

Study on Continuous Representation of Relational Data

秦, 希望

<https://hdl.handle.net/2324/4784417>

出版情報 : Kyushu University, 2021, 博士 (数理学), 課程博士
バージョン :
権利関係 :

氏 名 : 秦 希望

論 文 名 : Study on Continuous Representation of Relational Data
(関係データの連続的表現に関する研究)

区 分 : 甲

論 文 内 容 の 要 旨

グラフ構造に代表される関係データは離散的なデータ構造である一方、連続的なデータは一般にベクトルで扱われる。関係データに対する連続的な表現とは、2つの対象の関係を復元・予測できるように対象や関係ごとにベクトルを割り当てることを指す。この表現により、関係データに含まれる欠損部分の予測や、対象のクラスタリング、情報検索などさまざまな応用に対してベクトル同士の演算を活用できる。本研究では関係データの表現手法に関し、一般性及び表現可能性に着目した研究を行なった。

本論文の前半部分では、一般性に関し、既存の表現手法を統一的に扱うためのフレームワークである **Nested Subspace Arrangement (NSS Arrangement)** を提案した。**NSS Arrangement** においては、グラフの点などの関係データの対象は集合列として表現され、関係は集合同士の包含関係によって表現される。複数の集合間の包含関係を参照することにより、**NSS Arrangement** は既存手法の多くを一般化することができる。また **NSS Arrangement** の応用例として、有向グラフの表現手法である **Disk-ANChor ARrangement (DANCAR)** を構成した。数値実験の結果、**DANCAR** は 20次元ユークリッド空間でほぼ完全に **WordNet** を再構成可能であった。また **DANCAR** を用いたグラフの可視化により、グラフの階層性やクラスター性といったグラフの特徴が反映されることも視覚的に検証した。

ある手法の表現可能性とは、その手法が表現できる関係データの多さのことを指す。例えばグラフの点をベクトルで表現し、内積で枝を復元するモデルの場合、有向グラフを表現できないという意味で表現可能性は低い。本論文の後半部分では、グラフの一般化である知識グラフに対する表現手法の表現可能性を解析した。知識グラフはグラフの枝にラベルがついたものであり、既存の表現手法の多くはラベルごとに点の埋め込みを操作することによって知識グラフを表現している。この操作として、既存手法では平行移動や回転などの操作が用いられているが、これらを用いて表現できる知識グラフには理論的限界が存在することが指摘されている。本論文では操作ごとの表現能力の差を明らかにすることにより、既存手法同士の実験的な性能の差を裏付けする。また、この分析をもとに新たな表現手法である **DiskE** を提案した。**DiskE** は知識グラフに内在する様々なパターンを表現に反映できるとともに、理論的には任意の知識グラフを正確に表現できる能力を持った表現可能性の高いモデルである。本論文内では知識グラフの欠損補完に関する数値実験により **DiskE** の有効性を検証したとともに、実験結果が知識グラフ内のパターンを表現していることも示した。