

# 知識グラフ補完のためのモデル予測に基づくサブサンプリング

Xincan Feng<sup>1</sup> 上垣外英剛<sup>1</sup> 林克彦<sup>2</sup> 渡辺太郎<sup>1</sup>

<sup>1</sup> 奈良先端科学技術大学院大学 <sup>2</sup> 北海道大学

{feng.xincan.fy2, kamigaito.h, taro}@is.naist.jp katsuhiko-h@ist.hokudai.ac.jp

## 概要

知識グラフの埋め込み (KGE) において、サブサンプリングは知識グラフ (KG) データセットの疎性により生じる過学習を減少させる効果がある。しかし、現在のサブサンプリングは、エンティティとその関係からなるクエリの頻度を考慮するだけであるため、頻度の少ないクエリに含まれるエンティティや関係の頻度が高い場合、そのようなクエリの出現確率を低く見積もる可能性がある。この問題を解決するために、KGE モデルの予測を利用することでクエリの出現確率を推定する、モデルに基づくサブサンプリング (MBS) および混合サブサンプリング (MIX) を提案する。FB15k-237 および WN18RR での評価を行った結果、提案法であるサブサンプリングは代表的な KGE モデルである ComplEx, RotatE, および HAKE において知識グラフの補完性能が改善されることを示した。

## 1 はじめに

知識グラフ (KG) とは、エンティティとその関係を枝として含むグラフである。知識グラフは、対話 [1], 質問応答 [2], 自然言語生成 [3] などの様々な自然言語処理タスクにおいて重要なリソースとして利用されている。しかし、KG 内のエンティティの関係を人手で全て網羅するには多くのコストがかかる。知識グラフ補完 (KGC) は、知識グラフ中の関係を基に、含まれていない関係を自動的に補完することでこの問題を解決することができる。 $e_i$  と  $e_k$  をエンティティ、 $r_j$  をそれらの関係とすると、KGC は  $(e_i, r_j, ?)$  や  $(?, r_j, e_k)$  などのクエリに対して  $?$  を埋めて回答することにより枝の存在を予測する。

現在、KGC では知識グラフの埋め込み (KGE) を用いる手法が主流である。KGE はエンティティとその関係を連続ベクトルとして表現する。このベクトルの数は KG 内の枝の数に比例して増加するため、KGE では学習時の計算量を減らすためにネガ

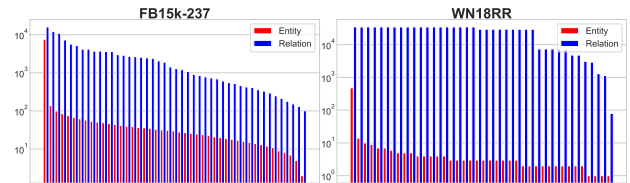


図1 FB15k-237 と WN18RR の学習データにおいて、一度だけしか出現しないクエリに含まれるエンティティと関係を持つ出現頻度<sup>1)</sup>。

ティブサンプリング (NS) を用いることが一般的である。NS では、KGE モデルが KG 内の枝からサンプリングして作成された偽の枝と真の枝とを識別することで KG を学習する。NS は計算量を削減できる一方で、サンプリングされた枝は元の KG の偏りを反映してしまうという問題がある。

その解決策として、KGE の NS に対し、サブサンプリング法 [4] が導入された [5]。サブサンプリングは、頻度の高い枝の頻度を減らし、頻度の低い枝の頻度を増やすことで、KG の偏りを緩和する。

しかし、KGE における現在のサブサンプリングは、クエリの頻度のみを考慮する。そのため、出現頻度の低いクエリに含まれるエンティティや関係の出現頻度が高い場合に、そのようなクエリの出現確率を過小評価する可能性がある。図1は、学習データに1回だけ出現したクエリに含まれるエンティティと関係の出現頻度を示している。この結果から、現在のカウントに基づくサブサンプリング (CBS) では、出現頻度の低いクエリに含まれるエンティティや関係の出現頻度が十分であるにもかかわらず有効に利用されていない。

この問題に対処するため、我々はモデルに基づくサブサンプリング (MBS) と混合分布に基づくサブサンプリング (MIX) を提案し、KGE モデルの予測結果に基づいてクエリの出現確率を推定することでこのような頻度の低いクエリに対処する。

1) このグラフでは紙面の都合で全ての値を表示が難しいため、エンティティと関係を頻度で降順に並べた後、2,000件ごとに値を示している。

本稿では、一般的に使用されている KGE モデル ComplEx [6], RotatE [5], HAKE [7] に我々の MBS と MIX を適用し, FB15k-237 [8] と WN18RR [9] データセットを用いた評価を実施した. 評価の結果, MBS と MIX はそれぞれの設定において, KGC の評価尺度である MRR, Hits@1, Hits@3, Hits@10 においてベースラインである CBS よりも高いスコアを達成可能であることが分かった.

## 2 KGE に適したサブサンプリング

### 2.1 KGE の定式化

本稿ではエンティティ  $h$  と  $t$  とその関係  $r$  を表す枝 (トリプレット) を  $(h, r, t)$  と表記する. KGC ではクエリ  $(h, r, ?)$  や  $(?, r, t)$  が与えられ, モデルは  $?$  に対応するエンティティを予測する. 入力されたクエリを  $x$ , 予測すべきエンティティを  $y$  とすると, モデルパラメータ  $\theta$  に基づくスコア関数  $s_\theta(x, y)$  は  $x$  から  $y$  が予測される確率  $p_\theta(y|x)$  を計算する. 一般的には,  $|D|$  個の枝を用いて,  $p_\theta(y|x)$  を予測することで,  $\theta$  を学習させる. ここで  $D = \{(x_1, y_1), \dots, (x_{|D|}, y_{|D|})\}$  は  $p_d(x, y)$  に従う観測データとする.

### 2.2 NS におけるサブサンプリング

計算効率の高さから NS は KGE モデルの学習によく使われる. 近年, Kamigaito ら [10] は KGE の NS において使用されている様々なサブサンプリングを説明可能な一般的な定式化を行った. この定式化においてサブサンプリングは二つの項  $A$  と  $B$  を用いて次のように表記される:

$$-\frac{1}{|D|} \sum_{(x,y) \in D} \left[ A \log(\sigma(s_\theta(x, y) + \gamma)) + \frac{1}{\nu} \sum_{y_i \sim p_n(y_i|x)} B \log(\sigma(-s_\theta(x, y_i) - \gamma)) \right]. \quad (1)$$

ここで  $\sigma$  はシグモイド関数を,  $p_n(y_i|x)$  は負例をサンプリングするためのノイズ分布を,  $\nu$  は一つの正例  $(x, y)$  に対してサンプリングする負例の数を,  $\gamma$  はマージン項をそれぞれ表す.

表 1 に, KGE において現在提案されているサブサンプリング手法を列挙する. これらは Sun ら [5] によって提案された word2vec [4] のサブサンプリングを用いる手法 (Base), 頻度に基づく手法 (Freq) [10], そして, ユニーク性を考慮した手法 (Uniq) [10] である. ここで  $\#$  は頻度,  $\#(x, y)$  は  $(x, y)$  の頻度,  $\alpha$  は頻度を平滑化するための温度を各々表している.

表 1 現在の KGE で使用されているサブサンプリング法と対応する項  $A$  と  $B$  の一覧.

手法	$A$	$B$
Base	$\frac{\#(x, y)^{-\alpha}  D }{\sum_{(x', y') \in D} \#(x', y')^{-\alpha}}$	$\frac{\#(x, y)^{-\alpha}  D }{\sum_{(x', y') \in D} \#(x', y')^{-\alpha}}$
Freq	$\frac{\#(x, y)^{-\alpha}  D }{\sum_{(x', y') \in D} \#(x', y')^{-\alpha}}$	$\frac{\#x^{-\alpha}  D }{\sum_{x' \in D} \#x'^{-\alpha}}$
Uniq	$\frac{\#x^{-\alpha}  D }{\sum_{x' \in D} \#x'^{-\alpha}}$	$\frac{\#x^{-\alpha}  D }{\sum_{x' \in D} \#x'^{-\alpha}}$

各  $(x, y)$  の頻度は KG 上で高々 1 であるため, 先行研究では次のようなバックオフ [11] を行い,  $(x, y)$  の頻度をクエリの頻度から近似的に導いている:

$$\#(x, y) \approx \#(x, y)_{cbs} = \frac{\#(h_i, r_j) + \#(r_j, t_k)}{2}, \quad (2)$$

ここで,  $(h_i, r_j, t_k)$  は  $(x, y)$  に対応する枝であり,  $(h_i, r_j), (r_j, t_k)$  はそのクエリである. 式 (2) は頻度情報に重度に依存しているため, 我々は以降, この手法を頻度に基づくサブサンプリング (CBS) と呼ぶ.

## 3 提案手法

CBS は式 (2) に示すように, エンティティと関係の組の頻度の平均で  $\#(x, y)$  を近似する. そのため, 少なくとも 1 つの組の頻度が低いと, CBS は  $\#(x, y)$  をうまく推定できない. このスパースネス問題に対処するために, 以下のように, モデルに基づくサブサンプリング (MBS) と, 混合分布に基づくサブサンプリング (MIX) を提案する.

### 3.1 モデルに基づくサブサンプリング

頻度の低いエンティティと関係のペアに起因する問題を回避するため, MBS では, 学習済みモデル  $\theta$  によって推定される確率を用いて, 以下のように各トリプレットとクエリの頻度を計算する.

$$\#(x, y) \approx \#(x, y)_{mbs} = |D| p_{\theta'}(x, y), \quad (3)$$

$$\#x \approx \#x_{mbs} = |D| \sum_{y_i \in D} p_{\theta'}(x, y_i), \quad (4)$$

$$p_{\theta'}(x, y) = \frac{e^{\text{score}_{\theta'}(x, y)}}{\sum_{(x', y') \in D} e^{\text{score}_{\theta'}(x', y')}}. \quad (5)$$

以下, MBS に使用するモデルをサブモデルと呼ぶ. 式 (2) のカウントによって決定される頻度とは異なり, 式 (5) の  $\text{score}_{\theta'}(x, y)$  は実際の頻度に関わらず推論によって推定される. したがって, MBS は CBS で生じるスパースネスの問題に対処することが期待できる. ただし, MBS の能力は選択するサブモデルに依存するため, その限界については §4 の実験により評価する.

表 2 FB15k-237 での結果. 太字は各サブサンプリング (Base, Freq, Uniq) で最も良い結果であることを, † は各モデルでの最良の結果を示す.

モデル	サブサンプリング	MRR	Hits@		
			1	3	10
Base	CBS	33.7	53.3	<b>37.6</b>	53.3
	MBS	33.9	<b>53.5</b>	37.5	<b>53.5</b>
	MIX	<b>34.0</b>	53.4	<b>37.6</b>	53.4
RotatE	CBS	34.1	53.2	37.6	53.2
	MBS	34.3	<b>53.6</b>	38.0	<b>53.6</b>
	MIX	<b>34.6†</b>	<b>53.6</b>	<b>38.1</b>	<b>53.6</b>
Uniq	CBS	34.0	53.0	37.5	53.0
	MBS	34.3	53.6	37.9	53.6
	MIX	<b>34.6†</b>	<b>53.7†</b>	<b>38.4†</b>	<b>53.7†</b>
Base	CBS	32.2	51.2	35.4	51.2
	MBS	31.3	51.7	34.8	51.7
	MIX	<b>32.5</b>	<b>52.3†</b>	<b>35.8</b>	<b>52.3†</b>
Freq	CBS	32.7	51.2	36.0	51.2
	MBS	31.8	49.9	34.9	49.9
	MIX	<b>32.8†</b>	<b>51.4</b>	<b>36.2†</b>	<b>51.4</b>
Uniq	CBS	<b>32.6</b>	51.0	<b>35.8</b>	51.0
	MBS	31.9	50.6	34.9	50.6
	MIX	<b>32.6</b>	<b>51.2</b>	35.6	<b>51.2</b>
Base	CBS	<b>34.6</b>	54.1	38.4	54.1
	MBS	34.4	<b>54.2</b>	38.0	<b>54.2</b>
	MIX	34.5	<b>54.2</b>	<b>38.5</b>	<b>54.2</b>
Freq	CBS	35.2	<b>54.5</b>	38.7	<b>54.5</b>
	MBS	<b>35.3</b>	54.4	<b>39.1</b>	54.4
	MIX	<b>35.3</b>	<b>54.5</b>	39.0	<b>54.5</b>
Uniq	CBS	35.2	54.6	38.7	54.6
	MBS	35.3	<b>54.7†</b>	<b>39.2†</b>	<b>54.7†</b>
	MIX	<b>35.4†</b>	54.6	39.0	54.6

### 3.2 混合分布に基づくサブサンプリング

言語モデルに関する研究 [12] で言及されているように, 頻度に基づく出現確率とモデル予測に基づく出現確率には異なる長所と短所が存在する. MIX では, CBS と MBS の短所を緩和して長所を高めるために表 1 中の  $\frac{\#(x,y)^{-\alpha}|D|}{\sum_{(x',y') \in D} \#(x',y')^{-\alpha}}$  を, 下記の CBS と MBS による混合分布に置き換える.

$$\lambda \frac{\#(x,y)_{mbs}^{-\alpha}|D|}{\sum_{(x',y') \in D} \#(x',y')_{mbs}^{-\alpha}} + (1-\lambda) \frac{\#(x,y)_{cbs}^{-\alpha}|D|}{\sum_{(x',y') \in D} \#(x',y')_{cbs}^{-\alpha}}. \quad (6)$$

そして表 1 中の  $\frac{\#x^{-\alpha}|D|}{\sum_{x' \in D} \#x'^{-\alpha}}$  を下記で置き換える:

$$\lambda \frac{\#x_{mbs}^{-\alpha}|D|}{\sum_{x' \in D} \#x'_{mbs}^{-\alpha}} + (1-\lambda) \frac{\#x_{cbs}^{-\alpha}|D|}{\sum_{x' \in D} \#x'_{cbs}^{-\alpha}}. \quad (7)$$

表 3 WN18RR での結果. 表記は表 2 と同様.

モデル	サブサンプリング	MRR	Hits@		
			1	3	10
Base	CBS	47.9	55.7	49.5	55.7
	MBS	<b>48.0</b>	<b>57.4†</b>	<b>49.8†</b>	<b>57.4†</b>
	MIX	47.9	56.7	49.7	56.7
RotatE	CBS	47.9	56.8	49.6	56.8
	MBS	<b>48.0</b>	<b>57.0</b>	<b>49.8†</b>	<b>57.0</b>
	MIX	47.9	56.9	49.5	56.9
Uniq	CBS	47.9	<b>57.3</b>	49.7	<b>57.3</b>
	MBS	48.0	56.8	49.6	56.8
	MIX	<b>48.2†</b>	56.8	<b>49.8†</b>	56.8
Base	CBS	46.8	<b>55.6</b>	48.5	<b>55.6</b>
	MBS	47.2	54.7	48.8	54.7
	MIX	<b>47.3</b>	55.3	<b>49.0</b>	55.3
Freq	CBS	47.2	56.2	49.3	56.2
	MBS	<b>48.6†</b>	56.8	<b>50.4†</b>	56.8
	MIX	<b>48.6†</b>	<b>57.0†</b>	50.3	<b>57.0†</b>
Uniq	CBS	47.5	56.2	49.0	56.2
	MBS	48.3	<b>56.5</b>	<b>49.9</b>	<b>56.5</b>
	MIX	<b>48.4</b>	56.4	<b>49.9</b>	56.4
Base	CBS	<b>49.5</b>	58.0	<b>51.3</b>	58.0
	MBS	49.1	57.9	<b>51.3</b>	57.9
	MIX	49.4	<b>58.4</b>	<b>51.3</b>	<b>58.4</b>
Freq	CBS	<b>50.0</b>	58.2	51.9	58.2
	MBS	49.9	<b>58.3</b>	<b>52.0†</b>	<b>58.3</b>
	MIX	49.8	58.1	51.6	58.1
Uniq	CBS	49.7	58.2	51.5	58.2
	MBS	<b>50.1†</b>	<b>58.5†</b>	<b>51.8</b>	<b>58.5†</b>
	MIX	50.0	<b>58.5†</b>	51.7	<b>58.5†</b>

$\lambda$  は混合比率を決定するハイパーパラメータである. なお MIX はマルチタスク学習の一種である<sup>2)</sup>.

## 4 実験

データセット: FB15k-237 と WN18RR を使用した<sup>3)</sup>.

比較手法: ComplEx, RotatE, HAKE に対し, 表 1 の Base, Freq, Uniq を適用し, さらに CBS (§2.2) と提案手法である MBS (§3.1) 及び MIX (§3.2) を適用した.

評価尺度: MRR, Hits@1, Hits@3, Hits@10 を使用.

実装およびハイパーパラメータ: ComplEx, RotatE については先行研究 [5] の実装とハイパーパラメータを継承し, HAKE についても先行研究 [7] の実装とハイパーパラメータを継承した. 温度パラメータ  $\alpha$  について, CBS では, Base の場合は Sun ら [5] を, Freq と Uniq は Kamigaito ら [10] の設定を引き継いだ. MBS と MIX については, 各手法の検証デー

2) 付録 A を参照.

3) 各データセットの統計については付録 B を参照.

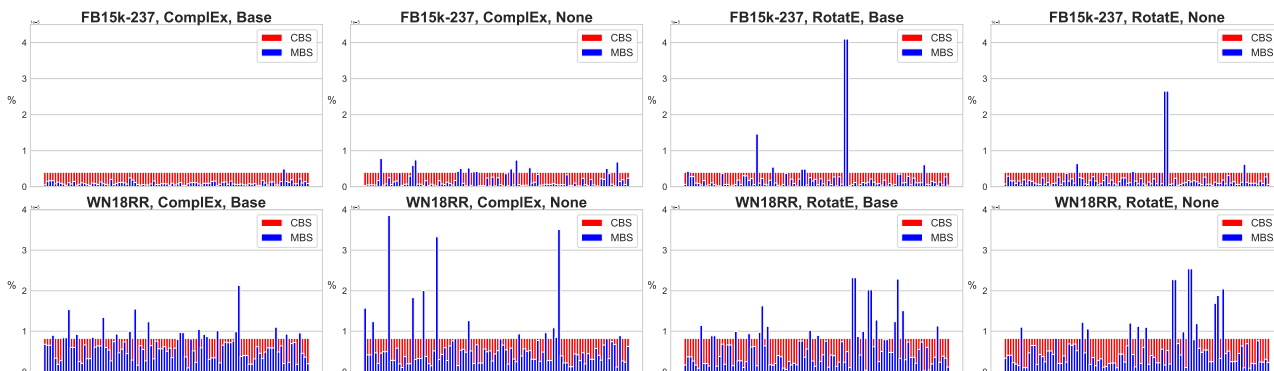


図2 CBSとMBSの各設定において、CBSの頻度で下位100件のクエリの出現確率をCBSの頻度で降順に左から右に整列した際のMBSでの出現確率。

タでのMRRの値を基に $\{2.0, 1, 0.5, 0.1, 0.05, 0.01\}$ から $\alpha$ を、 $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ から混合比 $\lambda$ を選択した。さらに、MBSとMIXのサブモデルを、開発データのMRRを元にComplEx, RotatEから選択した。サブモデルの学習ではサブサンプリング法のBaseを適用した場合とサブサンプリングを用いない場合の二種類を用意した<sup>4)</sup>。

#### 4.1 実験結果および分析

**実験結果** 表2, 3の結果はそれぞれFB15k-237とWN18RRにおけるKGの補完性能を示している。この結果から、MIXやMBSを用いたモデルは、全てのモデルにおいて最も高い性能を達成していることが分かる。この結果は§1で説明したように、MBSとMIXによってKGの補完性能が向上するという考えと一致する。しかし、各評価尺度に対して個別に手法を比較した場合には、CBSがMIXやMBSの補完性能を上回る場合も存在している。これは、MIXやMBSにおいて推定される頻度が選択したサブモデルに依存して変化するためである。

以上より、MIXとMBSは頻度を推定するためのサブモデルを適切に選択することで、KG補完の性能を向上させる可能性があることが分かった。

**分析** 残された課題、すなわち、MBSのためのどのモデルを選択することが好ましいのかについてを分析する。表4に各MBSで選択されたサブモデルを示す。この表からComplExが他のモデルより頻繁に選ばれていることが分かる。その理由を知るために、CBSにおける頻度が下位100件のクエリのMBSにおける頻度を図2に示した。この結果では、FB15k-237においてRotatEではComplExにはない頻度のスパイクがいくつか見られる。WN18RR

表4 開発データで選択されたサブモデルの一覧。

モデル	サブサンプリング	サブモデル			
		FB15k-237		WN18RR	
RotatE	Base	ComplEx	None	ComplEx	None
	Freq	ComplEx	None	ComplEx	None
	Uniq	ComplEx	None	ComplEx	None
ComplEx	Base	ComplEx	None	ComplEx	None
	Freq	ComplEx	Base	ComplEx	None
	Uniq	ComplEx	Base	ComplEx	None
HAKE	Base	ComplEx	None	ComplEx	None
	Freq	ComplEx	Base	ComplEx	None
	Uniq	RotatE	Base	ComplEx	None

ではサブサンプリングを用いないComplExにおけるMBSの頻度の最大値が他の手法に比べて大きくなっている。

この結果はFB15k-237とWN18RRにおいてモデルがそれぞれ平滑化の過剰と不足の問題に遭遇し、MBSがこの問題に対処したことを示している。通常、データが少ないとスパース性が問題になるため、この推測はFB15k-237の学習データがWN18RRより大きいという事実に沿う。このように、MBSでは対象とするデータセットに適したサブモデルを選択することが重要であると考えられる。

#### 5 おわりに

本論文では、モデルを用いてエンティティと関係の組の頻度を推定することで、CBSにおける低頻度な組が起こす問題に対処することが可能な新しいサブサンプリング法である、MBSとMIXを提案した。FB15k-237とWN18RRに対する評価結果から、MBSとMIXによってKG補完の性能が向上することが示された。さらに、分析の結果、データセットに適したサブモデルを選択することがKG補完の性能向上に重要であることを明らかにした。

4) その他の詳細は付録Cに記載。

## 謝辞

本研究は JSPS 科研費 JP21K17801, JP21H03491 の助成を受けたものです。

## 参考文献

- [1] Seungwhan Moon, Pararth Shah, Anuj Kumar, and Rajen Subba. OpenDialKG: Explainable conversational reasoning with attention-based walks over knowledge graphs. In **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**, pp. 845–854, Florence, Italy, July 2019. Association for Computational Linguistics.
- [2] Denis Lukovnikov, Asja Fischer, Jens Lehmann, and Sören Auer. Neural network-based question answering over knowledge graphs on word and character level. In **Proceedings of the 26th International Conference on World Wide Web**, WWW '17, p. 1211–1220, Republic and Canton of Geneva, CHE, 2017. International World Wide Web Conferences Steering Committee.
- [3] Jian Guan, Yansen Wang, and Minlie Huang. Story ending generation with incremental encoding and commonsense knowledge. **Proceedings of the AAAI Conference on Artificial Intelligence**, Vol. 33, No. 01, pp. 6473–6480, Jul. 2019.
- [4] Tomáš Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. **CoRR**, Vol. abs/1310.4546, , 2013.
- [5] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. In **Proceedings of the 7th International Conference on Learning Representations, ICLR 2019**, 2019.
- [6] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. **CoRR**, Vol. abs/1606.06357, , 2016.
- [7] Zhanqiu Zhang, Jianyu Cai, Yongdong Zhang, and Jie Wang. Learning hierarchy-aware knowledge graph embeddings for link prediction, 2019.
- [8] 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**, pp. 57–66, Beijing, China, July 2015. Association for Computational Linguistics.
- [9] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2d knowledge graph embeddings. In **AAAI**, 2018.
- [10] Hidetaka Kamigaito and Katsuhiko Hayashi. Comprehensive analysis of negative sampling in knowledge graph representation learning. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, **Proceedings of the 39th International Conference on Machine Learning**, Vol. 162 of **Proceedings of Machine Learning Research**, pp. 10661–10675. PMLR, 17–23 Jul 2022.
- [11] Slava Katz. Estimation of probabilities from sparse data for the language model component of a speech recognizer. **IEEE transactions on acoustics, speech, and signal processing**, Vol. 35, No. 3, pp. 400–401, 1987.
- [12] Graham Neubig and Chris Dyer. Generalizing and hybridizing count-based and neural language models. In **Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing**, pp. 1163–1172, Austin, Texas, November 2016. Association for Computational Linguistics.

## A マルチタスク学習

式(1)において  $A$  に CBS を適用したものを  $A_{cbs}$ ,  $B$  に CBS を適用したものを  $B_{cbs}$  とし,  $A$  に MBS を適用したものを  $A_{mbs}$ ,  $B$  に MBS を適用したものを  $B_{mbs}$  とする. ここで式(1)に CBS を適用した際の損失関数  $\ell_{cbs}$  は

$$\ell_{cbs} = -\frac{1}{|D|} \sum_{(x,y) \in D} \left[ A_{cbs} \log(\sigma(s_{\theta}(x,y) + \gamma)) + \frac{1}{v} \sum_{y_i \sim p_n(y_i|x)} B_{cbs} \log(\sigma(-s_{\theta}(x,y_i) - \gamma)) \right], \quad (8)$$

式(1)に MBS を適用した際の損失関数  $\ell_{mbs}$  は

$$\ell_{mbs} = -\frac{1}{|D|} \sum_{(x,y) \in D} \left[ A_{mbs} \log(\sigma(s_{\theta}(x,y) + \gamma)) + \frac{1}{v} \sum_{y_i \sim p_n(y_i|x)} B_{mbs} \log(\sigma(-s_{\theta}(x,y_i) - \gamma)) \right], \quad (9)$$

と定義される. また MIX を適用した際の損失関数  $\ell_{mix}$  は, 式(6)と式(7)より,

$$\begin{aligned} \ell_{mix} = -\frac{1}{|D|} \sum_{(x,y) \in D} & \left[ (\lambda A_{mbs} + (1-\lambda)A_{cbs}) \log(\sigma(s_{\theta}(x,y) + \gamma)) \right. \\ & \left. + \frac{1}{v} \sum_{y_i \sim p_n(y_i|x)} (\lambda B_{mbs} + (1-\lambda)B_{cbs}) \log(\sigma(-s_{\theta}(x,y_i) - \gamma)) \right], \quad (10) \end{aligned}$$

と定義される. ここで式(10)は次のように変形が可能である.

$$\ell_{mix} = \lambda \ell_{mbs} + (1-\lambda) \ell_{cbs} \quad (11)$$

式(11)より  $\ell_{mix}$  は  $\ell_{cbs}$  と  $\ell_{mbs}$  の両損失を混合した損失であることから, MIX は CBS と MBS を使用した場合に対するマルチタスク学習であることが示された.

## B データセットの統計量

表5に各データセットの統計を示す.

表5 各データセットの統計量. Entはエンティティを, Relは関係を示す.

データセット	訓練	開発	テスト	Ent	Rel
FB15K-237	272,115	17,535	20,466	14,541	237
WN18RR	86,835	3,034	3,134	40,943	11

## C 学習設定

表6に開発データの MRR で選択された各サブサンプリングの設定を示す. なお, MIX における  $\alpha$  は MBS における最良のものを引き継いでいる.

表6 開発データで選択されたサブモデルとその設定の一覧.

モデル	サブサンプリング	サブモデル							
		FB15k-237				WN18RR			
RotatE	Base	ComplEx	None	$\alpha: 0.5$	$\lambda: 0.9$	ComplEx	None	$\alpha: 1.0$	$\lambda: 0.5$
	Freq	ComplEx	None	$\alpha: 0.1$	$\lambda: 0.7$	ComplEx	None	$\alpha: 0.5$	$\lambda: 0.3$
	Uniq	ComplEx	None	$\alpha: 0.1$	$\lambda: 0.5$	ComplEx	None	$\alpha: 0.5$	$\lambda: 0.5$
ComplEx	Base	ComplEx	None	$\alpha: 1.0$	$\lambda: 0.5$	ComplEx	None	$\alpha: 2.0$	$\lambda: 0.7$
	Freq	ComplEx	Base	$\alpha: 0.5$	$\lambda: 0.1$	ComplEx	None	$\alpha: 0.5$	$\lambda: 0.9$
	Uniq	ComplEx	Base	$\alpha: 0.5$	$\lambda: 0.1$	ComplEx	None	$\alpha: 0.5$	$\lambda: 0.9$
HAKE	Base	ComplEx	None	$\alpha: 1.0$	$\lambda: 0.5$	ComplEx	None	$\alpha: 1.0$	$\lambda: 0.3$
	Freq	ComplEx	Base	$\alpha: 0.5$	$\lambda: 0.5$	ComplEx	None	$\alpha: 1.0$	$\lambda: 0.1$
	Uniq	RotatE	Base	$\alpha: 0.5$	$\lambda: 0.3$	ComplEx	None	$\alpha: 0.5$	$\lambda: 0.7$