関係パタンの分散表現の計算

Learning to Compose Distributed Representations of Relational Patterns

高瀬 翔^{*1}
Sho Takase

東北大学
Tohoku University
takase.sho@lab.ntt.co.jp, <http://www.kecl.ntt.co.jp/icl/lirg/members/takase/>

岡崎 直観
Naoaki Okazaki

(同 上)
okazaki@ecei.tohoku.ac.jp, <http://www.chokkan.org/>

乾 健太郎
Kentaro Inui

(同 上)
inui@ecei.tohoku.ac.jp, <http://www.cl.ecei.tohoku.ac.jp/~inui/>

keywords: relation extraction, semantic composition, neural network, data construction

Summary

Learning distributed representations for relation instances is a central technique in downstream NLP applications. In particular, semantic modeling of relations and their textual realizations (relational patterns) is important because a relation (e.g., causality) can be mentioned in various expressions (e.g., “X cause Y”, “X lead to Y”, “Y is associated with X”). Notwithstanding, the previous studies paid little attention to explicitly evaluate semantic modeling of relational patterns. In order to address semantic modeling of relational patterns, this study constructs a new dataset that provides multiple similarity ratings for every pair of relational patterns on the existing dataset [Zeichner 12]. Following the annotation guideline of [Mitchell 10], the new dataset shows a high inter-annotator agreement. We also present Gated Additive Composition (GAC), which is an enhancement of additive composition with the gating mechanism for composing distributed representations of relational patterns. In addition, we conduct a comparative study of different encoders including additive composition, RNN, LSTM, GRU, and GAC on the constructed dataset. Moreover, we adapt distributed representations of relational patterns for relation classification task in order to examine the usefulness of the dataset and distributed representations for a different application. Experiments show that the new dataset does not only enable detailed analyses of the different encoders, but also provides a gauge to predict successes of distributed representations of relational patterns in the relation classification task.

1. はじめに

タバコと癌の間には因果関係がある, ということや, マクベスはシェイクスピアと著作物関係にある, というように, エンティティ間にどのような関係があるかについての知識, すなわち, 関係インスタンスの知識は情報検索や質問応答, 含意関係認識など, 自然言語処理の様々な応用タスクにおいて重要である. 近年, 単語や文についての分散表現の有用性が示されているように [Cho 14, Kiros 15, Mikolov 13, Pennington 14, Sutskever 14], 関係インスタンスについても, 分散表現を学習することが有用である. 特に, 関係は様々な表現で言い表せる, 例えば因果関係は “X cause Y”, “X lead to Y”, “Y is associated with X” というような表現で言い表すことができ, さらに, “X increase the risk of Y” に *severe* を挿入することで, “X increase the severe risk of Y” という, 被害を強調した表現を作成することができる. このため, 関係, および, 関係を示す表現 (関係パタン) の意味のモデル化が必須と

なる.

関係パタンの意味をモデル化するために, 既存研究では, “X increase the risk of Y” のような関係パタンと, “X: smoking, Y: cancer” のようなエンティティペアとの, 共起頻度や自己相互情報量の行列を構築していた [Lin 01, Nakashole 12]. 分布仮説に基づく [Harris 54], この共起統計の行列から関係パタンの意味ベクトルを得ることができ, このベクトル間のコサイン類似度を計算することで, 関係パタン間の意味的な類似性を計算することができる. さらに, 近年では, 関係パタンの意味のモデル化のために, ニューラルネットワークを用いて計算した分散表現を利用する研究もある [Takase 16b, Yih 14].

しかしながら既存研究では, 関係パタンの意味ベクトルを応用タスクに適用し, その性能を検証するばかりであり, 関係パタンの意味のモデル化, すなわち, 関係パタン間の意味的な類似性を計算できているかを陽に評価することは軽視されていた. 本研究では, 関係パタンの意味のモデル化の評価のために, 関係パタンのペアに人手で類似度を付与したデータセットを構築する. このデータ

^{*1} 現所属は NTT コミュニケーション科学基礎研究所

図 1 研究の概要.

セットは, Mitchell らのガイドラインを参考にすることにより [Mitchell 10], 高い作業員間の一致率を達成した.

さらに, 関係パタンの分散表現計算手法として, 加法構成とゲート機構を組み合わせた, Gated Additive Composition という手法を提案する. 既存のニューラルネットワークを用いた様々な手法(エンコーダ), および, Gated Additive Composition について, 関係パタンの分散表現計算の性能を, 上記のデータセットを用いて比較する. 各エンコーダのパラメータは Skip-gram [Mikolov 13] の目的関数を用い, 大量のラベルなしコーパスから学習する.

また, 関係パタンの分散表現を, 応用タスク(関係分類)に適用する. 新たに構築したデータセット上での性能と応用タスクでの性能から, 構築したデータセットは関係パタンの分散表現計算における, 詳細な分析を可能とするだけでなく, 各エンコーダで計算した関係パタンの分散表現が応用タスクにどの程度有用かの推測もできることを示す. これら, 本研究の概要を図示したものが図 1 である.

なお, 本論文は, ACL2016^{*1}の予稿集に収録されている, Takase らの論文 [Takase 16a] を日本語化したうえで, Takase らの提案している, Recurrent Neural Network を応用した関係パタンの分散表現計算手法 [Takase 16b] との比較を行っている. さらに関係分類タスクにおいて, 実験を追加することにより, 提案手法である Gated Additive Composition および, 構築したデータセットの有用性を詳しく調査した.

2. データ構築

2.1 アノテーション対象

本研究では, 関係パターン間に含意関係が付与されたデータセットを基に [Zeichner 12], 新たなデータセットを構築する. ふたつの関係パターンについて, 一方が成立しているとき, もう一方も成立するならば, 前者は後者と含意関係にあると言う. 例えば, “X have a great influence on Y” という関係パターンは “X have an influence on Y” と

いう関係パターンと含意関係にある. Zeichner らの構築したデータセットは [Zeichner 12], Reverb [Fader 11] を用いてウェブ上の文書から抽出した関係パターンから, 5,555 個のペアをつくり^{*2}, 含意関係を付与したもので, 2,447 個の含意関係にあるペア, および, 3,108 個の含意関係にないペアから構成されている.

関係パタンの意味のモデル化の評価のために, 素朴な発想として, 含意と類似を同等と考え, この含意関係の付与されたデータセットを用いることが考えられる. しかしながら, 含意関係は意味的な類似性とは大きく異なる性質を持っている. 例えば, “X be the part of Y” と “X be an essential part of Y” というペアにおいて, X が “the small intestine”, Y が “the digestive system” である場合を考える. “X be the part of Y” は ‘essential’ という意味を含まないため, “X be an essential part of Y” を含意しないが, この 2 つのパターンは類似性が高く, PART-OF という関係を表すという点では同じである. 次に, “X fall down Y” と “X go up Y” というペアにおいて, X が “the dude”, Y が “the stairs” である場合を考える. 階段から落ちるためには, まず階段を上らなければならず, 実際, Zeichner らのデータセットでは前者は後者を含意としているが, 2 つのパターンは反対の動作を表している. これらの理由から, 本研究では, 各関係パターンペアについて, 意味的類似度を付与したデータセットの構築に取り組む.

2.2 アノテーション基準

意味的類似度を付与する作業員間の一致率を高く保つため, 本研究でも先行研究と同様 [Zeichner 12], インスタンスに基づく判定を行う. インスタンスに基づく判定では, 各作業員は関係パターンペアのそれぞれの変数が同一のエンティティペアで埋まった状態で判定を行う. すなわち, 彼らは “X prevent Y” と “X reduce the risk of Y” というような, 変数のある状態で判断するのではなく, “Cephalexin prevent the bacteria” と “Cephalexin reduce the risk of the bacteria” (X が ‘Cephalexin’, Y が ‘the bacteria’) というように, インスタンス化された状態で関係パターンペアの意味的類似性を判定する. 本研究でも, 先行研究と同一のインスタンスペアを用いる [Zeichner 12].

類似度付与の作業員には, 各関係インスタンスのペアについて, 1 (全く似ていない) から 7 (極めて似ている) の間の整数値を選択するよう指示した. 具体的には, 単語間の意味的類似性付与の手順を参考に [Mitchell 10], 以下のような指示を行った (1) “Palmer team with Jack Nicklaus” と “Palmer join with Jack Nicklaus” のように 2 つのインスタンスが同じか, ほとんど同じ内容の場合

^{*1} <http://acl2016.org/>

^{*2} より正確には, このデータセットは 5,555 個のペアに加え, 1,012 個の無意味なペアも含んでいる. 含意関係を評価する際に, 作業員がパタンの意味を理解できなかった場合, そのパターンを含むペアは無意味なものとされる. 本研究では, これら無意味なペアは無視することとした.

図2 各類似度の判定結果の数．類似度判定結果の総数は 27,775 (5,555 ペア × 5 作業者)．

に 6 か 7 を選択する (2) “the kids grow up with him” と “the kids forget about him” のように 2 つのインスタンスが全く異なるか、無関係な内容の場合に 1 か 2 を選択する (3) “Many of you know about the site” と “Many of you get more information about the site” について、意味的に同じではないが、知るという点について似ているというように、2 つのインスタンスの内容にある程度関連がある場合に 3, 4, 5 のいずれかを選択する．

2.3 アノテーション手順

類似性の判定にはクラウドソーシングサービスである、CrowdFlower^{*3}を利用した．CrowdFlower では評価データ、すなわち、正解とする類似度のスコアが付与された関係パタンのペアを類似度の判定作業中に挿入することで、各作業者の信頼性を測定することができる^{*4}．5,555 個の関係パタンのペアから 100 個をランダムに抽出し、本論文の著者らのうち 2 人が類似度を付与し、一致率の高い 80 個のペアを評価データとして利用した．なお、評価データの類似度は作業者の信頼性の評価のみに使用した．言い換えれば、評価データとして使用した類似度は最終的に破棄し、作業者の付与した類似度のみをデータセットに含めることとした．

質の高いデータセットを構築するために、CrowdFlower によって計算された信頼度のスコアが 75% 以上の作業者の判定結果のみを使用した．加えて、各ペアについて、少なくとも 5 人以上の作業者に判定させた．この結果、60 人の作業者が参加し、最終的に、各ペアについて、信頼度の上位 5 人の類似度の判定結果を採用した．

図 2 は各類似度の判定結果の数を示したものである．類似度 7 を付与した作業者の数は少ないが、これは、ほとんどの関係パターンペアにおいて、少なくとも 1 単語の内容語の違いがあるためと思われる．全ての類似度の判定結果について、標準偏差は 1.16 であり、また、各パターンペアの 5 個の類似度判定結果の平均と各作業者のピアマンの順位相関係数 ρ の平均は 0.728 であった．これらの結果は、データセットの構築において作業者の一致率が高いことを示している．

^{*3} <http://www.crowdfunder.com/>

^{*4} 作業者の信頼性を測定する際、評価データとのスコアについて、 ± 1 の違いは許容することとした．

3. 分散表現へのエンコーダ

前章で構築したデータセットに関連して、2 つの疑問が考えられる．ひとつ目は関係パタンの分散表現を計算するために、最も良い手法 (エンコーダ) は何かというものであり、ふたつ目は構築したデータセットは関係パタンの分散表現の応用タスクでの有用性を評価することができるか、というものである．これらの疑問に答えるため、本章では、様々な関係パタンの分散表現計算手法を探索する．

3.1 ベースライン

関係パタンの分散表現を得る素朴な手法としては、関係パターンを 1 個の記号 (単語) とみなし、単語の分散表現の学習と同様にして獲得する手法が考えられる．実際、Mikolov らはこの手法を採用し、訓練コーパス中でよく共起する単語同士をあらかじめまとめてしまっている [Mikolov 13]．しかしながら、この手法ではデータスパースネス問題によって、分散表現の質が低くなると予想される．

単純ながら効果的な手法として、加法構成も考えられる [Mitchell 10]．これは、関係パターンを構成する単語の分散表現の平均を関係パタンの分散表現とするものであり、より正確には、関係パターンが T 個の単語から構成される、すなわち、 $w_1, \dots, w_T$ と表せるとし、 $x_t \in \mathbb{R}^d$ を単語 w_t の分散表現とすると、 $\frac{1}{T} \sum_{t=1}^T x_t$ を関係パタンの分散表現とする．加法構成は、いくつかのニューラルネットワークを用いた手法よりも良い性能であることが、Muraoka らの研究によって示されている [Muraoka 14]．

関係パタンの分散表現計算手法として、Takase らは Recursive Neural Network を応用した手法を提案した [Takase 16b]．これは、関係パターン内の単語を、行列で表現すべき単語とベクトルで表現すべき単語とに分け、ベクトルで表現すべき単語を加法構成で合成した後、行列を再帰的に掛け合わせる手法である．具体的には、関係パターン p が n 単語の行列で表現すべき単語列 $w_1, \dots, w_n$ と $T - n$ 単語のベクトルで表現すべき単語列 $w_{n+1}, \dots, w_T$ からなるとしたとき、まず、ベクトルで表現すべき単語列の分散表現 v を加法構成により、以下の式で計算する．

$$v = \frac{x_{n+1} + x_{n+2} + \dots + x_T}{T - n} \quad (1)$$

次に、 v に対し、 w_n から w_1 までの行列を順に掛け合わせ、関係パタンの分散表現 h_p を得る．

$$h_p = g(W_1(\dots g(W_n g(v)))) \quad (2)$$

ここで、 W_i は単語 w_i に対応した、 $d \times d$ の行列 (パラメータ) であり、 $g(\cdot)$ は活性化関数 (本研究では $\tanh$) である．

3.2 リカレントニューラルネットワーク

リカレントニューラルネットワーク (Recurrent Neural Network, 以下 RNN とする), あるいはその発展型を用いて, 句や文の意味のモデル化を行う研究が増えている [Sutskever 14, Tang 15]. 位置 t の分散表現 x_t が与えられた際, 最も単純な RNN [Elman 90] では, 隠れ状態 $h_t \in \mathbb{R}^d$ を以下の再帰式によって計算する^{*5}.

$$h_t = g(W_x x_t + W_h h_{t-1}). \quad (3)$$

ここで, W_x と W_h は $d \times d$ の行列 (パラメータ)^{*6}, $g(\cdot)$ は活性化関数 (Takase らと同様, 本研究では $\tanh$) であり, $t = 1$ においては $h_0 = 0$ とする. 端的に言えば, RNN は単語の分散表現 x_t と, ひとつ前の位置の隠れ状態 h_{t-1} を用いて隠れ状態 h_t を計算する. 式 (3) を $t = 1$ から T まで適用し, 終端での隠れ状態 h_T を関係パタンの分散表現として使用する.

3.3 RNN の発展型

RNN に加え, その発展型である Long Short-Term Memory (LSTM) や Gated Recurrent Unit (GRU) も関係パタンのエンコードとして利用する. LSTM は単語分割, 依存構造解析, 機械翻訳, 評判分析など, 様々な NLP のタスクにおいて高い性能を示している [Chen 15, Dyer 15, Sutskever 14, Tai 15]. GRU もまた, 機械翻訳や文間の類似度計算, 言い換え認識, 評判分析など, 様々なタスクに用いられている [Cho 14, Kiros 15].

LSTM と GRU は RNN の勾配の消失や爆発に対処するため, ゲート機構を持っているという点で似ており, LSTM には入力, 忘却, 出力ゲートが, GRU にはリセットと更新ゲートがそれぞれ備わっている. GRU は LSTM よりも性能が良いと報告している研究もあるが [Chung 14], 優劣について統一的な見解はまだまだなく, 数単語からなる関係パタンの分散表現へのエンコードにどちらが適しているかは明らかでない. 本研究では, RNN, LSTM, GRU を同一の学習手順, コーパスを用いて学習し, 実験的に比較を行う. RNN と同様, LSTM^{*7} と GRU^{*8} でも終端での隠れ状態 h_T を関係パタンの分散表現として利用する.

^{*5} 本研究では, Skip-gram モデルの目的関数で学習を行う. このためには, 関係パタンの分散表現と単語の分散表現とが同一の次元数である必要がある. 隠れ状態に変換行列 (パラメータ) を適用することで, 隠れ状態の次元数がどの値であっても, 単語の分散表現の次元数にそろえることができるが, いたずらにパラメータを増やすと過学習の問題など, 学習を難しくしてしまうおそれがある. このため, 本研究では, 隠れ状態の次元数を単語の分散表現の次元数 d と同一にした.

^{*6} 一般的に, RNN ではバイアス項を用いる. しかしながら, 本研究の実験では, バイアス項を使用しない方がわずかに性能が良かったため, バイアス項は使用しないこととする.

^{*7} 本研究では, peephole を使用しない. また, RNN と同じく, バイアス項を使用しない方が実験においてわずかに性能が良かったため, バイアス項は使用しない.

^{*8} RNN, LSTM と同じく, 実験においてバイアス項を使用しない方がわずかに性能が良かったため, バイアス項は使用しない.

図3 Skip-gram モデルでの GAC の学習概要. GAC は入力ゲートと忘却ゲートを用いて関係パタンの分散表現を計算し, 周囲の単語を予測できるようにパラメータを学習する (Skip-gram モデル).

3.4 Gated Additive Composition

勾配の問題への対処に加え, LSTM と GRU は入力単語と隠れ状態の重要性をゲートによって調整できるため, 関係パタンの分散表現計算に適していると考えられる. 例として, “X have access to Y” という関係パタンについて考える. このパタン中での ‘have’ は軽動詞であり, 意味は “X access Y” というパタンと極めて近い. このため, ‘have’ や機能語である ‘to’ の分散表現を関係パタンの分散表現に組み込んでしまうことは望ましくないと考えられる. 加法構成や RNN にはこれらの単語を無視するような機構がないが, 加法構成にゲート機構を組み込むことで単語の重要性を考慮可能にするなど, 加法構成と LSTM や GRU との折衷を探ることは興味深い課題である.

本研究では, LSTM の入力ゲートと忘却ゲートに着想を得た, 新たな RNN を提案する. 位置 t において, 入力ゲート $i_t \in \mathbb{R}^d$ と忘却ゲート $f_t \in \mathbb{R}^d$ を用い, h_t を計算する際に, x_t と h_{t-1} をどの程度重視するかを調整する.

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1}) \quad (4)$$

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1}) \quad (5)$$

$$h_t = g(f_t \odot h_{t-1} + i_t \odot x_t) \quad (6)$$

ここで, $W_{ix}, W_{ih}, W_{fx}, W_{fh}$ は $d \times d$ の行列である. 式 (6) は現在位置の単語のベクトル x_t とひとつ前の隠れ状態 h_{t-1} との重み付き和とみなせる. 要素ごとの重みは入力ゲート i_t および忘却ゲート f_t で調整する. 例えば現在位置の単語が軽動詞や機能語の場合, 入力ゲートは閉じ (値が 0 に近くなり), 忘却ゲートは開く (値が 1 に近くなる) ことが期待される. 本手法を *Gated Additive Composition* (GAC) と呼ぶこととする.

3.5 パラメータ推定 (Skip-gram モデル)

各エンコードのパラメータをラベルなしコーパスから学習するために, Skip-gram モデルを適用する [Mikolov 13]. 関係パタン p が L 個の単語からなる ($w_s, \dots, w_{s+L-1}$) とし, 関係パタン p の周囲に出現する δ 個の単語 (文脈単語) を $C_p = (s - \delta, \dots, s - 1, s + L, \dots, s + L + \delta)$ というインデックスで表すとしたとき, 関係パタン p から文脈

単語を予測した対数尤度 l_p は負例サンプリングを用いた Skip-gram モデル (Skip-gram with negative sampling, SGNS [Levy 14]) により, 以下の式で表せる.

$$l_p = \sum_{\tau \in C_p} \left(\log \sigma(h_p^\top \tilde{x}_\tau) + \sum_{k=1}^K \log \sigma(-h_p^\top \tilde{x}_{\tau'}) \right) \quad (7)$$

この式において, K はサンプルする負例の数, $h_p \in \mathbb{R}^d$ は RNN のような, 各エンコードによって計算された関係パターン p の分散表現, $\tilde{x}_\tau \in \mathbb{R}^d$ は文脈単語 w_τ についての文脈ベクトル^{*9}, $\tilde{x}_{\tau'} \in \mathbb{R}^d$ は $\sum_k$ の繰り返し操作において, 単語の生起確率からサンプリングされた単語^{*10}の文脈ベクトルである.

コーパス中に出現している各関係パターンについて, 確率的勾配降下法と Backpropagation through time (BPTT) を用いてエンコードのパラメータ (行列) を学習する. 具体的には, 単語ベクトル x_t および文脈ベクトル $\tilde{x}_t$ を事前に学習した値で初期化し, 式 (7) によって計算した勾配を用いてエンコードのパラメータを更新する. これにより, 各エンコードは周囲の単語を予測できるような, 関係パタンの分散表現の合成法を学習する. 単語の分散表現と関係パタンの分散表現とが同一のベクトル空間で表現できる点が, このパラメータ学習手法の利点である.

4. 実験

4.1 節では, 各エンコードの計算した関係パタンの分散表現がどの程度意味的類似性をとらえられるかを測定する. 4.2 節では, 関係パタンの分散表現を関係分類タスク (SemEval 2010 Task 8) に適用し, 新たに構築したデータセットと応用タスクとの関連性について議論する.

4.1 関係パタンの意味的類似度

2 章で構築したデータセット内の各ペアについて, 3 章で説明したエンコードを用いて分散表現を計算し, 各ペアの分散表現間のコサイン類似度を計算する. データセット内の全てのペアについてコサイン類似度を計算し, 人手で付与した類似度とのスピアマンの順位相関係数 ρ を測定する.

§1 学習手順

エンコードの学習用のコーパスとして, ukWaC^{*11}を用いた. このコーパスは uk ドメインの Web ページをク

図4 各次元数における, 関係パタンの意味的類似度推定の実験.

ロールしたものであり, 約 20 億語の単語を含んでいる. コーパス中の単語は TreeTagger^{*12}によって品詞とレンマが付与されている. 実験には小文字化したレンマを用いた. 単語ベクトル x_t と文脈ベクトル $\tilde{x}_t$ を, このコーパスに word2vec を適用することで事前に獲得し, 各エンコードでの関係パタンの分散表現計算および目的関数と勾配の計算に用いた. なお, 本実験では, 単語ベクトル x_t と文脈ベクトル $\tilde{x}_t$ は word2vec による事前学習の値で固定し, 学習しないものとする.

関係パタンの候補を抽出するため, ukWaC に Reverb を適用した [Fader 11]. ノイズを除去するため, Reverb の抽出結果として配布されているデータセット^{*13}の構築の際と同様のフィルタリングを行った. さらに, エンコードが未知の関係パタンの分散表現を計算可能かを調べるため, 評価データ中に出現している関係パターンを除去した. これらの処理の結果, 127,677 種類の関係パターンを得た.

また, Takase らの提案したエンコードを再現するにあたって, 彼らと同様, 関係パターン中での出現頻度をもとに, 行列で表現する単語とベクトルで表現する単語とを分けた [Takase 16b]. ただし, Takase らが内容語, すなわち, 形容詞, 名詞, 動詞のみを扱っていたのに対し, 本研究では, 他のエンコードと同様, 機能語を含めた全単語を扱い, 関係パタンの分散表現を計算する. また, L1 正則化の係数は, Takase らが報告している中で, 最も性能の高い, $\lambda = 10^6$ を使用した.

Skip-gram モデル学習時のハイパーパラメータは Mikolov ら [Mikolov 13] と同様の値を用いた. すなわち, 文脈の窓幅 $\delta = 5$, 負例のサンプル数 $K = 5$, サブサンプリングの係数 10^{-5} である. 学習率については, 0.025, 0.0025 および 0.00025 を初期学習率として, ukWaC の約 5% を用いて学習し, 各エンコードについて最も性能の良かった値を選択した. なお, 関係パターンを逆順にエンコードする (すなわち, 最後の単語から最初の単語に向かってエンコードする) 方がわずかに性能が良かったため, 実験では逆順にエンコードすることとした. ukWaC に出現している, 訓練対象の全関係パターンについて, 各エン

*9 Skip-gram モデルは単語 w_t について, 単語ベクトル x_t と文脈ベクトル $\tilde{x}_t$ という, 2 種類のベクトルを持つ. Skip-gram モデルを提案した論文 [Mikolov 13] の式 (2) では単語ベクトル x_t を v (入力ベクトル) とし, 文脈ベクトル $\tilde{x}_t$ を v' (出力ベクトル) としている. Skip-gram モデルを実装した word2vec では文脈 (出力) ベクトルは書き出さず, 単語 (入力) ベクトルのみ書き出している. このため, 実験では文脈ベクトルを出力できるように書き換え, それを式 (7) で使用している.

*10 Mikolov ら [Mikolov 13] と同様, コーパス中での単語の出現確率を $3/4$ 乗したものをを用いた.

*11 <http://wacky.sslmit.unibo.it/>

*12 <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

*13 <http://reverb.cs.washington.edu/>

表 1 パタンの長さを変えた際の，スピアマンの順位相関係数（分散表現の次元数 $d = 500$ ）．

長さ	#	合成なし	加法構成	LSTM	GRU	RNN	GAC
1	636	0.3186 $\pm$ 0.0052	0.3197 $\pm$ 0.0033	0.3197 $\pm$ 0.0033	0.3197 $\pm$ 0.0033	0.3197 $\pm$ 0.0033	0.3197 $\pm$ 0.0033
2	1,018	0.2140 $\pm$ 0.0018	0.3179 $\pm$ 0.0010	0.2518 $\pm$ 0.0069	0.2777 $\pm$ 0.0023	0.2794 $\pm$ 0.0041	0.3213 $\pm$ 0.0013
3	2,272	0.2344 $\pm$ 0.0013	0.3852 $\pm$ 0.0004	0.3398 $\pm$ 0.0031	0.3690 $\pm$ 0.0014	0.3826 $\pm$ 0.0032	0.4014 $\pm$ 0.0038
4	1,206	0.2085 $\pm$ 0.0002	0.3060 $\pm$ 0.0007	0.3038 $\pm$ 0.0071	0.3303 $\pm$ 0.0024	0.3256 $\pm$ 0.0050	0.3254 $\pm$ 0.0017
> 5	423	0.2762 $\pm$ 0.0016	0.3114 $\pm$ 0.0024	0.3732 $\pm$ 0.0029	0.3804 $\pm$ 0.0078	0.3956 $\pm$ 0.0051	0.3576 $\pm$ 0.0031
全ペア	5,555	0.2144 $\pm$ 0.0006	0.3399 $\pm$ 0.0004	0.3312 $\pm$ 0.0035	0.3555 $\pm$ 0.0004	0.3605 $\pm$ 0.0011	0.3692 $\pm$ 0.0011

コードのパラメータの更新を行った．すなわち，各エンコードは ukWaC を 1 エポック学習した．

§2 結果

分散表現の次元数を 100 から 500 まで 100 毎に変化させた際の，各エンコードのスピアマンの順位相関係数を図 4 に示した．この図から，GAC は全ての次元で最も良い性能を達成していることが分かる．

図 4 には，関係パタンを 1 つの記号（単語）とみなす，素朴な手法の結果も示してある（合成なし）．この手法では，学習時に，評価データ中に出現している関係パタンの除去を行わず，式 (7) における，関係パタン p の分散表現 h_p について，構成単語から合成したベクトルではなく，Skip-gram モデルを使って学習した単語ベクトルを用いた．この手法の性能が低いことには 2 つの理由があると考えられる．ひとつ目は ukWaC に出現しない関係パタンがあるため，評価データにあるペアのうち，1,744 ペアの類似度が測定できなかったこと，ふたつ目はデータスパースネス問題により，関係パタンについて，十分な統計値が得られなかったことである．

図 4 では，Takase らの手法 [Takase 16b] を再現した結果は，加法構成よりも性能が低くなってしまっている．Takase らとは評価用のデータや評価指標が異なるため，結果の傾向も異なるのであろう．また，本実験では，加法構成や RNN ベースの手法との比較のために，内容語以外に機能語も使って関係パタンの分散表現を計算している．このため，ひとつの関係パタンの分散表現を計算する際に用いるベクトル，および行列が Takase らよりも多くなっている．加えて，学習対象が単語毎に用意した行列となっており，パラメータ数が RNN に比べると 200 倍以上と膨大になっているため，学習が難しいのだと思われる．

図 4 から，LSTM や GRU は RNN より性能が低いことが分かる．LSTM，および，GRU について，所持しているゲート機構が全て開いているとき，すなわち，全てのゲートの要素の値が 1 のとき，これらは RNN に等しい．言い換えれば，LSTM や GRU は RNN と同等の性能を達成できるはずである．しかしながら，図 4 はそのような結果にはなっていない．これは，LSTM や GRU のパラメータ数が RNN に比べて多いため，学習が収束していない，あるいは，過学習している可能性を示唆している．

関係パタンの長さ毎にスピアマンの順位相関係数を測定したものを表 1 に示す．なお，ペアになっている関係

表 2 入力 / 忘却ゲートが特徴的な状態．

	w_t	w_{t+1} w_{t+2} ...
i_t が大きい （入力開）	reimburse payable liable	for in to
i_t が小さい （入力閉）	a a be	charter member of valuable member of an avid reader of
f_t が大きい （忘却開）	be be be	eligible to participate in require to submit request to submit
f_t が小さい （忘却閉）	coauthor capital center	of of of

パタンのうち，より長いものの長さを関係パタンペアの長さとした．また，表 1 の各値は，各エンコード毎に，異なる初期値で学習した 3 つのモデルのスピアマンの順位相関係数の平均と標準偏差である．単語ベクトル x_t は合成なし以外共通なため，長さ 1 については合成なし以外の全ての手法が同じ値となっている．表 1 から，加法構成は短い関係パタン（長さ 2 や 3）において性能が良く，長いパタン（長さ 4 と 5 以上）では性能が悪いことが分かる．反対に，RNN とその発展型（LSTM，GRU）は長いパタン（長さ 4 と 5 以上）で高い性能を達成している．GAC は加法構成と似た傾向だが，入力と忘却ゲートでベクトルを足すときの重みを調整することで，加法構成よりも性能が向上している．

GAC における入力ゲートおよび忘却ゲートの働きを調べるため，入力 / 忘却ゲートが開いているとき，および閉じているときの状況を調査する．より正確には，評価データ内の関係パタンのエンコードにおいて， $|i_t|_2$ や $|f_t|_2$ が小さい（0 に近い）とき，および，大きい（1 に近い）ときの位置の単語および読み込み済みの単語を調べた．なお，関係パタンの構成単語を逆順にエンコードしているため，例えば“be author of”というパタンについては，“of”，“author”，“be”という順で分散表現を計算していく．入力 / 忘却ゲートが閉じているとき，および，開いているときについて，それぞれ上位 3 つの例を表 2 に記した．この表から，読み込み済みの単語が前置詞のみであり，現在の位置の単語が内容語である場合，入力ゲートは開き，忘却ゲートは閉じることが分かる．すな

わち、この状況では、GAC は前置詞の分散表現を無視し、内容語の分散表現を読もうとしている。反対に、現在の位置の単語が ‘be’ や ‘a’ のような単語であり、かつ、読み込み済みの単語列が “charter member of” のように名詞句である場合や “eligible to participate in” のように補部である場合、あるいは “require(d) to submit” のように受動態の場合に、入力ゲートは閉じ、忘却ゲートは開いている。すなわち、GAC は機能語よりも、有益な単語群の情報を保持するように働いている。実際、“be defeated in” と “be defeated at” や “reduce the chance of” と “reduce the potential for” のように、意味的な類似度は高いが前置詞が異なる関係パターンペアに対し、GAC は加法構成よりも性能が向上している。

4.2 関係分類

§1 実験設定

構築したデータセット、および、分散表現の応用タスクでの有用性を検証するため、SemEval 2010 Task 8 データセット [Hendrickx 10] を用いた関係分類タスクに取り組む。言い換えれば、関係パタンの分散表現の質の高さが、エンティティペアの間の関係推定に役立つかを調べる。

関係分類タスクは、与えられたエンティティペアに対し、候補となる関係タイプから適切な関係タイプを選択するタスクである。例えば、“The *burst* has been caused by water hammer *pressure*” という文中の $e_1 = \text{‘burst’}$ と $e_2 = \text{‘pressure’}$ がエンティティペアとして与えられた場合、CAUSE-EFFECT(e_2, e_1) を予測できるようにする。SemEval 2010 Task 8 データセットは、10,717 個の関係インスタンスからなり、8,000 の訓練事例と 2,717 のテスト事例を含んでいる。このデータセットに含まれる関係タイプは、9 種類の、CAUSE-EFFECT のような有向の関係タイプと、1 種類の OTHER という無向の関係タイプであり、候補となる関係タイプは 19 種類 (2×9 有向 + 1 無向) である。

実験には、LIBSVM^{*14} に実装されている、線形カーネル、および、放射基底関数 (Radial Basis Function, RBF) カーネルを用いたサポートベクターマシン (Support Vector Machine, SVM) を利用した。具体的には、関係パターン (エンティティ間の単語列)、エンティティ、および、エンティティペアの前後 1 単語の分散表現 ($d = 500$) を素性として、線形カーネルを用いた SVM を学習した。なお、関係パタンの分散表現については、合成なし、および、4.1 節で構築した 5 つのエンコーダ、すなわち、加法構成、RNN、GRU、LSTM、および、GAC で計算したものをそれぞれ用いて比較した。さらに、エンティティペア間に出現している品詞 (TreeTagger によって付与)、表層形、レンマ、および、エンティティペアのレンマを素性として追加し、RBF カーネルを用いた SVM を学習

した。SVM のハイパーパラメータは訓練データの 5 分割交差検定で求めた。

§2 結果

SemEval 2010 Task 8 を用いた関係分類タスクの F1 値のマクロ平均を表 3 に示した。表の上側のグループは分散表現を素性としたときの、線形カーネルを用いた SVM での性能を示しており、ひとつ下のグループは、分散表現に加え、品詞など追加の素性を利用したときの、RBF カーネルを用いた SVM での性能を示している。最上段の、分散表現のみを素性としたときの結果から、RNN や GRU、GAC など、関係パタンの意味的類似度推定タスクで性能が高いモデルは、関係分類タスクでも性能が高い傾向があることが分かる。特に、関係パタンの意味的類似度推定タスクで最も性能の良かった GAC が最も高いスコアを達成している。

2 段目のグループからは、分散表現に加え、品詞や単語などの素性を追加することで、性能が向上することが分かる。この結果から、合成なしでの分散表現を利用した場合でさえ、性能が向上することが分かる (77.3% から 79.9%)。さらに、品詞と単語のみを素性としたときの性能の向上度合いは、関係パタンの意味的類似度推定タスクにおいて高い性能を示すものであるほど、高いことが分かる。また、ここでも、GAC が最高スコアを達成している。

各エンコーダによる性能の向上と関係パタンの意味的類似度推定タスクでの性能がおおよそ対応づいていることは極めて興味深く、この結果は、2 つの可能性を示唆していると考えられる。ひとつ目は関係パタンの分散表現は知識ベースの拡張のような応用タスクに容易に適用でき、かつ、効果的であるということであり、ふたつ目は本研究で構築した、関係パターン間に意味的類似度を付与したデータセットは、応用タスクでの性能向上の程度も予測できる、ということである。

また、SVM+GAC について、先行研究と同様に外部資源を導入することで、さらに性能を向上させることができた。具体的には、WordNet、固有名詞、依存構造のパスを素性として加え [Hashimoto 15]、F1 値のマクロ平均で 83.7 を達成した。さらに、ランキング学習、および、ノイズなタイプである OTHER の事例を推定するような学習を行わないという工夫を用い [Santos 15]、GAC のパラメータや単語の分散表現を再学習することで、スコア 84.2 を得た。Xu ら [Xu 15] の提案した負例サンプリング手法など、表 3 の下側のグループに示してある、より性能の高い手法で用いられている手法を導入することで、さらに性能の向上が期待できる。しかしながら、本実験では、関係パタンの分散表現、および、関係パターン間に意味的類似度を付与したデータセットの有用性を検証することが目的であるため、上記の、性能の向上に取り組むことは本研究のスコープ外である。

*14 <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

表 3 SemEval 2010 Task 8 での F1 値.

手法	素性	F1 値のマクロ平均
SVM (線形カーネル) + 合成なし	分散表現	55.2
SVM (線形カーネル) + LSTM	分散表現	73.7
SVM (線形カーネル) + Add	分散表現	71.9
SVM (線形カーネル) + GRU	分散表現	74.8
SVM (線形カーネル) + RNN	分散表現	73.8
SVM (線形カーネル) + GAC	分散表現	75.2
SVM (RBF カーネル)	Bag of Words (BoW), 品詞	77.3
SVM (RBF カーネル) + 合成なし	分散表現, BoW, 品詞	79.9
SVM (RBF カーネル) + LSTM	分散表現, BoW, 品詞	81.1
SVM (RBF カーネル) + Add	分散表現, BoW, 品詞	81.1
SVM (RBF カーネル) + GRU	分散表現, BoW, 品詞	81.4
SVM (RBF カーネル) + RNN	分散表現, BoW, 品詞	81.7
SVM (RBF カーネル) + GAC	分散表現, BoW, 品詞	82.0
	+ 依存構造, WordNet, NE	83.7
ランキング学習 + GAC (再学習)	分散表現, BoW, 品詞 + 依存構造, WordNet, NE	84.2
SVM [Rink 10]	BoW, 品詞, 依存構造, Google n-gram など	82.2
SVM [Takase 16b]	分散表現, BoW, 品詞 + WordNet, NE	79.0 82.1
MV-RNN [Socher 12]	分散表現, 句構造木 + WordNet, 品詞, NE	79.1 82.4
FCM [Gormley 15] (分散表現学習なし)	分散表現, 依存構造	79.4
	+ WordNet	82.0
(分散表現学習あり)	分散表現, 依存構造 + NE	82.2 83.4
RelEmb [Hashimoto 15]	分散表現 + 依存構造, WordNet, NE	82.8 83.5
CR-CNN [Santos 15]	分散表現, 単語の位置情報の分散表現	82.7
(Other を特別扱い)	分散表現, 単語の位置情報の分散表現	84.1
SPTree [Miwa 16]	分散表現, 依存構造 + WordNet	84.4 85.5
depLCNN [Xu 15]	分散表現, 依存構造 + WordNet	81.9 83.7
depLCNN + NS	分散表現, 依存構造 + WordNet	84.0 85.6
Att-Pooling-CNN [Wang 16]	分散表現, 単語の位置情報の分散表現	88.0

5. 関 連 研 究

本研究では関係パターンを対象として意味的類似度を付与したが、他の単位を対象として、データセットを構築した研究もいくつかある。Mitchell らは 2 単語からなる句の意味のモデル化を行うため、句の間に、人手で意味的類似度を付与したデータセットを構築した [Mitchell 10]。Korkontzelos らは 2 単語からなる句と単語間に意味的類似度を付与したデータセットを構築した [Korkontzelos 13]。Wieting らは PPDB [Ganitkevitch 13] の一部分に対

し、言い換え関係にあるかどうか、人手であらためて判断したデータセットを構築した [Wieting 15]。本研究では Mitchell ら [Mitchell 10] の用いた類似度付与の基準を参考にし、新たなデータセットを構築した。

本研究で取り組んでいるタスクは、2 文間の意味的類似度を推定するタスクである、Semantic Similarity Task (STS) とも関連がある [Agirre 12]。関係パターンは文の一部ではあるが、本研究で用いたエンコーダは入力の高さや複雑さによって異なる特徴を見せるかもしれない、本研究での知見がそのまま STS に適用できるとは限らない、

データセットの構築に加え、本研究では、関係パタンの意味をモデル化する手法の探求も行った。Nakashole らは関係パターンとエンティティペアとの間の共起頻度を基に、関係パターン間に含意関係を付与した [Nakashole 12]。Grycner らはエンティティの上位下位関係を用い、さらに、関係パタンの含意関係に推移律を適用することで、Nakashole らの手法を発展させた [Grycner 15]。しかしながら、これらの手法は構成単語から関係パタンの意味をモデル化することを考えておらず、図 4 での合成なしの結果と同様、データスパースネス問題を被ってしまう可能性がある。

単語の分散表現を句や文の分散表現にエンコードする手法には、Recursive Neural Network [Socher 11]、Matrix Vector Recursive Neural Network [Socher 12]、統語カテゴリごとに行列を用意した Recursive Neural Network [Socher 13] や単語ごとに行列を用意した Recursive Neural Network [Takase 16b]、RNN [Sutskever 11]、PAS-CLBLM [Hashimoto 14]、LSTM [Sutskever 14]、および、GRU [Cho 14] と、様々なものがある。3 章で述べたように、RNN、GRU、LSTM がエンコーダとして優れている、という報告があるため [Sutskever 14, Tang 15]、本研究では、関係パタンの分散表現を計算するために、これらのエンコーダを用いた。

分散表現、および、新たに構築したデータセットの有用性を検証するため、本研究では、応用タスクである、関係分類タスク (SemEval 2010 Task 8) にも取り組んだ。SemEval 2010 Task 8 においては、単語の分散表現と 2 値の素性 (単語がエンティティペアの間に出現しているかなど) との外積で素性を構築する手法 [Gormley 15] や Convolutional Neural Network (CNN) を用いるもの [Santos 15, Wang 16, Xu 15]、木構造をエンコードする、Tree-LSTM [Tai 15] を依存構造木に適用するもの [Miwa 16] などが提案されている。さらに、Wang らは、どの分散表現に着目すべきかをエンティティや関係から選択できるよう、アテンション機構を CNN に導入した [Wang 16]。

関係分類以外の応用タスクを概観すると、エンティティ、関係パターン、知識ベースに定義されている関係のタイプの潜在的なベクトル表現を同時に学習し、知識ベースを拡充する、Knowledge base population がある [Riedel 13]。Toutanova らは、この同時学習において、CNN を適用し、関係パタンの分散表現を構成要素から計算する手法を提案した [Toutanova 15]。また、質問文と知識ベース中のエンティティ、関係のタイプとの対応を学習し、質問応答を行うことも、応用タスクのひとつである [Yih 14]。これらの研究は各タスクにおいて良い性能を達成しているが、特定のタスクを解くために学習した分散表現が他のタスクへの応用性があるかどうか、明らかでない。反対に、本研究では、単語の分散表現の学習手法という、汎用的な分散表現を得る手法で関係パタンの分散表現を獲得

し、この分散表現の関係分類タスクへの有用性を示した。

6. お わ り に

本研究では、関係パタンの意味のモデル化に取り組んだ。含意関係データセット [Zeichner 12] 内に含まれる各関係パタンのペアについて、人手で意味的類似度を付与したデータセットを構築した。さらに、関係パタンの分散表現計算における、各エンコーダの性能を比較した。実験を通して、提案手法である、加法構成とゲート機構を組み合わせた GAC が関係パタンの分散表現計算に有効であること、関係パタンの分散表現が応用タスクに有効であること、構築したデータセットが関係分類タスクでの性能予測に貢献することを示した。

本研究では、関係パタンの意味的類似度推定のためにデータセットを構築したが、このタスクは、言い換え認識など、他の NLP タスクとの共通点も多い。今後は、新たに構築したデータセットや関連するタスクのデータセットを用いて LSTM、GRU、GAC の内部機構を解析することにより、どのようなタスクにおいてどのような機構が最適なのかを明らかにしていきたい。

謝 辞

本研究は科研費 JP15H01702, JP15H05318, JP14J05820, および、JST、CREST の支援を受けたものである。

◇ 参 考 文 献 ◇

- [Agirre 12] Agirre, E., Cer, D., Diab, M., and Gonzalez-Agirre, A.: SemEval-2012 Task 6: A Pilot on Semantic Textual Similarity, in *The First Joint Conference on Lexical and Computational Semantics (*SEM 2012)*, pp. 385–393 (2012)
- [Chen 15] Chen, X., Qiu, X., Zhu, C., Liu, P., and Huang, X.: Long Short-Term Memory Neural Networks for Chinese Word Segmentation, in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015)*, pp. 1197–1206 (2015)
- [Cho 14] Cho, K., Merriënboer, van B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y.: Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, pp. 1724–1734 (2014)
- [Chung 14] Chung, J., Gülcehre, Ç., Cho, K., and Bengio, Y.: Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling, in *Proceedings of the Deep Learning and Representation Learning Workshop in NIPS 2014* (2014)
- [Dyer 15] Dyer, C., Ballesteros, M., Ling, W., Matthews, A., and Smith, N. A.: Transition-Based Dependency Parsing with Stack Long Short-Term Memory, in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2015)*, pp. 334–343 (2015)
- [Elman 90] Elman, J. L.: Finding structure in time, *Cognitive Science*, Vol. 14, No. 2, pp. 179–211 (1990)
- [Fader 11] Fader, A., Soderland, S., and Etzioni, O.: Identifying Relations for Open Information Extraction, in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP 2011)*, pp. 1535–1545 (2011)
- [Ganitkevitch 13] Ganitkevitch, J., Van Durme, B., and Callison-Burch, C.: PPDB: The Paraphrase Database, in *Proceedings of the*

- 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2013), pp. 758–764 (2013)
- [Gormley 15] Gormley, M. R., Yu, M., and Dredze, M.: Improved Relation Extraction with Feature-Rich Compositional Embedding Models, in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015)*, pp. 1774–1784 (2015)
- [Grycner 15] Grycner, A., Weikum, G., Pujara, J., Foulds, J., and Getoor, L.: RELLY: Inferring Hypernym Relationships Between Relational Phrases, in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015)*, pp. 971–981 (2015)
- [Harris 54] Harris, Z.: Distributional structure, *Word*, Vol. 10, No. 23, pp. 146–162 (1954)
- [Hashimoto 14] Hashimoto, K., Stenetorp, P., Miwa, M., and Tsuruoka, Y.: Jointly Learning Word Representations and Composition Functions Using Predicate-Argument Structures, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, pp. 1544–1555 (2014)
- [Hashimoto 15] Hashimoto, K., Stenetorp, P., Miwa, M., and Tsuruoka, Y.: Task-Oriented Learning of Word Embeddings for Semantic Relation Classification, in *Proceedings of the 19th Conference on Computational Natural Language Learning (CoNLL 2015)*, pp. 268–278 (2015)
- [Hendrickx 10] Hendrickx, I., Kim, S. N., Kozareva, Z., Nakov, P., Ó Séaghdha, D., Padó, S., Pennacchiotti, M., Romano, L., and Szpakowicz, S.: SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations between Pairs of Nominals, in *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 33–38 (2010)
- [Kiros 15] Kiros, R., Zhu, Y., Salakhutdinov, R., Zemel, R. S., Torralba, A., Urtasun, R., and Fidler, S.: Skip-Thought Vectors, in *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, pp. 3276–3284 (2015)
- [Korkontzelos 13] Korkontzelos, I., Zesch, T., Zanzotto, F. M., and Biemann, C.: SemEval-2013 Task 5: Evaluating Phrasal Semantics, in *Second Joint Conference on Lexical and Computational Semantics (*SEM 2013)*, pp. 39–47 (2013)
- [Levy 14] Levy, O. and Goldberg, Y.: Neural word embedding as implicit matrix factorization, in *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, pp. 2177–2185 (2014)
- [Lin 01] Lin, D. and Pantel, P.: DIRT – Discovery of Inference Rules from Text, in *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 01)*, pp. 323–328 (2001)
- [Mikolov 13] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J.: Distributed Representations of Words and Phrases and their Compositionality, in *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pp. 3111–3119 (2013)
- [Mitchell 10] Mitchell, J. and Lapata, M.: Composition in Distributional Models of Semantics, *Cognitive Science*, Vol. 34, No. 8, pp. 1388–1439 (2010)
- [Miwa 16] Miwa, M. and Bansal, M.: End-to-End Relation Extraction using LSTMs on Sequences and Tree Structures, in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016)*, pp. 1105–1116 (2016)
- [Muraoka 14] Muraoka, M., Shimaoka, S., Yamamoto, K., Watanabe, Y., Okazaki, N., and Inui, K.: Finding The Best Model Among Representative Compositional Models, in *Proceedings of the 28th Pacific Asia Conference on Language, Information, and Computation (PACLIC 28)*, pp. 65–74 (2014)
- [Nakashole 12] Nakashole, N., Weikum, G., and Suchanek, F.: PATTY: A Taxonomy of Relational Patterns with Semantic Types, in *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL 2012)*, pp. 1135–1145 (2012)
- [Pennington 14] Pennington, J., Socher, R., and Manning, C.: Glove: Global Vectors for Word Representation, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, pp. 1532–1543 (2014)
- [Riedel 13] Riedel, S., Yao, L., McCallum, A., and Marlin, B. M.: Relation Extraction with Matrix Factorization and Universal Schemas, in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2013)*, pp. 74–84 (2013)
- [Rink 10] Rink, B. and Harabagiu, S.: UTD: Classifying Semantic Relations by Combining Lexical and Semantic Resources, in *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 256–259 (2010)
- [Santos 15] Santos, C. d., Xiang, B., and Zhou, B.: Classifying Relations by Ranking with Convolutional Neural Networks, in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2015)*, pp. 626–634 (2015)
- [Socher 11] Socher, R., Lin, C. C.-Y., Ng, A. Y., and Manning, C. D.: Parsing Natural Scenes and Natural Language with Recursive Neural Networks, in *Proceedings of the 28th International Conference on Machine Learning (ICML 2011)*, pp. 129–136 (2011)
- [Socher 12] Socher, R., Huval, B., Manning, C. D., and Ng, A. Y.: Semantic Compositionality through Recursive Matrix-Vector Spaces, in *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL 2012)*, pp. 1201–1211 (2012)
- [Socher 13] Socher, R., Bauer, J., Manning, C. D., and Andrew Y., N.: Parsing with Compositional Vector Grammars, in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)*, pp. 455–465 (2013)
- [Sutskever 11] Sutskever, I., Martens, J., and Hinton, G.: Generating Text with Recurrent Neural Networks, in *Proceedings of the 28th International Conference on Machine Learning (ICML 2011)*, pp. 1017–1024 (2011)
- [Sutskever 14] Sutskever, I., Vinyals, O., and Le, Q. V.: Sequence to Sequence Learning with Neural Networks, in *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, pp. 3104–3112 (2014)
- [Tai 15] Tai, K. S., Socher, R., and Manning, C. D.: Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks, in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2015)*, pp. 1556–1566 (2015)
- [Takase 16a] Takase, S., Okazaki, N., and Inui, K.: Composing Distributed Representations of Relational Patterns, in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016)*, pp. 2276–2286 (2016)
- [Takase 16b] Takase, S., Okazaki, N., and Inui, K.: Modeling semantic compositionality of relational patterns, *Engineering Applications of Artificial Intelligence*, Vol. 50, No. C, pp. 256–264 (2016)
- [Tang 15] Tang, D., Qin, B., and Liu, T.: Document Modeling with Gated Recurrent Neural Network for Sentiment Classification, in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015)*, pp. 1422–1432 (2015)
- [Toutanova 15] Toutanova, K., Chen, D., Pantel, P., Poon, H., Choudhury, P., and Gamon, M.: Representing Text for Joint Embedding of Text and Knowledge Bases, in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015)*, pp. 1499–1509 (2015)
- [Wang 16] Wang, L., Cao, Z., Melo, de G., and Liu, Z.: Relation Classification via Multi-Level Attention CNNs, in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016)*, pp. 1298–1307 (2016)
- [Wieting 15] Wieting, J., Bansal, M., Gimpel, K., and Livescu, K.: From Paraphrase Database to Compositional Paraphrase Model and Back, *Transactions of the Association for Computational Linguistics (TACL 2015)*, Vol. 3, pp. 345–358 (2015)
- [Xu 15] Xu, K., Feng, Y., Huang, S., and Zhao, D.: Semantic Relation Classification via Convolutional Neural Networks with Simple Negative Sampling, in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015)*, pp. 536–540 (2015)
- [Yih 14] Yih, W., He, X., and Meek, C.: Semantic Parsing for Single-Relation Question Answering, in *Proceedings of the 52nd An-*

nual Meeting of the Association for Computational Linguistics (ACL 2014), pp. 643–648 (2014)

[Zeichner 12] Zeichner, N., Berant, J., and Dagan, I.: Crowdsourcing Inference-Rule Evaluation, in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012)*, pp. 156–160 (2012)

〔担当委員：谷口 忠大〕

2016 年 9 月 30 日 受理

著 者 紹 介

高瀬 翔

2017 年東北大学大学院情報科学研究科博士後期課程修了。同年より、NTT コミュニケーション科学基礎研究所にて自然言語処理の研究に従事。博士（情報科学）。ACL 会員。

岡崎 直観(正会員)

2007 年東京大学大学院情報理工学系研究科・電子情報学専攻博士課程修了。東京大学院情報理工学系研究科特別研究員を経て、2011 年より東北大学大学院情報科学研究科准教授。自然言語処理、テキストマイニングの研究に従事。博士（情報理工学）。情報処理学会、ACL 各会員。

乾 健太郎(正会員)

1995 年東京工業大学大学院情報理工学研究科博士課程修了。東京工業大学大学院理工学研究科助手、九州工業大学助教授、奈良先端科学技術大学院大学助教授を経て、2010 年より東北大学大学院情報科学研究科教授、現在に至る。博士（工学）。自然言語処理の研究に従事。情報処理学会、ACL、AAAI 各会員。