人物同一性を考慮した深層学習によるメディアコンテンツの変換生成

鈴木惇

東京大学情報理工学系研究科

1. 序論

画像や脚本に登場する複数の人物のように，複数の対象に絡んだデータを深層学習によって生成する際に，それらの対象を低次元空間上の表現を得るのは有効である。しばしば対象間には複数関係データが存在し，低次元表現はそれを反映するのが望ましい。

複数関係データとは，対象の集合 $\mathcal{E}$ と 関係の集合 $\mathcal{R}$，対象の間の関係の存在を表す三つ組 (h,r,t) の集合 $\mathcal{T}$ からなる。ただしここで $h,t\in\mathcal{E},r\in\mathcal{R}$ であり，$(h,r,t)\in\mathcal{T}$ は，対象 h から対象 t に関係 r が存在していることを意味する。

既存手法はユークリッド空間を用いてよい性能を発揮してきたが，これらはユークリッド空間を利用するがゆえに，階層的なデータに対する埋め込みに必ずしも向いているものではない。例えば，TransE (Bordes et al. 2013) は，h,r,t の埋め込み p_h,p_r,p_t はすべてユークリッド空間上の点であり，$(h,r,t)\in\mathcal{T}$ は埋め込まれた空間上では埋め込まれた点のベクトルの加算による関係 $p_h+p_r=p_t$ に対応するというモデルであるが，このモデルでは 1 対多の関係を正しく表現できない。TransH (Wang et al. 2014) では線型部分空間への射影を TransE のモデルに組み込み 1 対多の関係を表現したモデルを用いているが，一対多の関係が複数階層連なるような階層的なデータに対しては正しく関係を表現できない。

階層構造における関係としては，より対象数の多い下位階層から対象数の少ない上位階層へのな関係，若しくは，その逆の上位階層から下位階層への関係，あるいは，上位階層を共有するような同階層内の関係などが考えられる。これらを埋め込み空間で表現することを考える。まず，r が下位階層から上位階層への関係を表すとき，h は下位階層に存在し，t は上位階層に存在する。従って，$(h,r,t)\in\mathcal{T}$ であるとき， p_h を上位の方向へ一定距離動かせば p_t に到達するようなモデルが自然である。 一方 r が上位階層から下位階層への関係を表すとき，p_t を上位の方向へ一定距離動かせば p_h に到達するようなモデルが自然である。さらに上位階層を共有するような同一階層間の関係に対しては，p_h,p_t 双方を上位の方向へ一定距離動かせば互いに近づくようなモデルが自然である。ところが，上位階層の一つの対象を共用する下位階層の数が複数でしかも階層の数が多い場合，空間の広がり方が多項式的であるユークリッド空間では対象数の指数的な爆発に対応できず，逆に下位階層になっても対象数が増えていかない場合，多項式的に空間が広がっていくユークリッド空間では広すぎて過学習の恐れがある。

本研究では，関係の存在を埋め込み空間でベクトルの加算によって表すのではなく，それぞれの関係 r が最上位階層に対応する点 q_r を有し，$(h,r,t)\in\mathcal{T}$ であるときには p_h,p_t が q_r に向かってそれぞれ動いたときに互いに近づくものとして，ユークリッド空間とは限らない多様体上でモデリングを行う。本研究の提案手法はシンプルな目的関数で構成されており，また計算できる限りにおいていかなる多様体を用いることが可能である。また，確率的勾配降下法で効率的な分散計算による最適化が可能であり，スーパーコンピュータとの相性も優れている。

2. 手法

以下では埋め込みのための距離空間を $\mathcal{M}$ とおく。本研究では，各関係 r が最上位階層に相当する埋め込み空間上の点 q_r を有し，$(h,r,t)\in\mathcal{T}$ であれば， p_h,p_t をそれぞれ長さ $\ell_r^{\mathrm{H}},\ell_r^{\mathrm{T}}$ だけ q_r に動かしたときに互いに近づくようにモデリングを行う。この p_h,p_t から q_r への長さ $\ell_r^{\mathrm{H}},\ell_r^{\mathrm{T}}$ の移動先の点を $c_{p_h\to q_r}(\ell_r^{\mathrm{H}}),c_{p_t\to q_r}(\ell_r^{\mathrm{T}})$ とかく。三つ組 (h,r,t) に対する得点関数 f を

次のように設計する。

$$f(h,r,t) \coloneqq d_{\mathcal{M}}(c_{p_h \to q_r}(\ell_r^{\mathrm{H}}), c_{p_t \to q_r}(\ell_r^{\mathrm{T}}))$$

ここで $d_{\mathcal{M}}$ は $\mathcal{M}$ 上定義された距離関数である。このとき，$(h,r,t) \in \mathcal{T}$ ならば得点は小さく，さもなくば得点は大きくなるべきである。従って次のような損失関数を考える。

$$\sum_{(h,r,t)\in\mathcal{T}} [\delta + f(h,r,t) - f(h',r,t')]_+$$

ただしここで $[\cdot]_+$ は ReLU 関数であり, δ はマージンと呼ばれるハイパーパラメータである。また，h',t' はランダムにサンプリングされた対象である。このとき，$(h',r,t') \notin \mathcal{T}$ が高い確率で成り立つことを想定している。上記の損失関数を $(p_e)_{e\in\mathcal{E}}$ および $(q_r,\ell_r^H,\ell_r^T)_{r\in\mathcal{R}}$ について最適化を行う。損失関数は三つ組の得点関数の総和でかけているため，確率的勾配降下法を用いることが可能である。本研究では，各々の埋め込みが多様体上の点であることを考慮して，最適化には Riemannian Stochastic Gradient Descent (Zhang and Sra 2016) を用いる。

3. 実験

ベンチマークデータによる実験で性能を比較した。ここでは一部の結果のみを紹介する。実験ではデータセット WN18RR (Dettmers et al. 2017) に対して，与えられた三つ組に対し関係が存在するかを判別する問題において各手法の性能を評価した。まず訓練データを用いて各モデルを学習し，検証データを用いてスコア閾値ハイパーパラメータ選択を行い，テストデータに対して関係の有無を判別する判別問題において, 各々の誤判別率を性能評価指標として用いた。なお，提案手法は検証データを用いて多様体を球面・ユークリッド空間・双曲空間から選択している。紙面の都合上 2 次元の結果のみを示すと， TransE の誤判別率 0.2325 に対し提案手法の誤判別率は 0.2118 であり，提案手法の性能の高さが示された。

4. 結論

本研究では複数関係データに対し，最上位階層への動きを関係としてモデリングした埋め込みモデルを提案した。提案手法はデータの階層構造に応じた多様体を使用することが可能であるため，階層構造をうまく捉えた埋め込みが可能である。提案手法はベンチマークによる判別誤差評価において，ユークリッド空間を用いた既存手法を上回る結果を示した。

参 考 文 献

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, 2013.

Z. Wang, J. Zhang, J. Feng, and Z. Chen. Knowledge graph embedding by translating on hyperplanes. In AAAI Conference on Artificial Intelligence, 2014.

H. Zhang and S. Sra. First-order methods for geodesically convex optimization. In Conference on Learning Theory, 2016.

T. Dettmers, P. Minervini, P. Stenetorp, and S. Riedel. Convolutional 2d knowledge graph embeddings. arXiv preprint arXiv:1707.01476, 2017.