# 知識グラフ埋め込みにおける負例サンプリング損失の分析

上垣外 英剛
東京工業大学
kamigaito@lr.pi.titech.ac.jp

林 克彦
群馬大学
khayashi0201@gmail.com

## 概要

知識グラフ埋め込み (KGE) の学習では扱うエンティティ数が膨大であるため，負例サンプリング (NS) 損失が重要な役割を果たす．しかし，NS 損失におけるマージン項，負例サンプル数，サブサンプリング法などを適切に検討しなければ，KGE の学習性能は著しく悪化してしまうことが知られている．これまでこの問題は経験的なハイパーパラメータチューニングによって対処されてきたが，本稿では NS 損失を理論的に分析し，KGE 学習における NS 損失の活用に関する理解を深める．

## 1 はじめに

知識グラフは質問応答・対話といった自然言語処理の課題に対処する上で重要な資源である．しかし，全てのエンティティに対する関係を人手で網羅することは困難であるため，知識グラフ補完 (KGC) に関する研究が進められている．KGC に対するアプローチとしては知識グラフ埋め込み (KGE) がよく研究されている [1, 2, 3, 4, 5].

知識グラフに含まれるエンティティ数は膨大であるため，計算効率性の観点から KGE の学習では負例サンプリング (NS) 損失が広く使用されている．しかし，NS 損失は本来，単語分散表現の学習のために提案された手法である [6]．本来の NS 損失に加えられたいくつかの違いにより，KGE の学習においてはスコア関数，使用する損失関数，ハイパーパラメータの組み合わせは多様なものとなる．これらの組み合わせを適切に選択することで，KGE の性能は大きく変化することが知られているが [7, 8]，どのような組み合わせが適切であるかは理論的に明確ではない．そのため，現状では経験的なハイパーパラメータチューニングに頼っており，様々な組み合わせを考慮するために膨大な計算時間を必要とする．

我々はこの問題に対処するために，KGE で用いられている NS 損失に対しての理論的な検証を行い，いくつかの理論的な事実を明らかにした．それらの理論的な事実を FB15k-237，WN18RR，YAGO3-10 を用いて実験を通して検証した結果，我々が導いた理論的な事実は実際のモデルでも観測されることを示した．さらに，理論的な観点から提案したサブサンプリング法を用いることでモデル性能を改善できることについても示した．

## 2 KGE における NS 損失

本節では KGE を定式化し，それに基づいて，文献 [6] で提案された元の NS 損失と KGE で使用されている NS 損失との差異について説明する．

### 2.1 KGE の定式化

本稿ではエンティティ $e_i$，$e_j$ とその関係 $r_k$ を表す triplet を $(e_i, r_k, e_j)$ と表記する．KGC ではクエリ $(e_i, r_k, ?)$ や $(?, r_k, e_j)$ 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 $s_\theta(x, y)$ の下，$x$ から $y$ が予測される確率 $p_\theta(y|x)$ は，Softmax 関数を用いて次のように定義される:

$$p_\theta(y|x) = \frac{\exp\left(s_\theta(x, y)\right)}{\sum_{y' \in Y} \exp\left(s_\theta(x, y')\right)}. \tag{1}$$

本稿では以後，損失関数が 0 となった際の式 (1) を目的分布と呼ぶ．

### 2.2 NS 損失

Softmax 関数は明示的に確率を表せるが，正規化項によって学習時間が増大してしまう．そのため，学習時に正規化を行うことなく Softmax 関数を近似する手法として NS 損失が提案された [6]．$p_d(x, y)$ に従う観測データ $D = \{(x_1, y_1), \cdots, (x_n, y_n)\}$ に対し，NS 損失は次で定義される．

$$-\frac{1}{|D|} \sum_{(x,y) \in D} \Bigg[\log(\sigma(s_\theta(x, y))) + \sum_{y_i \sim p_n(y|x)} \log(\sigma(-s_\theta(x, y_i)))\Bigg]. \tag{2}$$

$p_n(y|x)$ はノイズ分布，$\sigma$ はシグモイド関数を表す．

一方，近年は式 (2) と共に，マージン項 $\gamma$ を加えた次のような NS 損失が使用されている [4].

$$-\frac{1}{|D|}\sum_{(x,y)\in D}\Big[\log(\sigma(s_\theta(x,y)+\gamma)) + \frac{1}{\nu}\sum_{y_i\sim p_n(y|x)}\log(\sigma(-s_\theta(x,y_i)-\gamma))\Big]. \quad (3)$$

次節以降で両損失関数の差異を議論する．

## 3 理論的な分析

### 3.1 マージン項の役割

本節ではマージン項 $\gamma$ が持つ役割について理論的側面から論じる．式 (2) と式 (3) を比較することで，$\gamma$ の役割に関連する次の命題を示せる．

**命題 1** 式 (2) と式 (3) は同じ目的分布を持つ．[1)]

命題 1 より，$\gamma$ の存在は損失関数が最適解に到達した場合には影響を及ぼさない事が分かる．一方，上記の命題が成立するには，前提となる損失関数が最適解に到達できる必要がある．特に，スコア関数の値域が $(-\infty,+\infty)$ ではない場合，到達不可能な値が存在することになるため，この点を議論する．

値域に制限のあるスコア関数として，TransE [1]，RotatE [4]，HAKE [9] 等に代表される距離ベースのスコアリング法がある．距離ベースのスコアリング法は $p$-ノルムを用いて $-||f_\theta(x,y)||_p$ と表現される．$f_\theta(x,y)$ は (x,y) に対してベクトル値を返す関数である．従って，$f_\theta(x,y)$ が任意のベクトルを表現可能な時，距離ベースのスコアリング法の値域は $(-\infty,0]$ となる．これより次の命題が成立する．

**命題 2** 距離ベースのスコアリング関数は式 (2) において $\nu p_n(y|x) < p_d(y|x)$ となる $(x,y)$ が存在する時，最適解に到達できない．また，式 (3) においては，$\exp(\gamma)p_n(y|x) < p_d(y|x)$ となる $(x,y)$ が存在する時，最適解に到達できない．[2)]

命題 2 より，距離ベースのスコアリング法を NS 損失で学習する際にはノイズ分布を適切に選択する必要があることが分かる．しかし，式 (2) では，$\nu p_n(y|x) > p_d(y|x)$ となる $\nu$ 及びノイズ分布を使用する必要があるが，ノイズ分布の調整のみで全ての (x,y) に対してこの条件を満たすことは困難である．したがって式 (2) では，計算時間を犠牲にサンプル数 $\nu$ を増大させる必要がある．その一方で，式 (3) では，$\exp(\gamma)p_n(y|x) > p_d(y|x)$ を満たせば良いことから，単に $\gamma$ に十分に大きな値を設定すればこの問題を回避することができる．KGE では一般的に $p_n(y|x)$ に一様分布が用いられることから，確率の定義より $p_d(y|x)$ が 1 を超えないことを考慮すると，$\exp(\gamma)$ はラベル数 $|Y|$ よりも大きいことが望ましい．

本節における今までの議論に基づくと，マージン項に大きな値を設定することで欠点に直面することなく，モデルの学習が促進されるように考えられる．しかし，次の命題はマージン項を自由に設定できるわけではないことを示す．

**命題 3** マージン項 $\gamma$ は式 (3) の勾配に影響を与える．[3)]

命題 3 より，マージン項を変更した場合，勾配に関係するその他のハイパーパラメータも適切に設定しなければならないことが分かる．以上より，距離ベースのスコアリング法を用いる際にはマージン項に十分大きな値を設定した上で，勾配に関係するハイパーパラメータを適切に調整しなければならない．その一方で，ComplEx [3] や DistMult [2] のような値域に制限が存在しないスコアリング法ではマージン項の調整は不要であると考えられる．

### 3.2 負例サンプル数の影響

次に負例サンプル数 $\nu$ が学習に与える影響について理論的側面から議論する．命題 1 より，負例サンプル数はマージン項と同様に，式 (2) と式 (3) で表される両損失関数の目的分布に対しては影響しないことが分かる．なお，詳細な議論を付録 B に記載した．

### 3.3 KGE に適したサブサンプリング法

今までの議論は，NS 損失がモデルを観測データにより定義される分布 $p_d(y|x)$ に適合させるという前提に基づいていた．しかし，実際に NS 損失が行うべきことは，モデルを観測データの背後に存在する真の分布 $p'_d(y|x)$ に適合させることである．このためには，観測結果に基づく $p_d(y|x)$ と $p'_d(y|x)$ の

1) 証明は付録 A.1 に記載．
2) 証明は付録 A.2 に記載．
3) 証明は付録 A.3 に記載．

隔たりを埋めなければならない．ここでそのような役割を果たす関数 $A(x,y)$，$B(x)$ が存在すると仮定する．この両関数を用いて，真の分布への適合を考慮した NS 損失が次のように導かれる．

$$\begin{aligned}
&-\frac{1}{|D|}\sum_{(x,y)\in D}\Big[\log(\sigma(s_\theta(x,y)+\gamma))\Big]p'_d(x,y)\\
&+\sum_{y_i\sim p_n(y|x)}\Big[p_n(y_i|x)\log(\sigma(-s_\theta(x,y_i)-\gamma))\Big]p'_d(x)\\
=&-\frac{1}{|D|}\sum_{(x,y)\in D}\Big[\log(\sigma(s_\theta(x,y)+\gamma))\Big]A(x,y)p_d(x,y)\\
&+\sum_{y_i\sim p_n(y|x)}\Big[p_n(y_i|x)\log(\sigma(-s_\theta(x,y_i)-\gamma))\Big]B(x)p_d(x)\\
=&-\frac{1}{|D|}\sum_{(x,y)\in D}\Big[A(x,y)\log(\sigma(s_\theta(x,y)+\gamma))\\
&+\frac{1}{\nu}\sum_{y_i\sim p_n(y|x)}B(x)\log(\sigma(-s_\theta(x,y_i)-\gamma))\Big]. \qquad (4)
\end{aligned}$$

Mikolov ら [6] は式 (3) の NS 損失と同時に，サブサンプリングを提案している．サブサンプリングは学習事例である単語の出現確率を調整するものであり，$A(x,y)$，$B(x)$ に対応する役割を持つ．一方で，KGE においてサブサンプリングを NS 損失で使用することについて議論している論文は我々の知る限りでは存在しない．しかし，実装としては Sun ら [4][4] が word2vec を踏襲する形で頻度に基づいた割引を行う下記のものを使用している．

$$A(x,y)=B(x,y)=\frac{\frac{1}{\sqrt{\#(x,y)}}}{\sum_{(x',y')\in D}\frac{1}{\sqrt{\#(x',y')}}}. \qquad (5)$$

ここで # は頻度を表す記号であり，$\#(x,y)$ は $(x,y)$ の頻度を表す．なお，実際の $(x,y)$ は KG 上で高々 1 回しか出現しないため，$(x,y)=(e_i,r_k,e_j)$ とするとき，$\#(x,y)$ のバックオフ [10] を用いて次のように近似している．

$$\#(x,y)=\#(e_i,r_k)+\#(r_k,e_j) \qquad (6)$$

ここで式 (4) を元に理論的な観点から式 (5) のような頻度による割引を行う KGE に特化したサブサンプリング法を導出する．この導出は真の分布 $p'_d(y|x)$ に対する仮定をどのように置くかによって変化する．まず，真の分布において，$(x,y)$ は頻度を持っているが，観測されたものは高々 1 に過ぎなかったという仮定に基づいた導出を行う．この場合は式 (5) と同様に実際に $\#(x,y)$ の頻度を計算できないため，式 (6) の近似を用いる．ここで式 (4) より，$A(x,y)$ は $(x,y)$ の頻度に対する割引を，$B(x)$ は $x$ の頻度に対する割引をそれぞれ行うため，次のようなサブサンプリング法が導出される．

$$A(x,y)=\frac{\frac{1}{\sqrt{\#(x,y)}}}{\sum_{(x',y')\in D}\frac{1}{\sqrt{\#(x',y')}}},\quad B(x)=\frac{\frac{1}{\sqrt{\#x}}}{\sum_{x'\in D}\frac{1}{\sqrt{\#x'}}} \qquad (7)$$

一方で，真の分布においても，観測結果と同様に $(x,y)$ は高々頻度 1 であると仮定した場合，$p'_d(y|x)=p'_d(x,y)/p'_d(x)\propto 1/p'_d(x)$ となり，$p'_d(y|x)$ は同じ $x$ に対しては同じ値となる．従って，この仮定の下では $p_d(x)$ に対してのみ割引を行うことになり，次のサブサンプリング法が導出される．

$$A(x,y)=B(x)=\frac{\frac{1}{\sqrt{\#x}}}{\sum_{x'\in D}\frac{1}{\sqrt{\#x'}}}. \qquad (8)$$

式 (7) と式 (8) は理論的な側面から導出されているが，実際のタスクにおいては汎化の対象である真の分布はデータセットによって異なる．従って理論的側面からのみサブサンプリング法の優劣を論じることは出来ないため，使用に際しては開発データを用いた検証が必要である．

## 4 経験的な分析

本節では 3 節で論じた理論的な特性が実際のデータやモデルに対して成立するか検証する．データセットは FB15k-237，WN18RR，YAGO3-10 を用いた [5]．モデルは値域に制限を持たない ComplEx，DistMult と，$p$-ノルムを使用し値域に制限をもつ RotatE, HAKE [9] を比較した．特に断りが無い場合，ComplEx，DistMult，RotatE のハイパーパラメータは Sun ら [4] のものを，HAKE のハイパーパラメータは Zhang ら [9] のものを使用した．

### 4.1 マージン項の影響

マージン項を変更した際のモデルの予測精度の変化を調査するために，Mean Reciprocal Rank (MRR) を対象に異なる $\gamma$ を用いた際の結果を比較した．

まず初めに，3.1 節で議論した $p$-ノルムを用いたスコアリング法を適切に学習するためにマージン項が必要であるという理論的な結論が実用上でも成立しているのかを検証する．そのためにマージン項の有無による性能の変化について，一様雑音分布を用いて $p$-ノルムに基づく RotatE，HAKE と値域に制限

4） https://github.com/DeepGraphLearning/KnowledgeGraphEmbedding

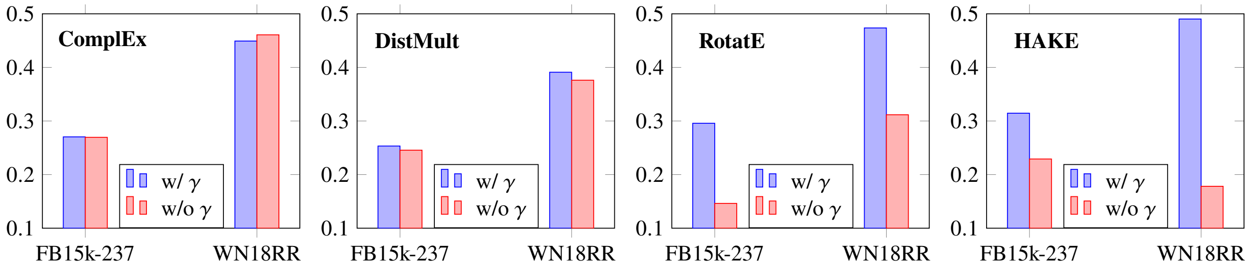


図 1: マージン項 $\gamma$ に 0 を用いた場合と報告値を用いた場合の各設定における MRR.

表 1: 各サブサンプリング法の評価結果．Sub. はサブサンプリングを，Base は式 (5) を，Freq は式 (7) を，Uniq は式 (8) をそれぞれ表す．太字は各設定における最高スコアを表す．

| Model | Sub. | FB15k-237 | | | | WN18RR | | | | YAGO3-10 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | MRR | Hits | | | MRR | Hits | | | MRR | Hits | | |
| | | | @1 | @3 | @10 | | @1 | @3 | @10 | | @1 | @3 | @10 |
| RotatE | - | 32.6 | 22.5 | 36.6 | **52.9** | 47.4 | 43.0 | 49.1 | 56.1 | 50.5 | 41.2 | 56.2 | 67.9 |
| | Base | 33.3 | 23.6 | 36.9 | 52.8 | **47.9** | **43.4** | **49.5** | 56.6 | 50.2 | 40.7 | 56.0 | 67.9 |
| | Freq | **34.0** | **24.5** | **37.5** | **52.9** | **47.9** | **43.4** | **49.5** | 56.6 | 51.1 | 41.8 | 56.8 | **68.0** |
| | Uniq | 33.8 | 24.3 | 37.3 | **52.9** | 47.7 | 43.1 | **49.5** | **56.7** | **51.5** | **42.5** | **57.0** | 67.8 |
| HAKE | - | 32.0 | 21.9 | 36.0 | 52.4 | 49.0 | 44.6 | 50.6 | 57.8 | 53.8 | 45.3 | 59.4 | 68.7 |
| | Base | 33.9 | 24.1 | 37.6 | **53.8** | 49.4 | 44.9 | 51.1 | 58.0 | 54.0 | 45.6 | 59.3 | 69.2 |
| | Freq | 34.4 | 24.7 | **38.1** | **53.8** | **49.7** | **45.2** | **51.5** | 58.2 | 53.6 | 44.7 | 58.9 | 69.3 |
| | Uniq | **34.5** | **24.9** | **38.1** | 53.7 | **49.7** | 45.1 | 51.4 | **58.8** | **54.7** | **46.3** | **59.9** | **69.9** |

が無い ComplEx，DistMult とで比較した．なお，$\gamma$ を元の値から 0 へと大幅に変更するに際し，モデルが適切に学習されるよう，Adam の初期学習率を最もよく使用されている 0.001 へと変更した．

図 1 にマージン項に先行研究の値を用いた際 (w/$\gamma$) の結果とマージンを使用しなかった際 (w/o $\gamma$) の結果を示す．この結果では，ComplEx，DistMult では大きな性能の変化が無く，これらのスコアリング法におけるマージン項の調整の不必要性を改めて確認することが出来る．そして，$p$-ノルムに基づく RotatE，HAKE ではマージン項を用いない場合に大幅な性能低下が確認でき，これらのスコアリング法においてマージン項を使用することが必要であることが改めて分かる．

### 4.2 サブサンプリング法の比較

3.3 節にて議論したサブサンプリング法の比較を行う．比較はサブサンプリングを行わない場合，現在使用されているサブサンプリング法である式 (5)，提案手法である式 (7, 8) との間で行う．比較のためのモデルとしては公開コード中にてサブサンプリングを行っている RotatE と HAKE を選択し，学習では元の論文と同様に自己敵対的負例サンプリング損失 [4] を用いた．評価尺度には MRR，Hits@1，Hits@3，Hits@10 を使用した．表 1 に各設定におけるスコアを示す．この結果より，知識グラフの補完においてサブサンプリングの使用が有効であることが分かる．また，我々の提案法である Uniq と Freq は RotatE を WN18RR で評価した場合以外は共に Base よりも高いかあるいは同等の結果を示しており，理論的側面から導出されたサブサンプリング法が実用上も適切なものであったと考えられる．

## 5 まとめ

本論文では，KGE で用いられている NS 損失に対しての理論的な検証を行い，いくつかの理論的な事実を明らかにした．また，その理論的な事実に基づき，KGE の損失関数に特化したサブサンプリング法についても提案した．それらの理論的な事実及び提案したサブサンプリング法の性能を検証するために，我々は FB15k-237，WN18RR，YAGO3-10 を用いた実験を実施した．実験の結果，我々が導いた理論的な事実は実際のモデルでも観測されることを示した．また，提案したサブサンプリング法が実際にモデルの知識グラフの補完における性能を改善できることについても示した．

## 参考文献


[1] 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**, pp. 2787–2795, 2013.

[2] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. In **Proceddings of the 3rd International Conference on Learning Representations, ICLR 2015**, 2015.

[3] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In **Proceedings of the 33nd International Conference on Machine Learning, ICML 2016**, Vol. 48 of **JMLR Workshop and Conference Proceedings**, pp. 2071–2080. JMLR.org, 2016.

[4] 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.

[5] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2d knowledge graph embeddings. In **Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18)**, pp. 1811–1818, 2018.

[6] Tomás 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.

[7] Daniel Ruffinelli, Samuel Broscheit, and Rainer Gemulla. You CAN teach an old dog new tricks! on training knowledge graph embeddings. In **Proceedings of the 8th International Conference on Learning Representations, ICLR 2020**, 2020.

[8] Mehdi Ali, Max Berrendorf, Charles Tapley Hoyt, Laurent Vermue, Mikhail Galkin, Sahand Sharifzadeh, Asja Fischer, Volker Tresp, and Jens Lehmann. Bringing light into the dark: A large-scale evaluation of knowledge graph embedding models under a unified framework. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, pp. 1–1, 2021.

[9] Zhanqiu Zhang, Jianyu Cai, Yongdong Zhang, and Jie Wang. Learning hierarchy-aware knowledge graph embeddings for link prediction. In **Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence, (AAAI20)**, pp. 3065–3072, 2020.

[10] 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.

[11] Omer Levy and Yoav Goldberg. Neural word embedding as implicit matrix factorization. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, editors, **Advances in Neural Information Processing Systems**, Vol. 27. Curran Associates, Inc., 2014.

[12] Hidetaka Kamigaito and Katsuhiko Hayashi. Unified interpretation of softmax cross-entropy and negative sampling: With case study for knowledge graph embedding. In **Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021**, pp. 5517–5531, 2021.

[13] 上垣外英剛, 林克彦. Unified interpretation of softmax cross entropy and negative sampling: With case study for knowledge graph embedding. 自然言語処理, Vol. 28, No. 4, pp. 1336–1341, 2021.

[14] Samuel Broscheit, Daniel Ruffinelli, Adrian Kochsiek, Patrick Betz, and Rainer Gemulla. LibKGE - A knowledge graph embedding library for reproducible research. In **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations**, pp. 165–174, 2020.

# A 証明

## A.1 命題 1 の証明

先行研究 [11, 12, 13] より，式 (2) が 0 となる際に

$$\exp(s_\theta(x,y)) = \frac{p_d(y|x)}{\nu p_n(y|x)}, \tag{9}$$

が成立．ここで式 (3) が 0 となる際を考える．式 (3) における $\nu$ への正規化により式 (9) の $\nu$ は消去される．さらにマージン項 $\gamma$ の存在により，式 (3) が 0 となる際に

$$\exp(s_\theta(x,y)) = \frac{p_d(y|x)}{\exp(\gamma) p_n(y|x)}, \tag{10}$$

が成立．式 (9) を式 (1) に代入した結果と，式 (10) を式 (1) に代入した結果は共に

$$\frac{p_d(y|x)}{\sum_{y'\in Y} p_d(y'|x))}, \tag{11}$$

となることから命題 1 が示された．

## A.2 命題 2 の証明

$s_\theta(x,y) \leqq 0$ のとき，$\exp(s_\theta(x,y)) \leqq 1$ が成立する．このとき，式 (9) の右辺が同じ値域を持たない場合，式 (2) は到達不可能な点を持つ可能性がある．従って式 (2) が到達不可能な点を持たない条件は次のように導かれる

$$\frac{p_d(y|x)}{\nu p_n(y|x)} \leqq 1$$
$$\Leftrightarrow p_d(y|x) \leqq \nu p_n(y|x). \tag{12}$$

式 (3) が到達不可能な点を持たない条件も同様に次のように導かれる

$$\frac{p_d(y|x)}{\exp(\gamma) p_n(y|x)} \leqq 1$$
$$\Leftrightarrow p_d(y|x) \leqq \exp(\gamma) p_n(y|x). \tag{13}$$

## A.3 命題 3 の証明

合成関数の微分の定義より式 (3) を微分した結果には $\gamma$ が含まれるため自明．

# B 負例サンプル数の影響

本節では 3.2 節で議論した負例サンプル数が学習に与える影響について調査する．具体的には，様々な負例サンプル数に基づきモデルを学習し，学習されたモデルの MRR を比較する．式 (2) に基づく NS 損失を使用するため，我々は LibKGE [14][5] を使用した．負例サンプル数以外のハイパーパラメータについては Kamigaito ら [12] と同じものを使用した．

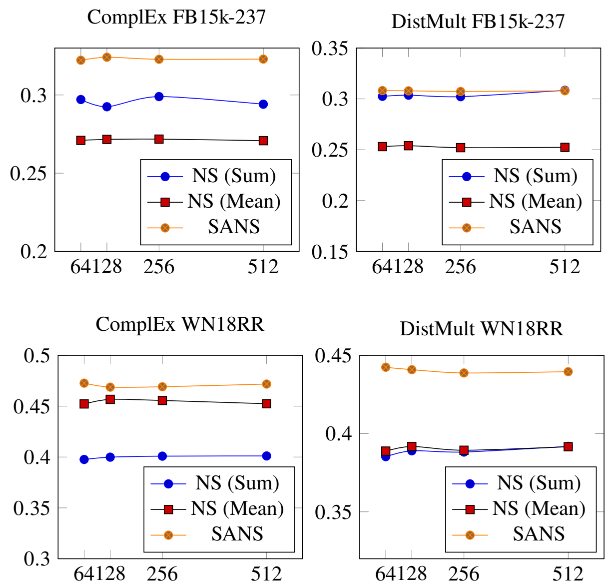


図 2: 負例サンプル数が変化した際の各設定における MRR．NS (Sum) は式 (2)，NS (Mean) は式 (3) の NS 損失をそれぞれ表している

図 2 に負例サンプル数を変化させた際の各設定における MRR を示す．．この結果より，損失関数によって示す値こそ異なるものの，負例サンプル数の違いによる MRR の変動幅はどの損失関数においてもほぼ同じであることが分かる．また，負例サンプル数の影響を強く受けると考えられる SANS においても負例サンプル数の違いによるスコアの大きな変動は観測されない．この結果は目的分布の観点からは負例サンプル数による影響はない，という 3.2 節の内容に沿うものである．

5） https://github.com/uma-pi1/kge