ACL2018読み会@名大

# Towards Understanding the Geometry of Knowledge Graph Embeddings

**Chandrahas**
Indian Institute of Science
chandrahas@iisc.ac.in

**Aditya Sharma**
Indian Institute of Science
adityasharma@iisc.ac.in

**Partha Talukdar**
Indian Institute of Science
ppt@iisc.ac.in

http://aclweb.org/anthology/P18-1012

http://anthology.aclweb.org/attachments/P/P18/P18-1012.Notes.pdf

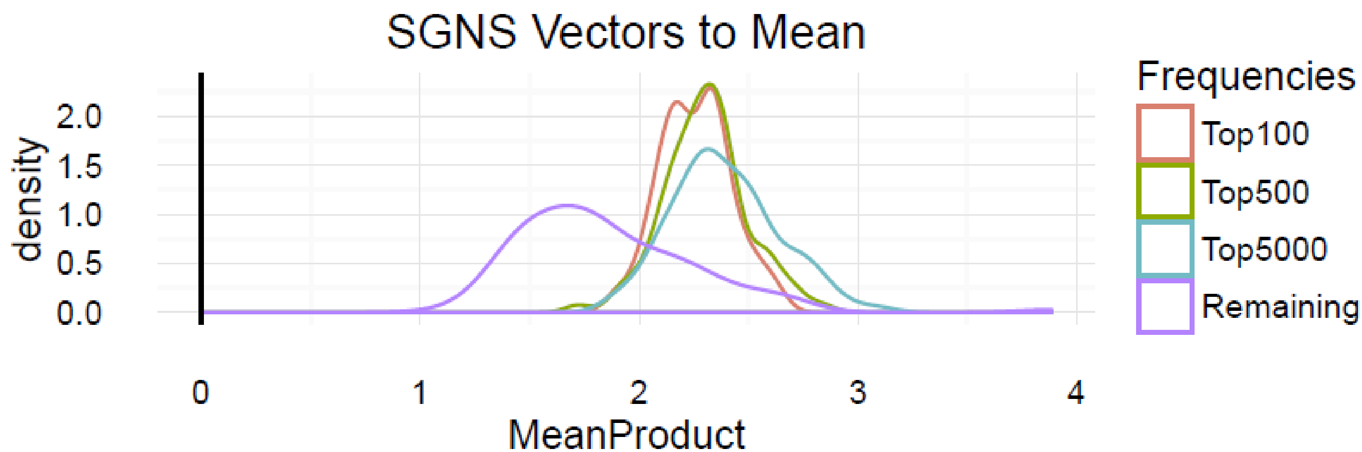
紹介者: 笹野遼平（名大）

# はじめに

- なぜこの論文を選んだか:
  - Embeddingの形状(geometry)分析に興味（how to?）
    - cf. SGNSの形状分析 [Mimno & Thompson ACL’17]
  - 最先端NLP勉強会＠東京8/3,4で紹介されなかった
- 概要:
  - 6つの知識グラフ(Knowledge Graph)の埋め込み手法でそれぞれ得られた要素/形状ベクトルの形状を始めて分析
  - 主に中心ベクトルとの類似度（ATM）を分析
    - Additiveな手法とMultiplicativeな手法（後述）で大きな違い
    - 前者では0を中心に広く分散
    - 後者では正の値を偏り、分散は小さめ（SGNSと近い傾向）
  - パフォーマンス等の分析はかなり微妙な印象

# 形状分析の先行研究: The strange geometry of skip-gram with negative sampling [Mimno & Thompson'17]

Figure 2



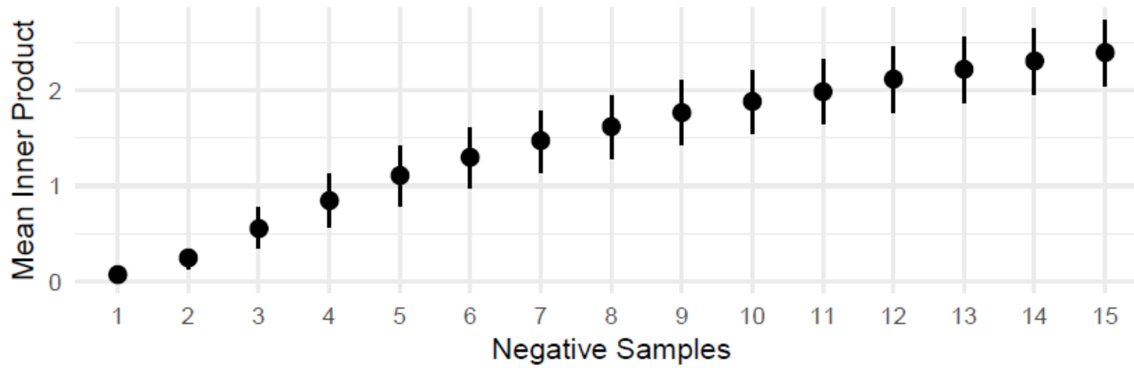
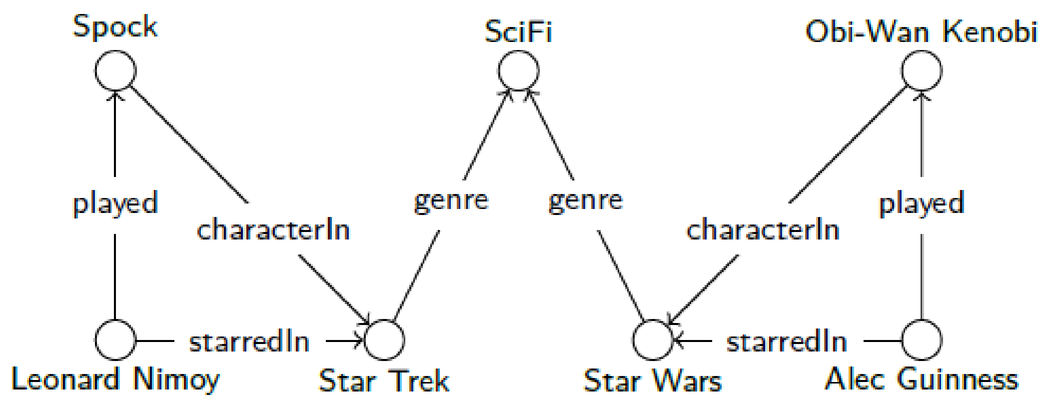




Figure 6: The number of negative samples affects the inner product between vectors and the mean vector. Results are indistinguishable across 10 initializations for each value.

# 知識グラフ(KG: knowledge graph)

- 有向グラフの形で表された知識(e.g. FreeBase)
  - それぞれのノードがエンティティ
  - ノード間にエッジを張り、関係をラベルで表現
  - ($h$; $r$; $t$)の形(三つ組: triplet)で表現($h$:始点, $r$: 関係, $t$: 終点)
    - e.g. J. K. Rowling [influenced by] C. S. Lewis



# Graph Embedding (グラフ埋め込み)

- 知識グラフをベクトル空間上に埋め込む手法
- 代表的な手法: TransE [Bordes+’13]
  - 知識ベースに含まれる三つ組(h; r; t) に対し、
    ***h* + *r* = *t*** が成立するようにベクトルを学習
  - 他の三つ組に対しては成立しないようにしたい
    ⇒ negative sampling で負例を生成し学習
- 応用: Link Prediction, Entity Classification

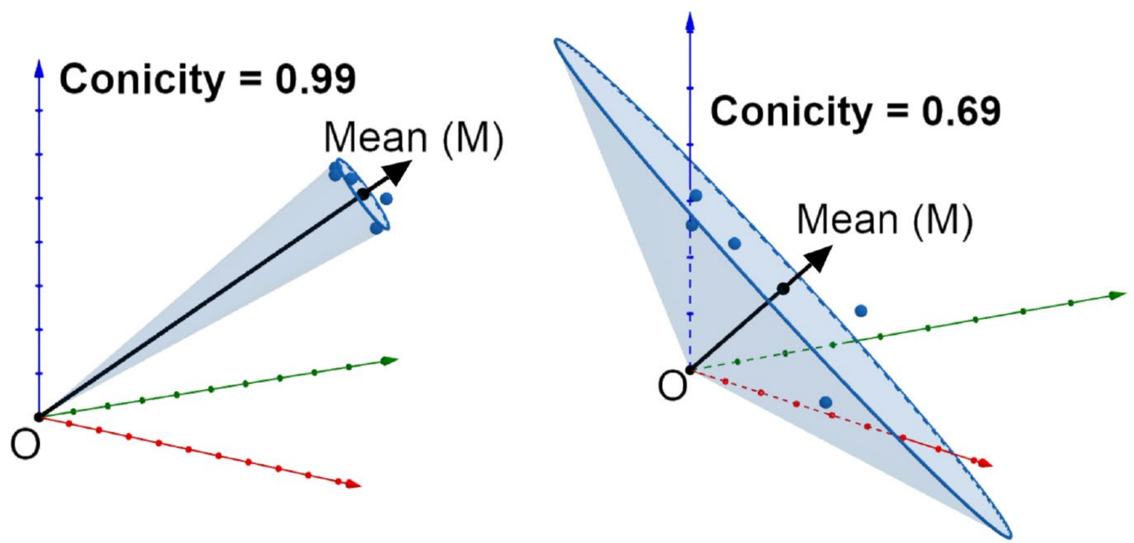
# 分析対象のグラフ埋め込み手法

| Type | Model | Score Function $\sigma(h, r, t)$ |
|---|---|---|
| Additive | TransE (Bordes et al., 2013) | $-\|\mathbf{h}+\mathbf{r}-\mathbf{t}\|_1$ |
| | TransR (Lin et al., 2015) | $-\|M_r\mathbf{h}+\mathbf{r}-M_r\mathbf{t}\|_1$ |
| | STransE (Nguyen et al., 2016) | $-\left\|M_r^1\mathbf{h}+\mathbf{r}-M_r^2\mathbf{t}\right\|_1$ |
| Multiplicative | DistMult (Yang et al., 2014) | $\mathbf{r}^\top(\mathbf{h}\odot\mathbf{t})$ |
| | HolE (Nickel et al., 2016) | $\mathbf{r}^\top(\mathbf{h}\star\mathbf{t})$ |
| | ComplEx (Trouillon et al., 2016) | $\mathbf{Re}(\mathbf{r}^\top(\mathbf{h}\odot\bar{\mathbf{t}}))$ |

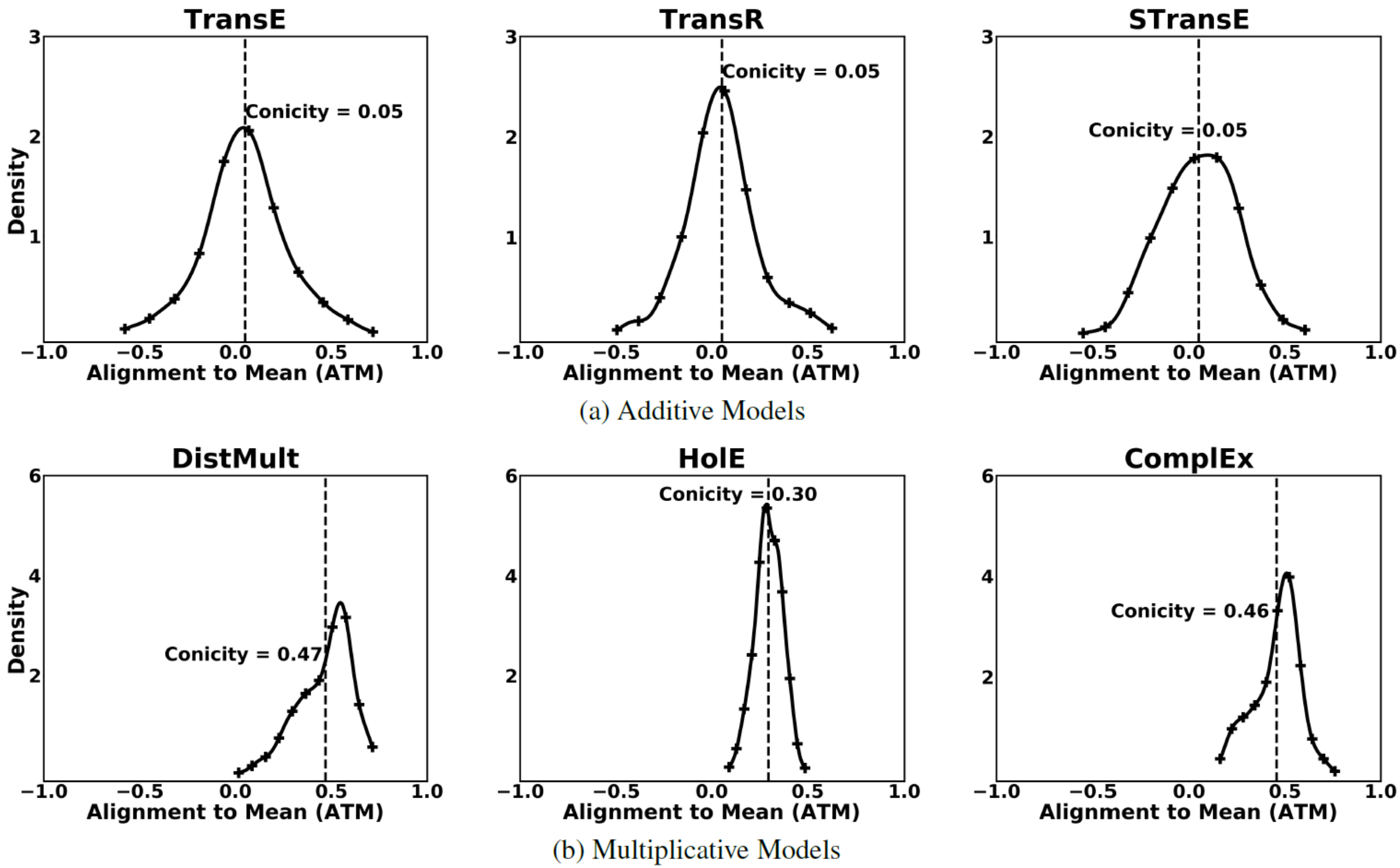
cf. 知識グラフの埋め込みとその応用[林克彦’17]
http://www.kecl.ntt.co.jp/icl/lirg/members/hayashi/files/stairs-hayashi17.pdf

# ベクトルの分析指標

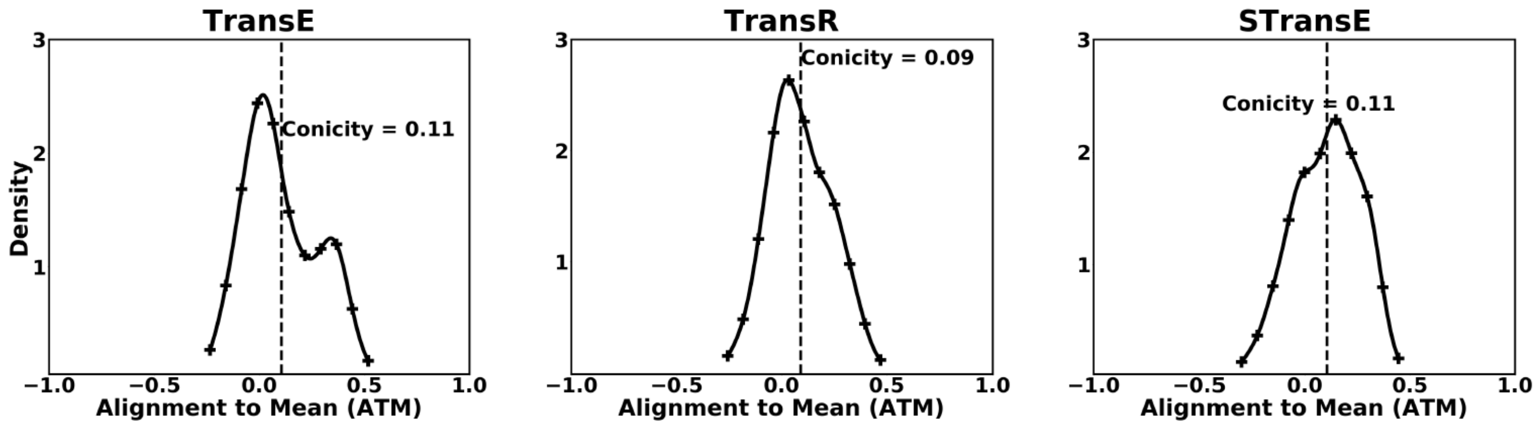
- 平均ベクトルとの類似度: $\mathrm{ATM}(\mathbf{v},\mathbf{V}) = \mathrm{cosine}\left(\mathbf{v}, \frac{1}{|\mathbf{V}|}\sum_{\mathbf{x}\in\mathbf{V}}\mathbf{x}\right)$
  (alignment to mean)
- ATMの平均(円錐形度合): $\mathrm{Conicity}(\mathbf{V}) = \frac{1}{|\mathbf{V}|}\sum_{\mathbf{v}\in\mathbf{V}}\mathrm{ATM}(\mathbf{v},\mathbf{V})$
- ATMの分散(vector spread): $\mathrm{VS}(\mathbf{V}) = \frac{1}{|\mathbf{V}|}\sum_{\mathbf{v}\in\mathbf{V}}\left(\mathrm{ATM}(\mathbf{v},\mathbf{V}) - \mathrm{Conicity}(\mathbf{V})\right)^2$
- ベクトルの平均長: $\mathrm{AVL}(\mathbf{V}) = \frac{1}{|\mathbf{V}|}\sum_{\mathbf{v}\in\mathbf{V}}\|\mathbf{v}\|_2$



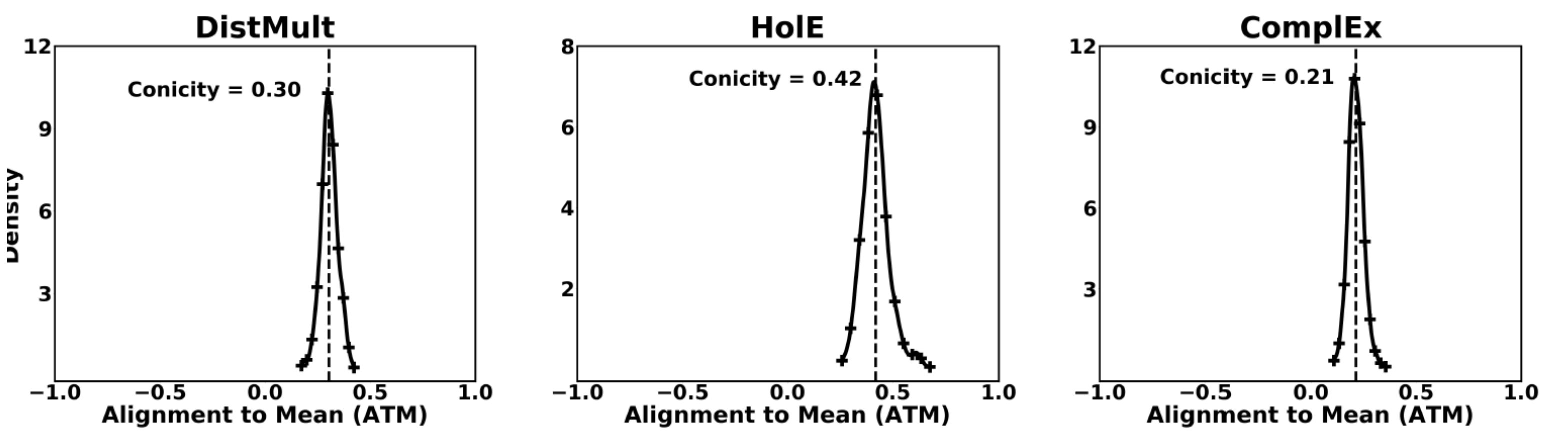
# 手法ごとの要素ベクトルのATM



# 手法ごとの関係ベクトルのATM




(a) Additive Models




(b) Multiplicative Models

# negative sampleの数の影響

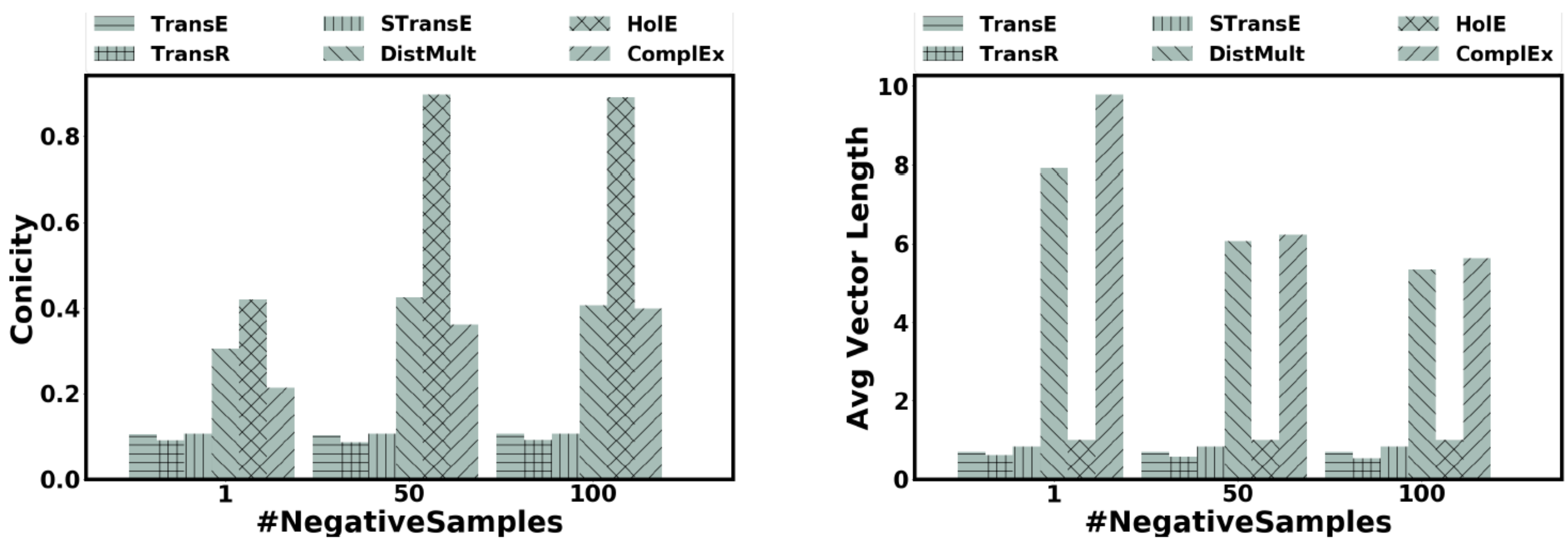



Figure 4: Conicity (left) and Average Vector Length (right) vs Number of negative samples for entity vectors learned using various KG embedding methods. In each bar group, first three models are additive, while the last three are multiplicative. Main findings from these plots are summarized in Section 6.2

# ベクトルの次元数の影響

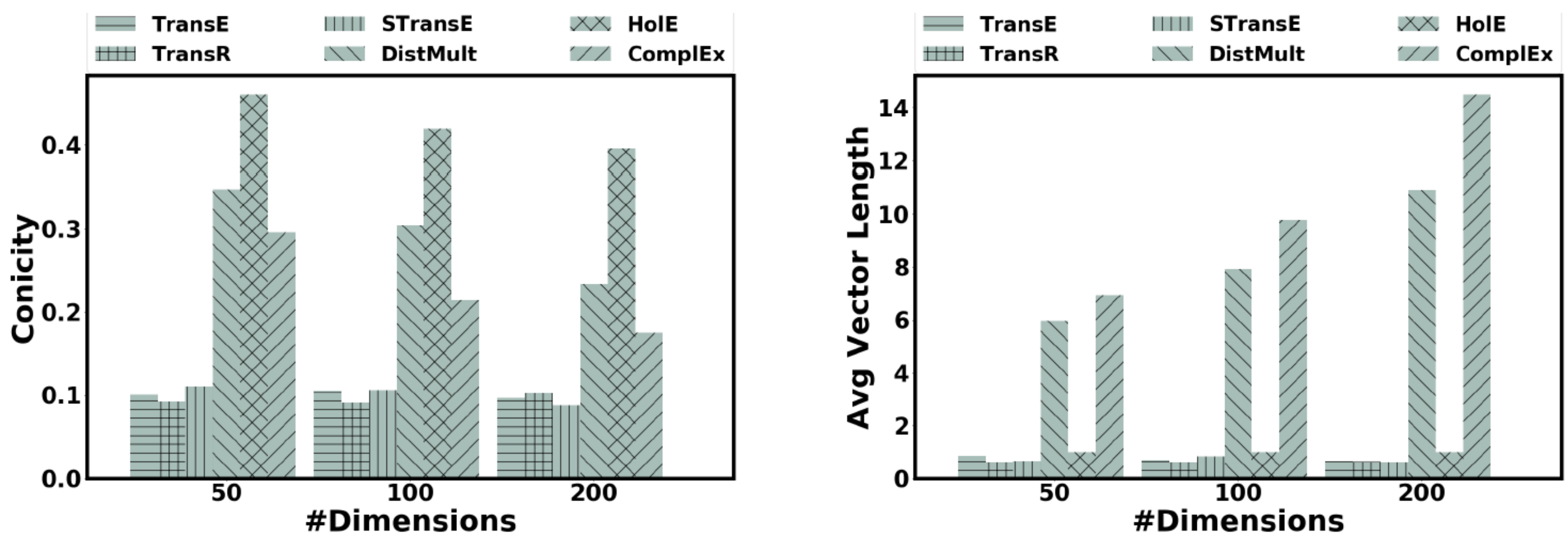



Figure 5: Conicity (left) and Average Vector Length (right) vs Number of Dimensions for entity vectors learned using various KG embedding methods. In each bar group, first three models are additive, while the last three are multiplicative. Main findings from these plots are summarized in Section 6.3.

# パフォーマンスとの関係

- In Figure 6 (right), for all multiplicative models except HolE, a higher average entity vector length translates to better performance, while the number of negative samples is kept fixed

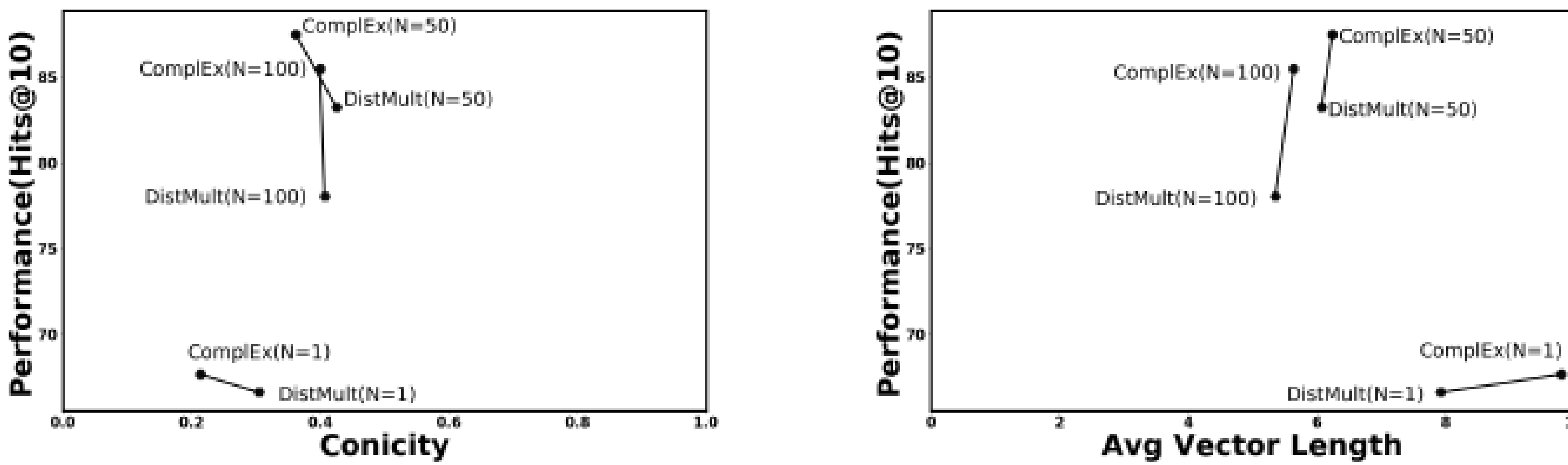



Figure 3: Relationship between Performance (HITS@10) on a link prediction task vs Conicity (left) and Avg. Vector Length (right) on FB15k dataset. For each point, $N$ represents the number of negative samples used. Models with same number of negative samples are connected by line segment. This demonstrates that model performance has negative correlation with Conicity while positive correlation with average vector length for fixed number of negatives. Main findings are summarized in Section 3.

# 読んでみた印象

- 「(最初に)やってみました」という印象が強い論文
- 少し面白いと感じた分析
  - Additive手法で得られたベクトルは、要素ベクトル、関係ベクトルともに0を中心に分布 ⇔ SGNS
  - Multiplicativeな手法で得られたベクトルの性質はSGNSで得られたベクトルの性質に類似(cf. $\mathbf{r}^{\top}(\mathbf{h} \odot \mathbf{t})$ )
- その他の考察はかなり微妙な印象(難しいとは言え…)
  - 異なる手法間で得られるベクトルの傾向が違うのは当然なので、単純に比較する意味がどのくらいあるのか？
  - パフォーマンスに関する分析は実質2事例から結論！