

修士論文

オートエンコーダとの同時学習によるパラメータ共有

高橋 諒

2018年2月13日

東北大学 大学院
情報科学研究科 システム情報科学専攻

本論文は東北大学大学院情報科学研究科システム情報科学専攻に
修士(工学)授与の要件として提出した修士論文である。

高橋 諒

審査委員：

乾 健太郎 教授 (主指導教員)

木下 賢吾 教授

塩入 諭 教授

岡崎 直観 教授 (東京工業大学)

オートエンコーダとの同時学習によるパラメータ共有*

高橋 諒

内容梗概

エンティティとその関係に関する実世界の事実を集めた知識ベースは、さまざまな自然言語処理タスクのための有用な資源である。しかし、知識ベースは一般的に不完全であるため、知識ベース補完、すなわち、知識ベースにない関係が真であるかどうかを予測することが有用であり、盛んに研究されてきた。これらの研究では、エンティティと関係を含む知識ベースの要素を連続ベクトル空間に埋め込むことで、格納された事実の背後に存在する統計的な規則性をモデル化し、記述漏れした事実を導き出す。

この連続空間埋め込みを行う際の研究課題の一つに、関係間のパラメータ共有がある。直感的に、例えば「(人物が) (作品を) 監督した」と「(人物が) (作品を) 製作した」という関係はどちらも「人物」と「作品」を結びつける概念であるように、多くの関係はいくつかの概念を共有するので、このような直感を知識ベースのモデル化に取り入れることが望ましい。

そこで本研究は、知識ベースの埋め込みを訓練すると同時に、関係に対するオートエンコーダを合わせて学習することで、関係間のパラメータ共有を通して共有概念の学習を促す柔軟な枠組みを提案し、評価実験を行なった。知識ベース補完で一般的に用いられるベンチマークデータセットでの実験は、提案モデルの最先端の性能を示す。

キーワード

機械学習, 知識ベース, 情報抽出

* 東北大学大学院情報科学研究科システム情報科学専攻 修士論文, B6IM2029, 2018年2月13日.

目次

1	はじめに	1
2	関連研究	2
3	モデル	3
3.1	オートエンコーダとの同時学習	4
4	実験	4
4.1	設定	4
4.2	結果	6
4.3	分析	6
5	おわりに	10
	謝辞	11

図目次

- 1 オートエンコーダによる関係の行列のクラスタリング効果. 8
- 2 オートエンコーダとの同時学習を行わないモデルの関係の行列. 9

表目次

1	実験に用いたデータセットの統計情報. #E と #R はそれぞれエンティティと関係の種類数を表す. #Train, #Valid, #Test はそれぞれ訓練, 開発, 評価セットに含まれる三つ組の数を表す.	5
2	WN18 と FB15k データセットにおける知識ベース補完評価結果.	6
3	WN18RR と FB15k-237 データセットにおける知識ベース補完評価結果.	7
4	オートエンコーダとの同時学習を行わないモデル (-autoenc) との比較.	7
5	FB15k-237 において関係 filmWrittenBy に対応する行列とコサイン類似度が高い関係.	9
6	FB15k-237 において関係 awardWinningFilm に対応する行列とコサイン類似度が高い関係.	10

1 はじめに

WordNet [1], YAGO [2], Freebase [3] などの**知識ベース**は, 意味解析 [4], 情報抽出 [5], 質問応答 [6] などに広く応用されてきた. 知識ベースはノードとしての**エンティティ**と異なるタイプのエッジとしての**関係**とで構成されるマルチリレーショナルグラフであり, **事実**を (ヘッドエンティティ, 関係, テールエンティティ) の三つ組で表現する. 事実を完全網羅するのが難しいため, 欠損されたエンティティを予測する**知識ベース補完**が盛んに研究されてきた [7]. これらの研究では, エンティティと関係を含む知識ベースの要素を連続ベクトル空間に埋め込むことで, 格納された事実 (三つ組) の背後に存在する統計的な規則性をモデル化し, 記述漏れした事実を導き出す.

この連続空間埋め込みを行う際の研究課題の一つに, 関係間の知識共有がある. 直感的に, 例えば「(人物が) (作品を) 監督した」と「(人物が) (作品を) 製作した」という関係はどちらも「人物」と「作品」を結びつく概念であるように, 多くの関係はいくつかの概念を共有するので, このような直感を知識ベースのモデル化に取り入れることが望ましい. 連続空間への埋め込みはこのような概念の共有を促すが, 一方で知識データベースのモデル化の性質上, 関係はエンティティ間を写像する演算として機能しなければならないので, 関係を埋め込む際の選択肢は大幅に制限される. 例えば, 関係をエンティティベクトル間の線形変換としてモデル化するのは自然な方法であるが, この場合に関係はエンティティベクトルの次元の 2 乗分ものパラメータを持ち, 埋め込みによる概念共有の促進作用が非常に弱いと思われる. 逆に, 関係をエンティティベクトル間の平行移動としてモデル化する場合, 関係の埋め込み次元はエンティティベクトルの次元と等しいが, この場合に概念共有の促進作用が非常に強いけれど, 複雑な関係に対しては表現力が不足すると思われる [7]. このように, 空間埋め込みに頼るだけでは, どの選択肢も関係の概念共有に最適であるとは限らない.

そこで本研究は, 知識ベースの埋め込みを訓練すると同時に, 関係に対するオートエンコーダを合わせて学習することで関係間の知識共有を促す柔軟な枠組みを提案し, 評価実験を行なった. この手法は, 関係のモデル化自体に何の制約も与えずに, 知識共有の強さ度合いをオートエンコーダのコード長で調整できる利点を持つ.

知識ベース補完で一般的に用いられるベンチマークデータセットでの実験は、提案モデルの最先端の性能を示す。関係に対応する行列同士の類似度を比較することで、似た関係は行列同士も似ることを観測し、オートエンコーダによるクラスタリングを効果的に使っていることを明らかにする。

2 関連研究

知識ベース補完のためにこれまで多くの埋め込みモデルが提案されてきた。このようなモデルの近年の包括的な調査は Nguyen (2017) [7] や Wang ら (2017) [8] を参照されたい。

この分野を草分けしたモデル TransE [9] は、関係を単純にエンティティベクトル間の平行移動としてモデル化したものである。しかし関係をこのように表現すると、一対一関係に対しては有効であるが、多対多関係のような複雑な概念に対しては表現力が不足していると思われる [7]。そこで、TransR [10] や STransE [11] などの拡張では、平行移動を行う前に、関係に応じてエンティティベクトルに線形変換を施すことで関係の多様性に対処したが、これによってまた関係に対するたくさんのパラメータ（行列）を導入することになり、関係間の知識共有は課題となった。

このような背景に踏まえ、Xie ら (2017) [12] の研究は本研究とモチベーションを共有し、関係間の知識共有を明示的にモデル化した。具体的には、彼らのモデルは STransE をベースとするが、関係に応じて施す線形変換は数少ない共通概念を表す行列の和で表せるというハードな制約をモデルに加えた。このような制約付きで学習するために、疎なアテンション機構によるアルゴリズムを提案し、関係に対する行列間のパラメータ共有を実現させた。これに対して本研究での提案はハードな制約がないため、通常の高率勾配降下法 (SGD) による学習ができ、より柔軟なモデルであると言える。

知識ベース補完にはパス情報が有用であることも知られている [13,14] が、本稿で提案する同時学習手法は訓練スキームに特別な制約を入れていないので、パス情報を利用する訓練を取り入れることも簡単である。パス情報を使った訓練は、多段階推論によってより多くの欠損事実を補完する可能性を持つ。例えば、バラク・オバマがホノルルで生まれ、

かつホノルルがアメリカ合衆国の都市であることを使えば、オバマの国籍がアメリカ合衆国であることを推測できるかもしれない。したがって、本研究ではパス情報も取り入れた訓練モデルを実装し、評価した。将来的には更に、Riedel ら (2013) [15] の研究のように、自然言語のテキストも使った同時学習で知識ベース補完を行いたい。

オートエンコーダは、高次元な入力データを一旦低次元のコードに変換してから、コードを使って元の入力データを復元するように訓練されたニューラルネットワークである。これまでその次元削減効果が注目され、教師なしの学習手法として事前訓練に使われたり [16]、分散表現の学習手法として単体で使われたり [17–19] してきたが、同時学習で知識共有を促す手段としてオートエンコーダを使ったのは、私の知る限りこの研究が初めてである。

3 モデル

知識ベース補完のモデルとして、Guu らのパス情報を取り入れた双線型モデル [13] と vecDCS [20] の訓練手法をベースにする。このモデルでは、エンティティを d -次元のベクトル、関係を $(d \times d)$ の行列として表現する。学習は、一つのエンティティ h から出発し、いくつかの関係 r_1, \dots, r_n からなるパスを経由して辿り着いたもう一つのエンティティ t に対して、エネルギー関数 $f(h, r_1, \dots, r_n, t)$ を最大化することでエンティティベクトル h, t と行列 M_{r_1}, \dots, M_{r_n} を推定する。エネルギー関数は

$$f(h, r_1, \dots, r_n, t) := \exp(\top \mathbf{h} \mathbf{M}_{r_1}, \dots, \mathbf{M}_{r_n} \mathbf{t})$$

と定義し、知識ベースからこのようなデータが取れる尤度に相当する。推定時は、知識ベースからのデータと合わせ、ランダムに生成された k 個の負例 h, r'_1, \dots, r'_n, t' を使って

$$\frac{f(h, r_1, \dots, r_n, t)}{k + f(h, r_1, \dots, r_n, t)} \cdot \prod_k \frac{k}{k + f(h, r'_1, \dots, r'_n, t')} \quad (1)$$

を最大化する。これは、「 (h, r_1, \dots, r_n, t) が正例で $(h, r'_1, \dots, r'_n, t')$ が負例」であるイベントの尤度に相当する。

3.1 オートエンコーダとの同時学習

ここで、行列 M_r を d^2 -次元のベクトルとして見なした時に記号 m_r を使う。オートエンコーダは、まず行列 A によって m_r を d^2 より遥かに低い次元数 l のコードベクトルに射影し、得られたコード $A m_r$ に非線形変換の $\text{ReLU}(A m_r)$ を掛けた後、行列 B をもって元の行列を復元しようとする：

$$m_r \approx B \text{ReLU}(A m_r)$$

復元したベクトルと m_r とのコサイン類似度を最大化するために、エネルギー関数

$$g(m_r) := \exp(m_r \cdot B \text{ReLU}(A m_r))$$

を定義し、最適化の際にランダムに生成された k 個の負例 $m_{r'}$ と合わせて

$$\frac{g(m_r)}{k + g(m_r)} \cdot \prod_k \frac{k}{k + g(m_{r'})} \quad (2)$$

を最大化する。また、式(2)の最大化において m_r に対する勾配も計算し、これと式(1)で計算された m_r の勾配と合わせてパラメータ m_r の更新を行う。オートエンコーダとの同時学習によって、 m_r が低次元のコードから「復元されやすい位置」、つまり類似した関係同士がクラスタしているような空間位置に動くことが期待される。また、全ての関係 r に対して m_r が同じ行列 A , B によってエンコード・デコードされるので、異なる m_r 同士が行列 A , B を介してパラメータを共有しているとの見方もできる。これによって、異なる関係間の知識共有が促されると思われる。

4 実験

4.1 設定

知識ベース補完の性能を評価するために4つのベンチマークデータセット上で実験を行う。各データセットの統計情報を表1に示す。WN18 (WordNet) と FB15k (Freebase) は Bordes ら (2013) [9] によって、FB15k-237 は Toutanova and Chen (2015) [21] によ

Dataset	#E	#R	#Train	#Valid	#Test
WN18	40,943	18	141,442	5,000	5,000
FB15k	14,951	1,345	483,142	50,000	59,071
WN18RR	40,943	11	86,835	3,034	3,134
FB15k-237	14,541	237	272,115	17,535	20,466

表1 実験に用いたデータセットの統計情報. #E と #R はそれぞれエンティティと関係の種類数を表す. #Train, #Valid, #Test はそれぞれ訓練, 開発, 評価セットに含まれる三つ組の数を表す.

って, WN18RR は Dettmers ら (2017) [22] によってそれぞれ導入された^{*1}. WN18RR と FB15k-237 の開発・評価セットは訓練セットに含まれない語彙を持つため, このような語彙を含む事実は予め除外した.

知識ベース補完では, ヘッドエンティティ (またはテールエンティティ) と関係が与えられたとき, もう一方のエンティティを予測するモデルの性能を評価する. 例えば, 三つ組 (h, r, t) における h と r が与えられたときのテールエンティティを予測するには, 知識ベースに含まれるエンティティ t' に対しエネルギー関数 $f_r(h, t')$ を計算し, その値に応じてエンティティをランク付けする. 予測された平均ランク (Mean Rank; MR) とランクの逆数の平均 (Mean Reciprocal Rank; MRR) と Hits@k (上位 k 位にランクされた正しいエンティティの割合) を報告する. ランク付けの対象とするエンティティを選ぶ手続きは, 先行研究と同様, Bordes ら (2013) [9] の *filtered* 設定に従う. すなわち, あるエンティティをランク付けするとき, 訓練, 開発, または評価セットで既存の三つ組の一部であることが分かっている全てのエンティティを削除する. これは, 評価対象のエンティティよりも他の正しいフィラーを高くランクするようなモデルにペナルティを課すことを回避するためである.

^{*1} WN18 と FB15k は, 評価セットの事例のほとんどがヘッドエンティティとテールエンティティを反転させた事例を訓練セットに含むため, 一般的な知識ベースをモデル化する能力を評価するためのデータセットとして現在では推奨されていない [22].

Method	WN18					FB15k				
	Hits					Hits				
	MR	MRR	@10	@3	@1	MR	MRR	@10	@3	@1
TransE [9]	251	-	89.2	-	-	125	-	47.1	-	-
TransH [23]	303	-	86.7	-	-	87	-	64.4	-	-
TransR [10]	225	-	92.0	-	-	77	-	68.7	-	-
STransE [11]	206	0.657	93.4	-	-	69	0.543	79.7	-	-
ITransF [12]	205	-	94.2	-	-	65	-	81.0	-	-
DistMult [24]	-	0.83	94.2	-	-	-	0.35	57.7	-	-
ComplEx [25]	-	0.941	94.7	-	-	-	0.692	84.0	-	-
ConvE [22]	504	0.942	95.5	94.7	93.5	64	0.745	87.3	80.1	67.0
提案手法	193	0.363	72.6	43.3	19.6	106	0.286	50.2	32.2	17.7

表2 WN18 と FB15k データセットにおける知識ベース補完評価結果.

4.2 結果

WN18 と FB15k における評価結果を表 2 に示す. 提案手法は WN18 上で MR の最先端の性能を達成しているが, MRR と Hits@k は既存手法に比べて低く留まっている.

WN18RR と FB15k-237 における評価結果を表 3 に示す. 提案手法は WN18RR では MRR と Hits@k の点で既存手法に比べて低い性能を示すが, FB15k-237 では全ての評価指標において最先端の性能を達成している.

4.3 分析

4.3.1 オートエンコーダの知識ベース補完への効果

オートエンコーダとの同時学習の知識ベース補完への効果を確認するために, FB15k-237 データセット上で ablation test を行った (表 4). 提案手法は同時学習を行わないモ

Method	WN18RR					FB15k-237				
			Hits					Hits		
	MR	MRR	@10	@3	@1	MR	MRR	@10	@3	@1
DistMult [24]	5110	0.425	49.1	43.9	38.9	254	0.241	41.9	26.3	15.5
ComplEx [25]	5261	0.444	50.7	45.8	41.1	248	0.240	41.9	26.3	15.2
ConvE [22]	7323	0.342	41.1	36.0	30.6	330	0.301	45.8	33.0	22.0
提案手法	1943	0.197	45.6	26.3	6.7	215	0.334	51.8	36.9	24.2

表3 WN18RR と FB15k-237 データセットにおける知識ベース補完評価結果.

Method	FB15k-237				
	Hits				
	MR	MRR	@10	@3	@1
提案手法	215	0.334	51.8	36.9	24.2
-autoenc	211	0.332	51.7	36.6	24.0

表4 オートエンコーダとの同時学習を行わないモデル (-autoenc) との比較.

デル (-autoenc) に比べて, MR の点で性能が下がっているが, MRR と Hits@k の点では性能を改善している.

4.3.2 関係の行列の可視化

関係の行列を t-SNE アルゴリズムで 2 次元に次元圧縮し可視化することで, オートエンコーダのクラスタリング効果を確認する. FB15k-237 上で提案手法により学習された関係の行列のプロットを図 1 に示す. 楕円で囲んだ領域は関係の行列同士が互いに類似したクラスタを表す. オートエンコーダとの同時学習により, 関係たちが複数のクラスタを形成していることが分かる. 図 2 に示したオートエンコーダとの同時学習を行わないモデルでは, 図 1 のようなクラスタ群は形成されていないため, オートエンコーダのク

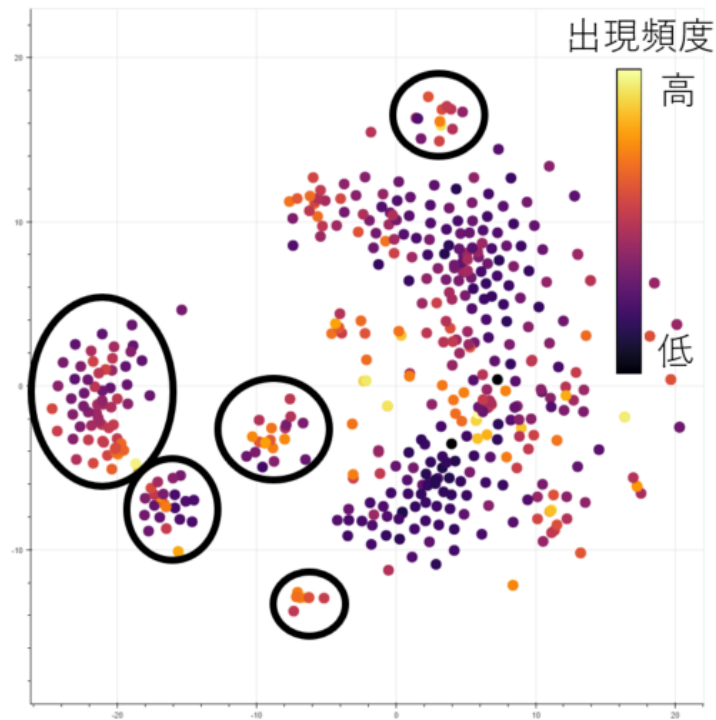


図1 オートエンコーダによる関係の行列のクラスタリング効果。

クラスタリング効果が作用しているものと考えられる。

4.3.3 学習された行列同士の類似度

オートエンコーダとの同時学習により，異なる関係に対応する行列たちは互いにパラメータを共有するため，意味的に似た関係はその行列同士も似るように学習が進むはずである．FB15k-237 において関係 `filmWrittenBy` に対応する行列とコサイン類似度が高い関係の行列上位 5 件を表 5 に示す．関係 `filmWrittenBy` は映画とそのライターを結びつける関係であるが，どちらのモデルにおいても，映画とそれに関わる人を結びつける点で意味的に近い関係たちの類似度が高く学習されている．特に提案手法は，オートエンコーダのパラメータ共有の効果により，意味的に近い関係たちの類似度の値が λ -autoenc モデルに比べて高く学習されている．また，FB15k-237 において関係 `awardWinningFilm` に対応する行列とコサイン類似度が高い関係の行列上位 2 件を表 6 に示す．関係 `awardWinningFilm` は人と賞を受賞した映画を結びつける関係であるが，

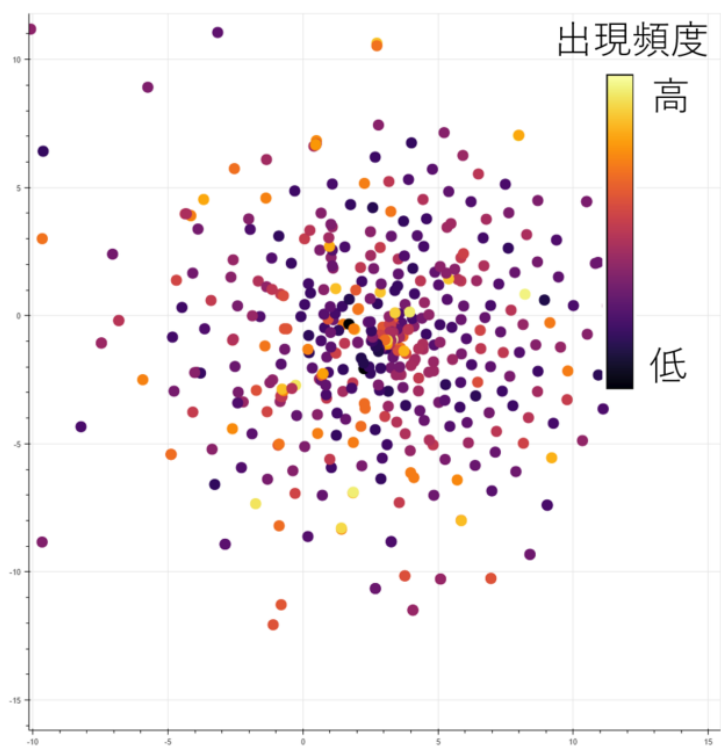


図2 オートエンコーダとの同時学習を行わないモデルの関係の行列.

-autoenc		提案手法	
関係	類似度	関係	類似度
filmAwardNominee	0.334	filmAwardNominee	0.348
filmAwardWinner	0.323	filmAwardWinner	0.329
filmDirector	0.300	filmDirector	0.311
filmProducedBy	0.271	filmActor	0.280
filmActor	0.269	filmProducedBy	0.278

表5 FB15k-237 において関係 filmWrittenBy に対応する行列とコサイン類似度が高い関係.

この関係に対する類似度においても前述の傾向がみられる。これらのような類似度の差が、知識ベース補完の性能評価にゲインをもたらしたものと考えられる。

-autoenc		提案手法	
関係	類似度	関係	類似度
awardNominatedFilm	0.721	awardNominatedFilm	0.759
actorOfFilm	0.509	actorOfFilm	0.543

表6 FB15k-237 において関係 awardWinningFilm に対応する行列とコサイン類似度が高い関係.

5 おわりに

知識ベースの埋め込みを訓練すると同時に、関係に対するオートエンコーダを合わせて学習することで関係間のパラメータ共有を促す柔軟な枠組みを提案した. 知識ベース補完で一般的に用いられるベンチマークデータセットでの実験は, 提案手法の最先端の性能を示した.

謝辞

本研究を進めるにあたり，多くの皆様のご協力，ご助言をいただきましたことに，ここに心より感謝申し上げます。

主指導教員である乾健太郎教授には，ご多忙の中，研究活動だけでなく進路に関することなど多くのご指導，ご助言を頂きましたことに心より感謝申し上げます。副指導教員である岡崎直観准教授には，同じく研究活動に関して多くのご助言を頂きましたことに心より感謝申し上げます。ご多忙の中審査委員をお引き受けくださいました，木下賢吾教授，塩入諭教授に心より感謝申し上げます。研究方針や研究手法，論文執筆に関しまして，直接のご指導を頂いた田然研究特任助教，井之上直也助教授，松林優一郎研究特任助教授に心より感謝申し上げます。研究会や日々の議論におきまして，多くのアドバイスを頂きました乾・岡崎研究室の皆様にご感謝申し上げます。

最後になりましたが，学校生活におきまして関わっていただきましたすべての皆様にご感謝致します。

参考文献

- [1] George A. Miller. Wordnet: A lexical database for english. *Commun. ACM*, 38(11):39–41, 1995.
- [2] Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: a core of semantic knowledge. In *Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 8-12, 2007*, pages 697–706, 2007.
- [3] Kurt D. Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2008, Vancouver, BC, Canada, June 10-12, 2008*, pages 1247–1250, 2008.
- [4] Jonathan Berant, Andrew Chou, Roy Frostig, and Percy Liang. Semantic parsing on freebase from question-answer pairs. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1533–1544, 2013.
- [5] Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1):11–33, 2016.
- [6] Ben Hixon, Peter Clark, and Hannaneh Hajishirzi. Learning knowledge graphs for question answering through conversational dialog. In *NAACL*, 2015.
- [7] Dat Quoc Nguyen. An overview of embedding models of entities and relationships for knowledge base completion. *CoRR*, abs/1703.08098, 2017.
- [8] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. Knowledge graph embedding: A survey of approaches and applications. *IEEE Trans. Knowl. Data Eng.*, 29(12):2724–2743, 2017.

- [9] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.*, pages 2787–2795, 2013.
- [10] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, pages 2181–2187, 2015.
- [11] Dat Quoc Nguyen, Kairit Sirts, Lizhen Qu, and Mark Johnson. Stranse: a novel embedding model of entities and relationships in knowledge bases. In *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016*, pages 460–466, 2016.
- [12] Qizhe Xie, Xuezhe Ma, Zihang Dai, and Eduard H. Hovy. An interpretable knowledge transfer model for knowledge base completion. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 950–962, 2017.
- [13] Kelvin Guu, John Miller, and Percy Liang. Traversing knowledge graphs in vector space. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015*, pages 318–327, 2015.
- [14] Yankai Lin, Zhiyuan Liu, Huan-Bo Luan, Maosong Sun, Siwei Rao, and Song Liu. Modeling relation paths for representation learning of knowledge bases. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language*

- Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015*, pages 705–714, 2015.
- [15] Sebastian Riedel, Limin Yao, Andrew McCallum, and Benjamin M. Marlin. Relation extraction with matrix factorization and universal schemas. In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 9-14, 2013, Westin Peachtree Plaza Hotel, Atlanta, Georgia, USA*, pages 74–84, 2013.
- [16] Dumitru Erhan, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent, and Samy Bengio. Why does unsupervised pre-training help deep learning? *J. Mach. Learn. Res.*, 11:625–660, March 2010.
- [17] Richard Socher, Jeffrey Pennington, Eric H. Huang, Andrew Y. Ng, and Christopher D. Manning. Semi-supervised recursive autoencoders for predicting sentiment distributions. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 151–161, Edinburgh, Scotland, UK., July 2011. Association for Computational Linguistics.
- [18] Carina Silberer and Mirella Lapata. Learning grounded meaning representations with autoencoders. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 721–732, Baltimore, Maryland, June 2014. Association for Computational Linguistics.
- [19] Ivan Titov and Ehsan Khoddam. Unsupervised induction of semantic roles within a reconstruction-error minimization framework. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1–10, Denver, Colorado, May–June 2015. Association for Computational Linguistics.
- [20] Ran Tian, Naoaki Okazaki, and Kentaro Inui. Learning semantically and additively compositional distributional representations. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1:*

- Long Papers*), pages 1277–1287, Berlin, Germany, August 2016. Association for Computational Linguistics.
- [21] Kristina Toutanova and Danqi Chen. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*, pages 57–66, 2015.
- [22] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2D Knowledge Graph Embeddings. In *Advances in Neural Information Processing Systems 30*, 2017.
- [23] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada.*, pages 1112–1119, 2014.
- [24] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. In *Proceedings of the 3rd International Conference on Learning Representations*, pages 1–12, 2015.
- [25] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In *Proceedings of the 33rd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, pages 2071–2080, 2016.