DEIM Forum 2019 B5-4

ゼロショット学習によるテキストからのエンティティプロパティ同定

Wiradee Imrattanatrai† 加藤 誠†† 吉川 正俊†

† 京都大学大学院情報学研究科
†† 京都大学国際高等教育院／ JST さきがけ
E-mail: †wiradee@dl.kuis.kyoto-u.ac.jp, ††mpkato@acm.org, †††yoshikawa@i.kyoto-u.ac.jp

あらまし 本研究では，テキストに記述されるエンティティのプロパティを抽出する方法を提案する．プロパティ抽出は文の関係分類と似た問題であり，各プロパティについて訓練データを用意することで教師あり学習により実現できる．しかしながら，すべてのプロパティに対して訓練データを用意することは現実的ではないため，各プロパティについて訓練データを必要としない，ゼロショット学習による方法を提案する．我々は，プロパティを知識ベースのグラフ構造から得られる埋め込みによって表現し，ディープニューラルネットワークモデルと組み合わせることによって，訓練データのないプロパティの抽出を可能とした．実験では，独自データセットおよび公開データセットにて，訓練データがある場合とない場合を比較し，訓練データがないプロパティであっても訓練データのあるプロパティに匹敵する精度で抽出できることを明らかにした．

キーワード ゼロショット学習，深層学習，関係分類

1. はじめに

多くのエンティティ関連のクエリ，および，Web において急速に増加するエンティティ情報に対応するために，近年さまざま新機能が検索エンジンに導入されてきている [6, 17, 26] ．最も成功している機能の 1 つとして，エンティティカードが挙げられる．これは，クエリに関連するエンティティの情報，例えば，エンティティの簡潔な説明やファクト等を，検索結果とともに表示する機能である．特に，ユーザの情報要求に合致するようなエンティティ情報を含む場合には，エンティティカードはユーザの興味を引き有用であることが示唆されている [2].

ユーザの情報要求に対して適合するようなエンティティの情報を提供するために，エンティティのプロパティ（例えば，エンティティ「テイラー・スウィフト」の「アルバム」や「ジャンル」，エンティティ「スタンフォード大学」の「学校区分」や「キャンパス」など）を，文書や文などのテキストから抽出することは重要なタスクである．このタスクでは，例えば,「2012 album *Red* took Taylor Swift's popularity to new levels and the universal appeal of *I Knew You Were Trouble* was a key part of that success」という文から，エンティティ「テイラー・スウィフト」のプロパティ「アルバム」と「トラック」を特定することを目的とする．

プロパティの特定は適合するエンティティ情報への効率的なアクセスを可能にする．ユーザにクリックされた文書や検索結果の上位の文書に基づいて適合するエンティティのプロパティを推定することが可能であり，より適したエンティティカードの提供や追加の適合文書を提示することに利用できる．例えば，エンティティ「テイラー・スウィフト」の「アルバム」や「トラック」に関する文書をユーザがクリックしたのであれば，そのユーザはこれらのプロパティに関心があると推定可能であり，これらのプロパティについてのエンティティカードを用意したり，これらのプロパティについて書かれた別の文書を提示することもできる．上記に加え，各文書が含むプロパティを検索結果中に表示させ，ユーザの効率的な検索結果閲覧を助けることもできる．例えばユーザが「Taylor Swift career highlights」というクエリで検索した場合，各検索結果に含まれるプロパティを検索結果中に提示できれば，適合する文書を選ぶのに役立つはずである．したがって，テキストからのプロパティ同定タスクはエンティティ指向の検索において中心的なタスクであり，様々な応用が考えられる．

そこで，本論文では，テキストからエンティティのプロパティを同定する問題に取り組む．この問題は関係抽出タスクと類似している．関係抽出タスクでは，遠距離教師あり学習によって文分類を行う方法が主流である [9, 15, 20, 24]．この方法では，知識ベース中のファクト（主語，述語，目的語により構成される 3 つ組）を表現する文を教師データとして分類器を学習させる．ある種のプロパティは知識ベース中の 1 つの述語に対応しており（例えば，エンティティ「テイラー・スウィフト」のプロパティ**album** は *music.artist.album* という述語に対応する），もしある文がその述語の主語と目的語を含めば正例として訓練データに用いられる．一方で，プロパティが述語パス，接続された述語集合，に対応することもある．例えば，プロパティ**track** は *music.artist.track_contributions* と *music.track_contribution.track* という 2 つの述語によって知識ベース中で表現される．この種のプロパティの場合は，述語パスを構成するファクト集合を用いて正例が特定される．この遠距離教師あり学習のアプローチは人手によるラベルづけが不要であり，多くの訓練データを用意するのに有用である．

しかしながら，訓練データ獲得の成否は知識ベース中のファクトの表現に大きく依存し，もしあるプロパティに関連するファクトが少なければ，ほとんど訓練用の文が得られないような場合も起こりうる．また，知識ベースには非常に多くの述語パス

が存在し，それらすべてに対して訓練データが得られることを前提とするのは現実的ではない．教師あり学習に頼る場合には，あるプロパティに対する訓練データが得られなければ，そのプロパティに対応するテキストパターンを学習することができないこととなる．この問題はゼロショット学習問題として知られており，これまでに関係抽出タスクにおいて取り組まれてこなかった問題である．特に，応用上，特定できるプロパティの網羅性は非常に重要であり，ユーザが必要とする情報を取りこぼさないためにも解決しなければならない問題である．

本論文では，テキストからプロパティを特定する問題を文のマルチラベル問題として扱い，ニューラルネットワークに基づくモデルを提案する．このモデルは知識ベースのグラフ埋め込み（例えば，TransE [1] など）を活用することによってゼロショット学習問題に対応することができる．また，我々は知識ベースのグラフ埋め込みに基づいた様々なプロパティ表現の方法についても提案を行う．実験では，新規に構築したデータセット，および，既存の関係抽出用のデータセットを用い，既存手法と提案手法の比較を行った．実験結果から，我々の提案手法が既存手法よりもより多くのプロパティを特定できることを示した．これに加え，訓練用の文が与えられていないようなプロパティも扱うことが可能であり，その性能は訓練用の文が与えられた場合に匹敵するものであった．

本論文における我々の貢献は下記の 3 点である．

- テキスト中のプロパティを同定する問題を文のマルチラベル問題として扱い，ニューラルネットワークに基づくモデルを提案した．
- ゼロショット学習問題へも対応できるように，知識ベースのグラフ埋め込みから構築されたプロパティの埋め込みを利用する方法を提案した．
- 既存のデータセットよりも多くのプロパティを含む，新規の文分類用データセットを構築し提案モデルの評価を行った．

2. 関 連 研 究

本節では，関係抽出とゼロショット学習問題に関する既存研究について述べる．

関係抽出は一般的に文分類問題として扱うことが可能であり，文をいずれかのプロパティに分類する問題である．この問題に対する初期のアプローチでは，人手によってラベルづけされた文を用いた教師あり学習手法を用いることであった [3, 4, 7, 16, 27]．このアプローチは高い適合率と再現率を達成できるものの，大量のラベルづけされた文を人手によって用意するのは困難であった．特に多くのプロパティが考慮されるべき状況においては，各プロパティに対して十分な量のデータを用意することができない．この問題を克服するために，遠距離教師あり学習が提案され最近の関係抽出研究で広く用いられている（ [9, 15, 20, 24] など）．遠距離教師あり学習では，知識ベース中のファクトに基づき，文とプロパティの対応づけが自動的に行われる．

遠距離教師あり学習によって，多くの種類のプロパティに対して大量の教師データを用意できるようになったものの，すべてのプロパティに対してこのアプローチが有効なわけではない．多くの既存手法はファクトを構成するエンティティ対の存在を仮定しているが，ファクトは日付や数量などの非エンティティからも構成されており，これらのファクトについては単に対象外とされてきている．また，非エンティティから構成されるファクトに対しては，実際に訓練データを発見することが困難であるため，前述のゼロショット学習問題が顕在化する．

ゼロショット学習は特に画像分類や物体認識問題などにおいて注目を集めてきている．最近では，ゼロショット学習に対して既知クラスおよび未知クラスの埋め込みを利用する方法が多く提案されており，これらの埋め込みはクラス属性 [11] や単語ベクトル [5, 22, 25]，テキスト情報 [12, 19] などから学習される．これらの埋め込みは訓練時に全く事例が与えられなかったクラスへの分類を行う際に用いられることになる．

我々の研究ではゼロショット学習問題に対応するために，クラス埋め込みの考え方を利用する．特に重要な課題として，プロパティをどのように表現するかという問題が挙げられる．知識ベースから得られるグラフ中において，プロパティはパスとして表現されるため，グラフ埋め込みによって得られるノードやエッジの埋め込みからどのようにプロパティを表現すべきかは明確ではない．そのため，我々は複数のプロパティ表現方法を提案し，実験によって適切なプロパティ表現を明らかにする．

3. 提 案 手 法

本節では，プロパティの埋め込みを利用することによってゼロショット学習問題に対応した，ニューラルネットワークに基づくマルチラベル分類モデルを示す．

3.1 問 題 設 定

プロパティ集合 Y を，文が分類されるクラスの集合であるとする．我々の問題は，各プロパティ $y \in Y$ がどの程度，あるエンティティ e を含む文 X によって表現されているかを推定する問題である．学習時には，いくつかのプロパティに対してそれらを表現する文が与えられており，これらのプロパティを既知プロパティと呼び，$Y_{\text{seen}} \subset Y$ と表現する．テスト時には，既知プロパティ Y_{seen} と未知プロパティ $Y_{\text{unseen}} \subset Y$ の両方を表すような文が分類対象となる．ここで，$Y_{\text{unseen}} \cap Y_{\text{seen}} = \emptyset$ である．したがって，訓練時に利用できるのは一部のクラス（すなわち，$Y_{\text{train}} = Y_{\text{seen}}$）の事例のみであり，テスト時には訓練時には事例が与えられなかったクラスを含む全クラス（すなわち，$Y_{\text{test}} = Y_{\text{seen}} \cup Y_{\text{unseen}}$）の事例を分類しなくてはならない．これはまさにゼロショット学習問題の設定と符合する．

3.2 モ デ ル

図 1 に示す通り，我々の提案モデルは下記の 5 つの主要なコンポーネントから構成されている．すなわち，(1) 低次元の単語ベクトルへ文中の単語を変換するための単語埋め込み，(2) 文単位の表現を生成するための両方向長期短期記憶 (BLSTM)，(3) 文表現からプロパティの埋め込みへの写像，(4) エンティティの型に基づく予測プロパティの絞り込み，(5) 各プロパティが表現されている確率の計算，である．

もっとも重要なアイデアは，ある文の表現がその文で表現されているプロパティの埋め込みと似るように，文表現からプロ

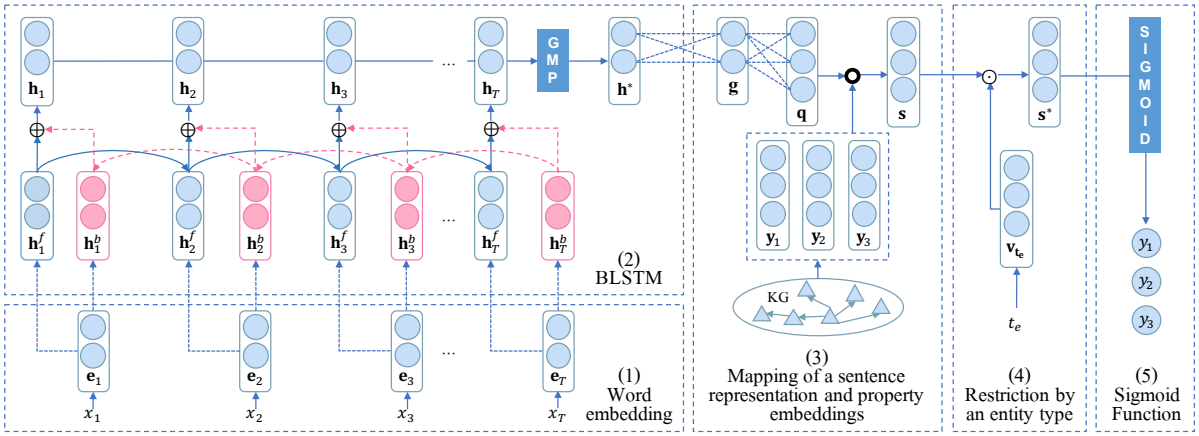

図 1　提案モデルの構造.

パティ表現への写像を学習することである．これが実現できれば，プロパティの埋め込みさえ用意できていれば，プロパティごとに訓練データを用意する必要はなくなり，ゼロショット学習問題に対応することができる．

以下の小節では，提案モデルの詳細について述べていく．

3.2.1　単語埋め込み

提案モデルの最初のコンポーネントは，入力文中の各単語を語の意味を捉えた単語ベクトルに写像する機能を有する．T 個の単語 $X = \{x_1, x_2, \ldots, x_T\}$ を含む文が与えられた時，語 x_t はその One-hot 表現である，$|V|$ 次元ベクトル $\mathbf{v}_t$ によって表現される．ここで，$\mathbf{v}_t$ は，語 x_t のインデックスに対応する次元は 1，それ以外が 0 であるようなベクトルであり，V は全語彙の集合である．各 One-hot ベクトル $\mathbf{v}_t$ は単語埋め込み行列 $\mathbf{E_w} \in \mathbb{R}^{d_e \times |V|}$ により単語ベクトル $\mathbf{e}_t$ に変換される：$\mathbf{e}_t = \mathbf{E_w}\mathbf{v}_t$．ただし，$d_e$ は単語ベクトルの次元を表す．

3.2.2　両方向長期短期記憶

文中の単語系列を扱うために，我々は両方向長期短期記憶 (BLSTM) を 2 つ目のコンポーネントとして採用している．長期短期記憶 (LSTM) は入力系列の長期的な依存関係を扱うために設計された再帰型ニューラルネットワークの一種である [8]．LSTM は入力，忘却，出力の 3 種類のゲートとメモリセルから構成されている．文 X から得られた単語埋め込み $\{\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e}_T\}$ が与えられた時，各ステップ t において単語埋め込み $\mathbf{e}_t$，および，前の隠れ状態 $\mathbf{h}_{t-1}$ が LSTM ユニットに入力される．LSTM ユニットの隠れ状態 $\mathbf{h}_t$ は以下のようにして得られる：

$$\mathbf{i}_t = \sigma\left(\mathbf{W_i}\left[\mathbf{h}_{t-1}, \mathbf{e}_t\right] + \mathbf{b_i}\right) \quad (1)$$

$$\mathbf{f}_t = \sigma\left(\mathbf{W_f}\left[\mathbf{h}_{t-1}, \mathbf{e}_t\right] + \mathbf{b_f}\right) \quad (2)$$

$$\mathbf{o}_t = \sigma\left(\mathbf{W_o}\left[\mathbf{h}_{t-1}, \mathbf{e}_t\right] + \mathbf{b_o}\right) \quad (3)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tanh\left(\mathbf{W_c}\left[\mathbf{h}_{t-1}, \mathbf{e}_t\right] + \mathbf{b_c}\right) \quad (4)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh\left(\mathbf{c}_t\right) \quad (5)$$

ここで，σ はシグモイド関数，tanh はハイパボリックタンジェント関数，$\odot$ は要素ごとの積，$\mathbf{W_i}$，$\mathbf{W_f}$，$\mathbf{W_o}$，$\mathbf{W_c} \in \mathbb{R}^{d_h \times (d_e + d_h)}$，および，$\mathbf{b_i}$，$\mathbf{b_f}$，$\mathbf{b_o}$，$\mathbf{b_c} \in \mathbb{R}^{d_h}$ はモデルのパラメータである．d_e と d_h は単語ベクトルの次元，ならびに，隠れ状態の次元である．

単方向の再帰型ニューラルネットワークを用いた場合，入力系列が順方向で与えられるため，入力系列の先頭付近の情報は LSTM ユニットの最後の隠れ状態に反映されにくい．最後の隠れ状態を入力系列の表現として用いる際にはこのことが問題となる．この問題を解決するために両方向再帰型ニューラルネットワークが提案されている [21]．両方向再帰型ニューラルネットワークを用いる場合，前向き，および，後ろ向き方向の入力系列を扱うために，再帰型ニューラルネットワーク層がもう 1 層追加されることになる．本研究では，両方向再帰型ニューラルネットワークとして BLSTM を用いることにする．各ステップ t における BLSTM の隠れ状態は前向き，および，後ろ向きの層の隠れ状態（それぞれ，$\mathbf{h}_t^f$，$\mathbf{h}_t^b$ とする）同士の要素ごとの積によって得られる：

$$\mathbf{h}_t = \mathbf{h}_t^f \oplus \mathbf{h}_t^b \quad (6)$$

コンポーネントの最後にて，すべての隠れ状態 $\mathbf{h}_t$ に対して大域的な最大プーリングを適用することによって，系列において発見された顕著な語のパターンのみを獲得する．大域的な最大プーリングは以下の式によって定義される：

$$h_i^* = \max_t \left\{\mathbf{h}_t(i)\right\} \quad (7)$$

ただし，h_i^* は，このコンポーネントの出力である $\mathbf{h}^*$ の i 次元目の値であり，$\mathbf{h}_t(i)$ は $\mathbf{h}_t$ の i 次元目の値である．

3.2.3　プロパティ埋め込みと文表現の対応づけ

3 番目のコンポーネントは，プロパティ埋め込みと文表現の対応づけを行うことによって，ゼロショット学習を実現する機能を備えている．プロパティ埋め込みについては 3.4 節にて詳しく述べる．モデルが未知のプロパティに対しても分類を行えるようにするため，まず前のコンポーネントの出力を非線型変換を行う全結合層によって射影する：

$$\mathbf{g} = \mathrm{ReLU}(\mathbf{W_g}\mathbf{h}^* + \mathbf{b_g}) \quad (8)$$

ただし，ReLU は Rectified Linear Unit であり，$\mathbf{W_g} \in \mathbb{R}^{d_h \times d_h}$

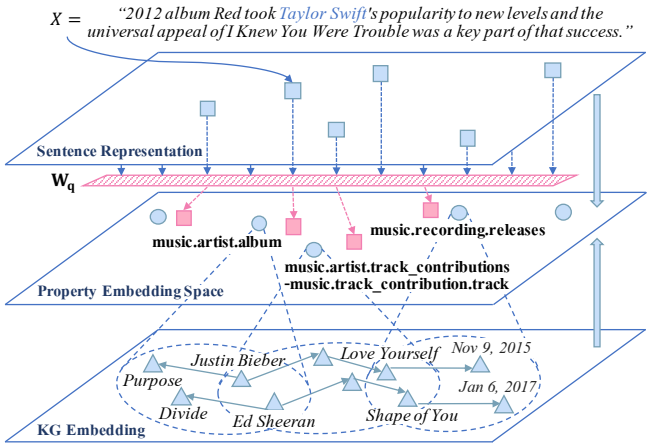

図 2　文表現とプロパティ埋め込みの対応づけ．

と $\mathbf{b_g} \in \mathbb{R}^{d_h}$ は全結合層のパラメータである．

次に，d_h 次元の文表現を d_y 次元のプロパティ埋め込み空間へ射影する：

$$\mathbf{q} = \mathbf{W_q g} + \mathbf{b_q} \tag{9}$$

ただし，$\mathbf{W_q} \in \mathbb{R}^{d_y \times d_h}$ と $\mathbf{b_q} \in \mathbb{R}^{d_y}$ はこの射影のパラメータである．

射影された文表現がどのプロパティに対応するかを定量化するために，射影された文表現とプロパティ埋め込みの類似度を内積によって求める．各行がプロパティの埋め込みを表すプロパティ埋め込み行列 $\mathbf{E_y} \in \mathbb{R}^{|Y'| \times d_y}$（$Y'$ は訓練時には Y_{train}，テスト時には Y_{test} が用いられる），および，射影された文表現 $\mathbf{q}$ が与えられた時，これらの類似度ベクトル $\mathbf{s}$ は以下のようにして得られる：

$$\mathbf{s} = \mathbf{E_y q} \tag{10}$$

図 2 に示すように，文中で表現されるプロパティの埋め込みがその文の表現と近くなるように学習された変換行列 $\mathbf{W_q}$ によって，各文表現はプロパティ埋め込み空間に射影される．結果として，学習されたモデルでは，ある文中で表現されるプロパティとその文の類似度が高くなる．

3.2.4　エンティティの型に基づく予測プロパティの絞り込み

次のコンポーネントではエンティティの型を利用することで，妥当なプロパティを絞り込み，予測結果の向上を図る．文中に含まれるエンティティの型はその文が意味しうるプロパティ集合を規定する．例えば，文中のエンティティの型が「映画」であれば，「ジャンル」や「公開日」などが妥当なプロパティとなりうるが，「職業」や「出身地」などは文が表すプロパティとして適切ではない．どのプロパティがどの型に対応するかについては，知識ベースの構造を見れば明らかである．文 X 中に含まれるエンティティ e に対して，エンティティの型を t_e と表記する．そして，型 t_e は適切なプロパティ集合 Y_{t_e} との対応関係が取れているとする．このとき，エンティティ型ベクトル $\mathbf{v_{t_e}} \in \{0,1\}^{|Y'|}$ は，$y_i \in Y_{t_e}$ ならば i 番目の値が 1，そうでなければ 0 であるようなベクトルとして定義できる．ただし，前述の通り，Y' は訓練時には Y_{train}，テスト時には Y_{test} であるようなプロパティ集合である．このエンティティ型ベクトルは前のコンポーネントの出力と要素ごとの積を取ることによって，妥当なプロパティを絞り込む：

$$\mathbf{s}^* = \mathbf{s} \odot \mathbf{v_{t_e}} \tag{11}$$

結果として，妥当なプロパティ以外の類似度が強制的に 0 になる．

3.2.5　出　　力

最後のコンポーネントによって，ある文が各プロパティをどのぐらいの確率で表現しているかを得る．これは，前のコンポーネントの出力の要素ごとにシグモイド関数を適用することにより達成することができる．

$$P(y_i|X) = \sigma(s_i^*) \tag{12}$$

ただし，s_i^* は $\mathbf{s}^*$ の i 番目の値である．

3.3　パラメータ学習

モデルの学習では，交差エントロピーを目的関数として用いる．y，および，$\hat{y}$ をそれぞれ正解プロパティを表すベクトル，予測プロパティの予測確率を表すベクトルとする．我々は目的関数を以下のように定義した：

$$L_{CLF} = -\sum_i \sum_j y_{i,j} \log \hat{y}_{i,j} \tag{13}$$

ここで，i は訓練データ中の文のインデックス，j はプロパティのインデックスである．

パラメータの最適化には Adam [10] を用いる．Adam は確率的勾配降下法の拡張であり，他の最適化方法と比べ優れていることが示されている．

3.4　プロパティ埋め込み

ここまで，モデルの各コンポーネントを説明し，提案モデルがゼロショット学習問題を克服するためにどのようにプロパティ埋め込みを利用して動作するかを説明してきた．本節では，プロパティ埋め込みをどのように構成するかについて述べる．まず，述語パスに対応するプロパティの特徴について議論を行い，TransE という知識ベース埋め込み手法に基づいて，プロパティ埋め込みを構成するいくつかの方法について説明する．

3.4.1　プロパティの特徴

エンティティのプロパティはファクトにより構成される知識ベースの構造によって定義される．各ファクトは，(s, r, o) の 3 つ組によって表現され，それぞれ，主語，述語，目的語を表す．知識ベースはグラフとしても表現され，主語と目的語はノード，述語はエッジに対応する．あるファクト集合 $\mathcal{K}$ が与えられた時，任意の主語と述語間に存在する述語パスにより構成される 3 つ組の集合を以下のように表すことができる：$\mathcal{P} = \{(s, \mathbb{r}, o) | (s, r_1, x_1) \in \mathcal{K} \wedge (x_1, r_2, x_2) \in \mathcal{K} \wedge \cdots \wedge (x_n, r_n, o) \in \mathcal{K}\}$．ここで，述語パス $\mathbb{r} = (r_1, r_2, \cdots, r_n)$ は主語 s と目的語 o のパス上に存在する述語の列であり，x_i はこのパス上の主語や目的語を表現し，n は述語パスの長さを表す．本論文では，$\mathcal{P}$ 中の主語をエンティティとし，述語列からなる述語パスをプロパティと定義する．

3.4.2 TransE による知識ベース埋め込み

TransE は知識ベース $\mathcal{K}$ 中のファクトの構成要素である主語，述語，目的語を表す，低次元空間への埋め込みを学習するモデルである．ファクト $(s,r,o) \in \mathcal{K}$ が与えられた時，翻訳ベクトル $\mathbf{r}$ によって表現される述語は主語の埋め込み $\mathbf{s}$ と目的語の埋め込み $\mathbf{o}$ の間をつなぐ役目を果たし，$\mathbf{s}+\mathbf{r} \simeq \mathbf{o}$ を満たすように学習される．そのような埋め込みは以下のマージンに基づくランキング損失を最小化することで求められる：

$$L_{\mathrm{KG}} = \sum_{(s,r,o)\in\mathcal{K}} \sum_{(s',r,o')\in\mathcal{K}'} [\gamma + d(\mathbf{s}+\mathbf{r},\mathbf{o}) - d(\mathbf{s}'+\mathbf{r},\mathbf{o}')]_+ \tag{14}$$

ただし，$\mathcal{K}'$ は $\mathcal{K}$ の主語か目的語を無作為に他の主語または目的語に置き換えたファクト集合であり，d は非類似度，γ はマージンパラメータ，$[x]_+$ は x が正のときは x で，負のときは 0 である．

3.4.3 プロパティ埋め込みの表現

各プロパティの埋め込みを得るために，我々は知識ベース埋め込みによって得られた，主語，述語，目的語の埋め込みを利用する．プロパティが意味的に似ていればその埋め込みも近くなり，また，逆も成り立つようなプロパティ表現が期待される．プロパティをそれに対応する述語パスに含まれる述語の埋め込みによって単純に表現することも可能ではあるが，プロパティをより良く特徴づけるためには主語や目的語の埋め込みも重要ではないかと仮定し，いくつかのプロパティ埋め込み構成方法を以下では説明する．プロパティ $y \in Y$ とそれに対応する述語パス $\mathbb{r}$ が与えられたとき，プロパティ y の埋め込み e_y を以下の 4 つの方法で表現する．

- **SUM-R**: 述語パスに含まれる述語の埋め込みの和，$\sum_{i=1}^{n} \mathbf{r}_i$．例えば，プロパティ **music.artist.track_contributions–music.track_contribution.track** の埋め込みは述語 *music.artist.track_contributions*，および，*music.track_contribution.track* の埋め込みの和で構成される．
- **CC-SO**: 主語と目的語の埋め込みを平均し結合，$[\bar{\mathbf{s}},\bar{\mathbf{o}}]$．例えば，主語「ジャスティン・ビーバー」と「エド・シーラン」の埋め込みと，それらの楽曲であるような目的語「Love Yourself」と「Shape of You」の埋め込みの平均をそれぞれに対して求め，結合することによって，プロパティ **music.artist.track_contributions–music.track_contribution.track** の埋め込みを表現する．
- **CC-RO**: 述語パスの最初の述語の埋め込みと目的語の埋め込みの平均を結合，$[\mathbf{r_1},\bar{\mathbf{o}}]$．述語パスの最初の述語は，述語パスの中で最も述語パス全体の意味を包含していると考えたため，このプロパティ埋め込み方法を提案している．
- **CC-SON**: 主語，目的語，および，述語パス中の最後の述語に隣接する述語の，目的語の平均をそれぞれ求め結合，$[\bar{\mathbf{s}},\bar{\mathbf{o}},\bar{\mathbf{n}}]$．この方法は 2 つ目の構成方法に似ているが，述語の埋め込みが適切に学習されなかった場合でもプロパティを頑健に特徴づけられるようにするために，述語パス中の最後の述語と主語を共有する述語の目的語を利用した．

これらの方法で用いられている要素は以下のように定義される：$\bar{\mathbf{s}} = \frac{\sum_{s\in S_{\mathbb{r}}} \mathbf{s}}{|S_{\mathbb{r}}|}$，$\bar{\mathbf{o}} = \frac{\sum_{o\in \mathcal{O}_{\mathbb{r}}} \mathbf{o}}{|\mathcal{O}_{\mathbb{r}}|}$，$\bar{\mathbf{n}} = \frac{\sum_{o\in N_{\mathbb{r}}} \mathbf{o}}{|N_{\mathbb{r}}|}$ if $n>1$, otherwise $\mathbf{0}$．ただし，$S_{\mathbb{r}} = \{s \mid (s,\mathbb{r},o) \in \mathcal{P}\}$, $\mathcal{O}_{\mathbb{r}} = \{o \mid (s,\mathbb{r},o) \in \mathcal{P}\}$, $N_{\mathbb{r}} = \{o \mid (s,(r_1,r_2,\cdots,r_{n-1},\tilde{r}_n),o) \in \mathcal{P}\}$, $\tilde{r}_n$ は述語 r_n に隣接する述語を表し，n は述語パスの長さを表す．

4. 実　　験

実験では提案手法に関する下記の疑問に答えることを目的とした．

- **RQ1**: ベースライン手法と比較して，提案モデルはゼロショット学習問題に対して対処可能であるか．
- **RQ2**: どのプロパティ埋め込みの構成方法が最も適しているか．
- **RQ3**: どのような状況において，提案モデルがゼロショット学習問題に対して有効に働くのか．

以下では，まず実験で用いたデータセット，および，評価指標について説明する．次に実験設定を示し，実験結果について議論を行う．

4.1 データセット

実験では 2 種類のデータセットを用いた．1 つ目のデータセットは，Riedel ら [20] によって公開された NYT10 データセットであり，New York Times コーパスに対して知識ベース Freebase のファクトを対応づけて生成されたデータセットである．このデータセットには **54** 種類のプロパティが含まれている．しかしながら，我々はプロパティの同定に高い網羅性を求めているため，多くのプロパティを含むようなデータセットが必要であった．加えて，既存のデータセットに含まれる少ないプロパティでは，ゼロショット学習問題に対して高い有効性を示せない可能性があった．そのため，我々は **271** プロパティを含む独自のデータセット，WEB19 を構築した．以下では，このデータセットの構築方法について述べる．

4.1.1 WEB19 データセット

このデータセットを構築するために，まず FB15k データセットに含まれる述語パスに基づいてプロパティ集合を選定した．FB15k データセットは Bordes らによって公開され，知識ベース埋め込みを生成するに用いられてきた [1]．我々は Freebase(注1) 中で 100 ファクト以上と対応しているような述語パスをプロパティとして選定した．

次に，各プロパティについて最大で 500 件の主語と目的語のペアを得て，Web 検索エンジン Bing を用いてこれらを含むような文を獲得した．これによって，各プロパティの正例であるような文を得ようとしたが，主語と目的語を含む文が必ずしもそれらのプロパティを表していないという問題がある．

上記の問題を解決するために，我々は Amazon Mechanical Turk を用いて，人手によって獲得した文のラベルづけを実施した．各文について 3 名のワーカを割り当て，各プロパティが文に含まれているかどうかを判定してもらった．ラベルづけの品質を確認するために，Free-Marginal Multirater Kappa [18]

(注1)：https://developers.google.com/freebase

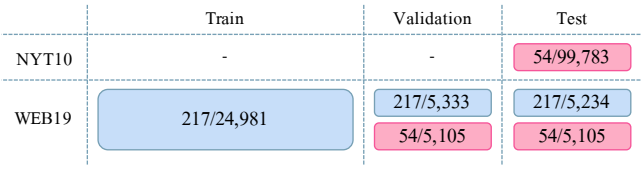

図 3　NYT10・WEB19 データセットにおける訓練，検証，テストデータのプロパティ数および文数．青と赤の四角はそれぞれ既知プロパティと未知プロパティを表す．

を計算したところ 0.602 となり，並みの一致度であることを確認した．過半数，すなわち，2 名以上が文に含まれていると判定したプロパティのみを採用した．

図 3 は 2 つのデータセットの分割方法，および，プロパティと文の数を図示している．NYT10 データセットではすべてのプロパティを未知であるとみなし，WEB19 データセットではランダムに選択されたプロパティを未知であるとみなし残りを既知のプロパティとみなした．WEB19 データセットにおいて，既知プロパティの文は訓練，検証，テストに分割し，未知プロパティの文は検証とテストに分割した．検証データはハイパーパラメータの決定に利用した．

4.2 評価指標

実験では，マルチラベル分類におけるサンプルに基づく評価指標とラベルに基づく評価指標を用いて評価を行った．

4.2.1 サンプルに基づく指標

マルチラベル分類の性能はサンプル単位（本研究の場合，文単位）の性能で評価することができる．具体的には，Y_s をある文 s に付与された正解プロパティの集合，$\hat{Y}_s$ を予測確率が 0.5 より高いような予測ラベルの集合として，下記の指標を文単位で計算する：精度 $\mathrm{A}_{\mathrm{sample}} = \frac{|Y_s \cap \hat{Y}_s|}{|Y_s \cup \hat{Y}_s|}$，適合率 $\mathrm{P}_{\mathrm{sample}} = \frac{|Y_s \cap \hat{Y}_s|}{|\hat{Y}_s|}$，再現率 $\mathrm{R}_{\mathrm{sample}} = \frac{|Y_s \cap \hat{Y}_s|}{|Y_s|}$，F 値 $\mathrm{F}_{\mathrm{sample}} = \frac{2|Y_s \cap \hat{Y}_s|}{|Y_s|+|\hat{Y}_s|}$．また，予測されたプロパティのうち，予測確率の高い順で上位 k 件に含まれるプロパティに，少なくとも 1 つの正解を含むような文の割合を hit@k として用いる．

4.2.2 ラベルに基づく指標

文単位の評価とは別に，各プロパティに対する予測性能を測るために，ラベルに基づく指標でも評価を行った．S_l をあるプロパティ l が付与された文の集合，$\hat{S}_l$ をプロパティ l が 0.5 より高いような予測確率で付与された文の集合として，ラベルに基づく F 値は $\mathrm{F}_{\mathrm{label}} = \frac{2|S_l \cap \hat{S}_l|}{|S_l|+|\hat{S}_l|}$ と計算できる．

4.3 実験設定

4.3.1 埋め込み

単語埋め込みを得るために，英語版 Wikipedia(注2) を用いてスキップグラムモデルを学習した [14]．単語埋め込みの次元数として 100, 200, 300 を用い，検証データにて最適な値を決定し評価を行った．知識ベースの埋め込みを得るために，我々は Bordes らによって公開されている FB15k データセットを用いた [1]．

4.3.2 比較手法

実験では提案するニューラルネットワークによるモデルとベースライン手法を比較した．ベースライン手法を下記に示す：

- **Rand と Rand-T**: これらの手法では n 個のプロパティを各文に対してランダムに選択する．ただし，Rand では全プロパティを用い，Rand-T は文中のエンティティのエンティティタイプによってプロパティを限定し，その上でランダムなプロパティ選択を行っている．プロパティ数 n は WEB19 におけるプロパティ数の分布を利用してランダムに選択した．
- **RNN**: RNN モデル [23].
- **CNN**: CNN モデル [28].
- **BRNN**: Bidirectional RNN モデル [29].
- **LSTM**: LSTM モデル.
- **BLSTM**: Bidirectional LSTM モデル.

4.4 実験結果

4.4.1 RQ1: ベースライン手法と比較して，提案モデルはゼロショット学習問題に対して対処可能であるか

表 1 はベースライン手法と提案手法を平均精度，適合率，再現率，F 値，hit@k ($k = 1$) によって評価した結果を示している．提案手法はいくつかの変種を評価しており，各手法は，① エンティティタイプを利用しているか，② プロパティ埋め込みを利用しているか，③ プロパティ埋め込みの構成方法，によって特徴付けられている．

RQ1 に回答するために，未知プロパティの性能に着目して考察する．NYT10 データセットでは提案手法 #7，プロパティ埋め込みを CC-SO によって構成した手法が最も高い精度，適合率，再現率，F 値を示しており，提案手法 #9，プロパティ埋め込みを CC-SON によって構成した手法が最も高い hit@1 を示している．一方で，WEB19 データセットでは提案手法 #6，プロパティ埋め込みを SUM-R によって構成した手法が全ての指標で高い性能を示している．また，提案手法 #6 は既知のプロパティに匹敵する性能を未知のプロパティにおいて達成できていることが読み取れる．

ベースライン手法は既知プロパティしか出力できないため，未知プロパティに対してほとんどの評価指標は 0 となっている．一方で，プロパティ埋め込みを用いる提案手法 #2-9 は，既知プロパティにおける性能と比肩する性能を示すことからゼロショット学習問題に対処できていると言える．提案手法 #1 はプロパティ埋め込みを用いていないためゼロショット学習問題には対処できていないが，既知プロパティに対しては高い適合率と hit@1 を示している．また，提案手法 #2-5 と提案手法 #6-9 を比較することによって，エンティティタイプを利用することによって大きな改善が得られることがわかる．

4.4.2 RQ2: どのプロパティ埋め込みの構成方法が最も適しているか

表 1 から，NYT10 データセットでは提案手法 #7 と提案手法 #9 が，WEB19 データセットでは提案手法 #6 が全ての指標で高い性能を示していることがわかる．

最も優れた手法がデータセットによって異なっていたため，我々は適した手法が訓練データ中の既知プロパティの数に依存しているのではないかという仮説を立てた．各エンティティタイプごとの既知プロパティの数は NYT10 データセットでは

(注2)：https://dumps.wikimedia.org/enwiki

表 1　NYT10・WEB19 データセットにおける各プロパティ同定手法の実験結果．

	手法				NYT10 未知プロパティ					WEB19 未知プロパティ					WEB19 既知プロパティ				
					A_{sample}	P_{sample}	R_{sample}	F_{sample}	hit@1	A_{sample}	P_{sample}	R_{sample}	F_{sample}	hit@1	A_{sample}	P_{sample}	R_{sample}	F_{sample}	hit@1
ベースライン手法	Rand				0.002	0.004	0.006	0.004	0.003	0.005	0.005	0.006	0.005	0.008	0.005	0.006	0.007	0.006	0.007
	Rand-T				0.268	0.323	0.402	0.344	0.336	0.162	0.175	0.178	0.172	0.179	0.129	0.181	0.173	0.159	0.181
	RNN				0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.031	0.041	0.032	0.035	0.212
	CNN				0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.145	0.188	0.162	0.164	0.270
	LSTM				0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.026	0.137	0.170	0.167	0.157	0.294
	BRNN				0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.021	0.147	0.183	0.178	0.168	0.339
	BLSTM				0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.161	0.198	0.197	0.185	0.315
	#ID	①	②	③															
提案手法	#1	✓	-	-	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.022	0.331	**0.423**	0.373	0.374	**0.518**
	#2	-	✓	SUM-R	0.015	0.015	0.413	0.030	0.002	0.028	0.028	0.336	0.051	0.017	0.154	0.194	0.193	0.179	0.308
	#3	-	✓	CC-SO	0.021	0.021	0.467	0.041	0.001	0.136	0.142	0.234	0.164	0.229	0.169	0.206	0.227	0.199	0.306
	#4	-	✓	CC-RO	0.013	0.013	0.307	0.025	0.001	0.056	0.058	0.230	0.090	0.012	0.162	0.197	0.224	0.192	0.302
	#5	-	✓	CC-SON	0.022	0.022	0.343	0.040	0.004	0.101	0.107	0.170	0.121	0.234	0.171	0.207	0.231	0.201	0.317
	#6	✓	✓	SUM-R	0.328	0.328	0.498	0.369	0.325	**0.395**	**0.414**	**0.450**	**0.418**	**0.420**	0.322	0.402	0.377	0.366	0.416
	#7	✓	✓	CC-SO	**0.441**	**0.441**	**0.683**	**0.511**	0.433	0.283	0.298	0.337	0.306	0.297	0.321	0.401	0.390	0.370	0.422
	#8	✓	✓	CC-RO	0.400	0.400	0.677	0.466	0.398	0.354	0.369	0.429	0.380	0.372	0.323	0.402	0.387	0.370	0.421
	#9	✓	✓	CC-SON	0.435	0.435	0.656	0.502	**0.471**	0.282	0.295	0.343	0.306	0.297	**0.333**	0.416	**0.392**	**0.380**	0.435

0.919，WEB19 データセットでは 7.692 であった．多くの既知プロパティが存在する場合，プロパティ同士が区別できるようにプロパティの埋め込みは精密に行われる必要がある．一方で，既知プロパティが少ない場合，未知プロパティと類似するプロパティがなければ提案手法は適切に動作しないため，プロパティの埋め込みは比較的粗く行われる方が望ましい．提案手法のうち，SUM-R や CC-RO などの埋め込み構成手法は，述語の埋め込みを利用するためプロパティの埋め込みの粒度は細かく，一方で CC-SO や CC-SON などの手法は，主語と目的語を用い述語を用いないため，SUM-R や CC-RO などよりも粒度が粗くなる傾向がある．そのため，異なるデータセットにおいて適したプロパティ埋め込み方法が異なっていたのではないかと考えられる．

プロパティがどのように埋め込まれているかを視覚化するために，図 4 に，SUM-R によって得られた WEB19 のプロパティ埋め込みを t-SNE [13] によって 2 次元平面に写像した結果を示す．同じエンティティタイプの類似するプロパティが似た埋め込みで表現されていること（**produced_by**，**directed_by**，**written_by** など），また，異なるエンティティタイプであっても類似するプロパティであれば似た埋め込みで表現されていること（**origin** と **film_regional_debut_venue** など）が読み取れる．

4.4.3　RQ3: どのような状況において，提案モデルがゼロショット学習問題に対して有効に働くのか

図 5 に WEB19 データセットにおいて最も優れた性能を示した，提案手法 #6 の未知プロパティに対する相対 F 値（ベースライン手法 Rand-T を 1 としたときの F 値）を，同じエンティティタイプに属する既知プロパティの数と共に示す．この図から既知プロパティの数が増加すると共に，相対 F 値が緩やかに増加していっていることがわかる．したがって，同じエンティティタイプに属する既知プロパティの数がゼロショット学習問題を解く上で重要な要因であると考えられる．

5.　ま　と　め

本研究では，テキストに記述されるエンティティのプロパティを抽出する方法を提案した．特に，各プロパティについて訓練データを必要としない，ゼロショット学習による方法を提案した．我々は，プロパティを知識ベースのグラフ構造から得られる埋め込みによって表現し，ディープニューラルネットワークモデルと組み合わせることによって，訓練データのないプロパティの抽出を可能とした．実験では，訓練データがないプロパティであっても訓練データのあるプロパティに匹敵する精度で抽出できることを明らかにした．

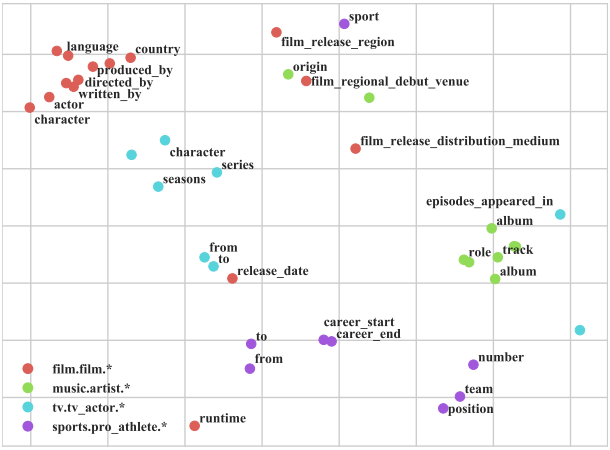

図 4　WEB19 データセットにおいて SUM-R によって得られたプロパティ埋め込み．

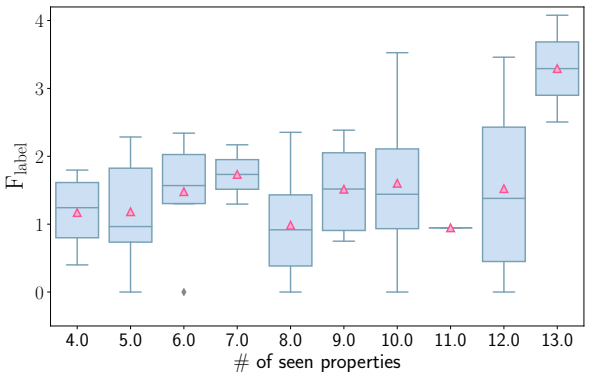

図 5　WEB19 データセットにおける提案手法 #6 の未知プロパティに対する相対 F 値と同じエンティティタイプに属する既知プロパティの数の関係．

謝辞 本研究は JSPS 科研費 18H03244，18H03243，および，JST さきがけ JPMJPR1853 の支援の助成を受けたものです．ここに記して謝意を表します．

文　献

[1] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*, pages 2787–2795, 2013.

[2] H. Bota, K. Zhou, and J. M. Jose. Playing your cards right: The effect of entity cards on search behaviour and workload. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, pages 131–140. ACM, 2016.

[3] R. C. Bunescu and R. J. Mooney. A shortest path dependency kernel for relation extraction. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pages 724–731. Association for Computational Linguistics, 2005.

[4] A. Culotta and J. Sorensen. Dependency tree kernels for relation extraction. In *Proceedings of the 42nd annual meeting on association for computational linguistics*, page 423. Association for Computational Linguistics, 2004.

[5] A. Frome, G. S. Corrado, J. Shlens, S. Bengio, J. Dean, T. Mikolov, et al. Devise: A deep visual-semantic embedding model. In *Advances in neural information processing systems*, pages 2121–2129, 2013.

[6] J. Guo, G. Xu, X. Cheng, and H. Li. Named entity recognition in query. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 267–274. ACM, 2009.

[7] Z. GuoDong, S. Jian, Z. Jie, and Z. Min. Exploring various knowledge in relation extraction. In *Proceedings of the 43rd annual meeting on association for computational linguistics*, pages 427–434. Association for Computational Linguistics, 2005.

[8] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[9] R. Hoffmann, C. Zhang, X. Ling, L. Zettlemoyer, and D. S. Weld. Knowledge-based weak supervision for information extraction of overlapping relations. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 541–550. Association for Computational Linguistics, 2011.

[10] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[11] C. H. Lampert, H. Nickisch, and S. Harmeling. Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):453–465, 2014.

[12] J. Lei Ba, K. Swersky, S. Fidler, et al. Predicting deep zero-shot convolutional neural networks using textual descriptions. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4247–4255, 2015.

[13] L. v. d. Maaten and G. Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[14] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[15] M. Mintz, S. Bills, R. Snow, and D. Jurafsky. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, pages 1003–1011. Association for Computational Linguistics, 2009.

[16] R. J. Mooney and R. C. Bunescu. Subsequence kernels for relation extraction. In *Advances in neural information processing systems*, pages 171–178, 2006.

[17] J. Pound, P. Mika, and H. Zaragoza. Ad-hoc object retrieval in the web of data. In *Proceedings of the 19th international conference on World wide web*, pages 771–780. ACM, 2010.

[18] J. J. Randolph. Free-marginal multirater kappa (multirater k [free]): An alternative to fleiss' fixed-marginal multirater kappa. *Online submission*, 2005.

[19] S. Reed, Z. Akata, H. Lee, and B. Schiele. Learning deep representations of fine-grained visual descriptions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 49–58, 2016.

[20] S. Riedel, L. Yao, and A. McCallum. Modeling relations and their mentions without labeled text. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 148–163. Springer, 2010.

[21] M. Schuster and K. K. Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681, 1997.

[22] R. Socher, M. Ganjoo, C. D. Manning, and A. Ng. Zero-shot learning through cross-modal transfer. In *Advances in neural information processing systems*, pages 935–943, 2013.

[23] R. Socher, B. Huval, C. D. Manning, and A. Y. Ng. 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*, pages 1201–1211. Association for Computational Linguistics, 2012.

[24] M. Surdeanu, J. Tibshirani, R. Nallapati, and C. D. Manning. Multi-instance multi-label learning for relation extraction. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*, pages 455–465. Association for Computational Linguistics, 2012.

[25] Y. Xian, Z. Akata, G. Sharma, Q. Nguyen, M. Hein, and B. Schiele. Latent embeddings for zero-shot classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 69–77, 2016.

[26] X. Yin and S. Shah. Building taxonomy of web search intents for name entity queries. In *Proceedings of the 19th international conference on World wide web*, pages 1001–1010. ACM, 2010.

[27] D. Zelenko, C. Aone, and A. Richardella. Kernel methods for relation extraction. *Journal of machine learning research*, 3(Feb):1083–1106, 2003.

[28] D. Zeng, K. Liu, S. Lai, G. Zhou, and J. Zhao. Relation classification via convolutional deep neural network. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 2335–2344, 2014.

[29] D. Zhang and D. Wang. Relation classification via recurrent neural network. *arXiv preprint arXiv:1508.01006*, 2015.