classification, *In Proc. Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 1–10 (2017).
- [25] Miyato, T., Dai, A. M. and Goodfellow, I.: Adversarial Training Methods for Semi-Supervised Text Classification, *In Proc. International Conference on Learning Representation (ICLR)* (2017).
- [26] Chen, X. and Cardie, C.: Multinomial Adversarial Networks for Multi-Domain Text Classification, *arXiv preprint arXiv:1802.05694* (2018).
- [27] Zhang, M., Liu, Y., Luan, H. and Sun, M.: Adversarial Training for Unsupervised Bilingual Lexicon Induction, *In Proc. Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 1959–1970 (2017).