

(19)日本国特許庁(JP)

(12)公開特許公報(A)

(11)公開番号
特開2023-4969
(P2023-4969A)

(43)公開日 令和5年1月17日(2023.1.17)

(51)国際特許分類

G 0 6 N	5/02 (2023.01)	F I	G 0 6 N	5/02
G 0 6 N	20/00 (2019.01)		G 0 6 N	20/00
G 0 6 F	16/90 (2019.01)		G 0 6 F	16/90

テーマコード(参考)
5 B 1 7 5

【特許請求の範囲】**【請求項 1】**

オントロジー（202）によって強化される知識グラフ（200）の知識グラフ埋め込みモデル（208）を訓練するためのコンピュータ実装された方法であって、

当該方法は、

前記知識グラフ埋め込みモデル（208）を、前記第1の訓練クエリと、前記第1の訓練クエリの所定の応答とを用いて、

前記知識グラフ埋め込みモデル（208）における前記応答の埋め込みと、前記知識グラフ埋め込みモデル（208）における前記第1の訓練クエリの埋め込みとの間の距離を減少させるために、特に最小化するために、かつ、

前記応答の埋め込みと、前記知識グラフ埋め込みモデル（208）における第2の訓練クエリの埋め込みとの間の距離を減少させるために、特に最小化するために、

訓練すること（2）

を含み、

前記第2の訓練クエリは、前記オントロジー（202）に依存して前記第1の訓練クエリから決定される（1）

ことを特徴とする、方法。

【請求項 2】

前記オントロジーに従って無矛盾である、可能性のある単項の連続的なクエリの集合と、前記知識グラフ（200）のエンティティの集合及び関係の集合とを決定すること（1）と、

前記単項の連続的なクエリの集合から前記第1の訓練クエリを選択することと、
を特徴とする、請求項1に記載の方法。

【請求項 3】

当該方法は、前記第1の訓練クエリを所定のクエリ形状に従って決定することを含む
ことを特徴とする、請求項2に記載の方法。

【請求項 4】

当該方法は、

クエリを、特にランダムにサンプリングすることと、

前記オントロジー（202）を用いて前記クエリの汎化を決定することと、

前記汎化、特に前記汎化の特化から前記第2の訓練クエリを決定することと、
を含む

ことを特徴とする、請求項1乃至3のいずれか一項に記載の方法。

【請求項 5】

汎化深さを提供し、前記汎化深さまで前記クエリの汎化を決定すること、
及び／又は

特化深さを提供し、前記特化深さまで前記クエリの特化を決定すること
を特徴とする、請求項4に記載の方法。

【請求項 6】

前記知識グラフ埋め込みモデル（208）を用いて、連言クエリに対する応答を提供す
ること

を特徴とする、請求項1乃至5のいずれか一項に記載の方法。

【請求項 7】

前記第1の訓練クエリの前記埋め込みと、前記第1の訓練クエリに対する応答ではない
所定のエンティティの少なくとも1つの埋め込みとの間の距離を増加させるために、特に
最大化するために、及び／又は、

前記第2の訓練クエリの前記埋め込みと、前記第2の訓練クエリに対する応答ではない
所定のエンティティの少なくとも1つの埋め込みとの間の距離を増加させるために、特に
最大化するために、

前記知識グラフ埋め込みモデル（208）を訓練すること

10

20

30

40

50

を特徴とする、請求項 1 乃至 6 のいずれか一項に記載の方法。

【請求項 8】

オントロジー（202）によって強化される知識グラフ（200）の知識グラフ埋め込みモデル（208）を訓練するための装置（300）であって、

当該装置（300）は、請求項 1 乃至 7 のいずれか一項に記載の方法における各ステップを実施するように構成されている

ことを特徴とする、装置（300）。

【請求項 9】

コンピュータプログラムであって、

当該コンピュータプログラムは、コンピュータによって実行された場合に、請求項 1 乃至 7 のいずれか一項に記載の方法を前記コンピュータに実施させるためのコンピュータ可読命令を含む

ことを特徴とする、コンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

背景

本発明は、知識グラフ埋め込みモデルを訓練するための装置、コンピュータプログラム、及び、コンピュータ実装された方法に関する。

【背景技術】

【0002】

知識グラフ埋め込みモデルは、クエリに対する応答を提供するように訓練可能である。不完全な知識グラフに対するクエリに応答する体系的な手法を有する方法を提供することが望ましい。

【発明の概要】

【課題を解決するための手段】

【0003】

発明の開示

オントロジーによって強化される知識グラフの知識グラフ埋め込みモデルを訓練するためのコンピュータ実装された方法は、知識グラフ埋め込みモデルを、第1の訓練クエリと、第1の訓練クエリの所定の応答とを用いて、知識グラフ埋め込みモデルにおける応答の埋め込みと、知識グラフ埋め込みモデルにおける第1の訓練クエリの埋め込みとの間の距離を減少させるために、特に最小化するために、かつ、応答の埋め込みと、知識グラフ埋め込みモデルにおける第2の訓練クエリの埋め込みとの間の距離を減少させるために、特に最小化するために、訓練することを含み、第2の訓練クエリは、オントロジーに依存して第1の訓練クエリから決定される。第2の訓練クエリは、知識グラフにおける応答を有する所定のクエリの特化である。これにより、不完全な知識グラフに対する連言クエリ（conjunctive query）に応答するために、知識グラフ埋め込みモデルを訓練することが可能となる。訓練は、元々の知識グラフだけでなく、知識グラフに付随するオントロジーにも依存している。第2の訓練クエリの埋め込みは、第2の訓練クエリを決定するために使用されたオントロジー公理を考慮している。

【0004】

第1の訓練クエリをサンプリングするために、本方法は、オントロジーに従って無矛盾である、可能性のある単項の連続的なクエリの集合と、知識グラフのエンティティの集合及び関係の集合とを決定することと、単項の連続的なクエリの集合から第1の訓練クエリを選択することとを含み得る。訓練するためのクエリをランダムにサンプリングする代わりに、オントロジーに依存してクエリを戦略的にサンプリングのための方法が提供される。

【0005】

本方法は、第1の訓練クエリを所定のクエリ形状に従って決定することを含み得る。ク

10

20

30

40

50

エリをランダムにサンプリングする代わりに、所定のクエリ形状に従ったクエリが考慮される。ランダムに選択された例は、オントロジーに基づいて互いに関連するクエリを何ら含まない可能性がある。このことは、オントロジーに基づくサンプリングによって回避される。

【0006】

本方法は、クエリを、特にランダムにサンプリングすることと、オントロジーを用いてクエリの汎化を決定することと、汎化から、特に汎化の特化から第2の訓練クエリを決定することとを含み得る。汎化は、複数の特化を記述し、オントロジーに従って第1の訓練クエリに類似する多くの異なる訓練クエリを決定することを可能にする。

【0007】

本方法は、汎化深さを提供し、汎化深さまでクエリの汎化を決定すること、及び／又は、特化深さを提供し、特化深さまでクエリの特化を決定することを含み得る。このことは、実質的に、計算労力を制限するためである。

【0008】

本方法は、知識グラフ埋め込みモデルを用いて、連言クエリに対する応答を提供することを含み得る。

【0009】

本方法は、第1の訓練クエリの埋め込みと、第1の訓練クエリに対する応答ではない所定のエンティティの少なくとも1つの埋め込みとの間の距離を増加させるために、特に最大化するために、及び／又は、第2の訓練クエリの埋め込みと、第2の訓練クエリに対する応答ではない所定のエンティティの少なくとも1つの埋め込みとの間の距離を増加させるために、特に最大化するために、知識グラフ埋め込みモデルを訓練することを含み得る。このようにすると、訓練の結果、第1の訓練クエリの埋め込みは、第1の訓練クエリの非応答の埋め込みによりも、第1の訓練クエリの応答の埋め込みの方に近づくこととなる。

【0010】

オントロジーによって強化される知識グラフの知識グラフ埋め込みモデルを訓練するための装置は、本方法における各ステップを実施するように構成されている。

【0011】

コンピュータプログラムは、コンピュータによって実行された場合に、本方法をコンピュータに実施させるためのコンピュータ可読命令を含む。

【0012】

以下の説明及び図面から、さらなる有利な実施形態を導き出すことができる。

【図面の簡単な説明】

【0013】

【図1】知識グラフパターンを示す図である。

【図2】知識グラフのための知識グラフ埋め込みモデルを訓練するための方法を示す図である。

【図3】知識グラフのための知識グラフ埋め込みモデルを訓練するための装置を示す図である。

【発明を実施するための形態】

【0014】

知識グラフKGは、エンティティの集合及び関係の集合を含む。KGは、特定の関心ドメインに関するファクトを、エンティティの集合のうちの少なくとも1つのエンティティを用いてそのファクトを表現することによって記述し、エンティティの集合のうちのこの少なくとも1つのエンティティは、関係の集合のうちの少なくとも1つの関係を介して、エンティティの集合のうちの少なくとも1つの他のエンティティに相互接続されている。

【0015】

KG表現においては、エンティティは、KGのノードによって表現され、2つのエンティティの間の関係が、これらのノードの間のKGのエッジによって表現される。

10

20

30

40

50

【 0 0 1 6 】

ファクトは、主語 (subject) と、述語 (predicate) と、目的語 (object) とのトリプルである。KGにおいては、主語は、エンティティであり、目的語は、エンティティであり、述語は、関係である。

【 0 0 1 7 】

KGの知識グラフ埋め込みKGEにおいては、エンティティが埋め込みによって表現される。KGEにおいては、関係が埋め込みによって表現される。ファクトの主語の埋め込みと、述語の埋め込みと、目的語の埋め込みとのトリプルが、KGEにおけるファクトを表現する。

【 0 0 1 8 】

KGは、第1の所与のエンティティと第2の所与のエンティティとの間の関係を予測するために使用可能である。関係は、関係の集合からスコアに依存して選択可能である。スコアは、KGEにおける第1のエンティティの埋め込みと、KGEにおける第2のエンティティの埋め込みと、KGEにおける関係の埋め込みとをスコアにマッピングするスコア関数を用いて決定可能である。

【 0 0 1 9 】

埋め込みは、ベクトル空間内のベクトルであるものとしてよい。スコア関数を用いてスコアを決定することは、ベクトル和を決定することを含み得る。ベクトル和を決定することは、第1のエンティティを表現するベクトルに、関係を表現するベクトルを加算することを含み得る。スコアを決定することは、ベクトル和から第2のエンティティを表現するベクトルまでの距離を決定することを含み得る。

【 0 0 2 0 】

エンティティの埋め込みは、第1のベクトル空間内のベクトルであるものとしてよい。関係の埋め込みは、第2のベクトル空間内のベクトルであるものとしてよい。スコアを決定することは、第1のベクトル空間内の第1のエンティティを表現する第1のベクトルと、第2のベクトル空間内の第1のベクトルとのマッピングを決定することを含み得る。スコアを決定することは、第1のベクトル空間内の第2のエンティティを表現する第2のベクトルと、第2のベクトル空間内の第2のベクトルとのマッピングを決定することを含み得る。スコア関数を用いてスコアを決定することは、ベクトル和を決定することを含み得る。ベクトル和を決定することは、第1のベクトルに、第2のベクトル空間内の関係を表現するベクトルを加算することを含み得る。スコアを決定することは、ベクトル和から第2のベクトルまでの距離を決定することを含み得る。

【 0 0 2 1 】

一例においては、距離は、ユークリッド距離である。

【 0 0 2 2 】

KGとの関係を予測するために、2つの所与のエンティティを含む入力を、関係を含む出力にマッピングすることができる。関係は、関係の集合から選択可能である。一例においては、選択された関係は、関係の集合のうちの少なくとも1つの他の関係よりも高いスコアを結果的にもたらす。好ましくは、関係の集合において最高のスコアの関係を結果的にもたらすような関係が選択される。

【 0 0 2 3 】

ニューラルネットワークは、KGEを表現するように訓練可能である。ニューラルネットワークは、埋め込みのトリプルを含む訓練データを用いて訓練可能である。訓練データは、KGの真のファクトを表現するトリプルを含み得る。訓練データは、KGの真のファクトではないトリプルを表現するトリプルを含み得る。

【 0 0 2 4 】

ニューラルネットワークは、エンティティの集合のうちの第1の所与のエンティティの第1の埋め込みと、第2の所与のエンティティの第2の埋め込みとを、関係の集合のうちのそれぞれの関係ごとのスコアにマッピングするように訓練可能である。関係のスコアは、関係について、この関係が第1の所与のエンティティと第2の所与のエンティティとの

10

20

30

40

50

間の関係である確率を表す。

【 0 0 2 5 】

ニューラルネットワークは、所与のエンティティの埋め込みと、関係の集合のうちの所与の関係の埋め込みとを、エンティティの集合のうちのそれぞれのエンティティごとのスコアにマッピングするように訓練可能である。エンティティのスコアは、エンティティに関して、このエンティティが所与のエンティティに対する所与の関係を有するエンティティである確率を表す。

【 0 0 2 6 】

K G は、自然質問応答、ウェブ検索及びデータ分析のために広く使用されている。K G は、数百万のファクトに関する情報を格納している。

10

【 0 0 2 7 】

K G は、自動的に、半自動的に又は少なくとも部分的に手動により、例えばクラウドソーシング法を使用することによって構築可能である。

【 0 0 2 8 】

訓練においては、K G 又は K G E 、特にニューラルネットワークは、利用可能な知識を表現するために、訓練データを用いて訓練可能である。訓練データは、真のファクトを表現する正のトリプルと、正しくないファクトを表現する負のトリプルとを含み得る。

【 0 0 2 9 】

K G 又は K G E 、特にニューラルネットワークは、正のトリプル及び負のトリプルを用いて訓練可能である。

20

【 0 0 3 0 】

本方法は、正しいトリプル即ち正のトリプルと、正しくないトリプル即ち負のトリプルとを識別する。

【 0 0 3 1 】

K G は、ファクトに基づいた情報の、相互リンクされた集合体を表現する。K G は、(主語 ; 述語 ; 目的語)のトリプル、例えば(j o h n ; w o r k s A t ; b o s c h)の集合としてエンコード可能である。このようなトリプルのうちの主語又は目的語は、エンティティと称され、述語は、関係と称される。K G のトリプルの集合は、有向グラフとして表現可能であり、この有向グラフの頂点及び辺にはラベルが付されている。K G トリプルは、ファクトと称される。K G ファクトは、m a n (j o h n) , w o r k s A t (j o h n ; b o s c h) のように、単項又は二項の基底述語として表現可能である。

30

【 0 0 3 2 】

本例においては、K G G のシグネチャ

【 数 1 】

$$\Sigma_G = \langle \mathcal{E}, R \rangle$$

は、G に出現するエンティティ

【 数 2 】

$$\mathcal{E}$$

40

の集合及び関係 R の集合を定義する。シグネチャ Σ_G 内の関係 R は、述語、即ち、K G のエッジを表現する。シグネチャ Σ_G 内の定数は、エンティティ

【 数 3 】

$$\mathcal{E}$$

、即ち、K G のノード、又は、関係 R を表現する。

【 0 0 3 3 】

K G E は、K G のエンティティ及び関係を、ユーザ指定された次元 n を有する連続的なベクトル空間に埋め込むことに関する。より具体的には、K G E モデルは、K G トリプル

50

の集合を入力として受信し、KG構造を反映するいくつかの特徴が保持されるように、エンティティ及び関係をn次元のベクトル空間にマッピングすることを目的とする。これらの特徴は、それぞれのKEGモデルの目的関数によって捕捉される。このようにして、リレーションナルデータから数値ベクトルの集合が得られる。

【0034】

オントロジーは、公理の集合として表現される関心ドメインを概念化したものである。オントロジーは、KGが従うべきスキーム、例えば、

【数4】

$$\mathcal{O} = \left\{ \begin{array}{l} \text{worksAt}(X, Y), \\ \text{type}(Y, \text{company}) \rightarrow \text{type}(X, \text{employee}); \\ \text{managerAt}(X, Y) \rightarrow \text{worksAt}(X, Y) \end{array} \right\}$$

10

を反映している。

【0035】

第1の公理は、企業で働く人は従業員であると陳述し、その一方で、第2の公理は、関係managerAtが関係worksAtよりも特有であることを示す。

【0036】

本発明は、Artale, A., Calvanese, D., Kontchakov, R., Zakharyashev, M.: The DL-lite family and relations. CoRR abs/1401.3487 (2014)によるDL-lite_Aにおけるオントロジーに、ルールの右側の実存論的な制約なしに基づいている。サポートされる例示的なルール形式の概要は、

- (1) type(X, A) type(X, B)
- (2) p(X, Y) type(X, A)
- (3) p(X, Y) type(Y, A)
- (4) p(X, Y) s(X, Y)
- (5) p(X, Y) s(Y, X)

である。

【0037】

連言クエリは、 $q(X_1, X_2, \dots, X_k) \quad B$ 又は $\langle X_1, X_2, \dots, X_k \rangle \quad B$ の形式で表現したものであり、ここで、Bは、ルールにおいて定義されたボディであり、 X_1, X_2, \dots, X_k は、応答変数、即ち、クエリの応答を保持する変数である。単項のCQ(monadic conjunctive query)は、単一の応答変数を有するCQである。

【0038】

KG及びオントロジーオについて、CQに対する特定の応答は、KG及びオントロジーオの結果として生じる全てのファクトによって強化されたKGに対して得られる応答である。

【0039】

ユーザによって自然言語で定式化された情報ニーズは、例えば、Yahya, M., Berberich, K., Elbassuoni, S., Ramanath, M., Tresp, V., Weikum, G.: Deep answers for naturally asked questions on the web of data. In: Proceedings of the 21st World Wide Web Conference, WWW 2012, Lyon, France, April 16-20, 2012 (Companion Volume). pp. 445-449 (2012)に開示されている方法を使用して、そのような形式的なCQに翻訳される。例えば、人とその職場とに関する情報を格納しているKGの場合、ユーザは、Bosch社のいくつかのIT部門で働いている全ての人に興味がある可能性がある。形式的には、このようなクエリは、

$Q(X) \quad \text{workAt}(X, \text{bosch}) ; \text{employedIn}(X, Y) ; \text{type}(Y, \text{it_department})$

のように定式化されるべきであり、これは、単項のCQである。

20

30

40

50

【0040】

連言クエリは、KGパターンとして自然に表現可能である。上記のクエリに対応するKGパターンについて、例示的なKGパターンが図1に示されている。

【0041】

以下に、不完全なKGに対してCQに応答するための方法を説明する。本方法は、元々のKGのみならず、KGが付随するオントロジーにも依存している。

【0042】

本方法は、オントロジーに基づく訓練戦略を含み、オントロジー公理を考慮することを可能にする損失関数を拡張し、拡張された損失関数を使用する。このような損失関数は、例えば、Ren, H., Hu, W., Leskovec, J.: Query2box: Reasoning over knowledge graphs in vector space using box embedding. In: ICLR. OpenReview.net (2000)に記載されている。10

【0043】

エンティティは、d次元のベクトル空間内に点として埋め込まれる。

【0044】

クエリは、ボックスとして埋め込まれる。この文脈におけるボックスとは、d次元のベクトル空間内の、座標軸に平行な超長方形 (axis-aligned hyper-rectangle) を指す。

【0045】

($G ; Q_G$)についてのd次元の埋め込みは、20

【数5】

$$c \in \mathcal{E}$$

を $c \in R^d$ にマッピングし、かつ、 $q \in Q_G$ を

【数6】

$$q = (cen_q, off_q) \subseteq R^d \times R_{\geq 0}^d$$

にマッピングする関数 F であり、ここで、シグネチャ Σ_G は、KG G の所与のシグネチャ

【数7】

$$\Sigma_G = \langle \mathcal{E}, R \rangle$$
30

であり、 Q_G は、シグネチャ Σ_G に対する単項のCQの集合であり、 cen_q は、ボックスの中心であり、 off_q は、ボックスのオフセット cen_q である。ボックスの中心 cen_q までの距離 off_q の内部にある、d次元ベクトル空間の任意の点は、ボックスの内側にあるとみなされる。

【0046】

オントロジールール B_1, \dots, B_n H は、 B_1, \dots, B_n が保持されるときにはいつでも H を強制的に保持するように、KGEモデルを構成することができる場合に、KGEモデルに注入可能である。40

【0047】

本方法は、クエリボックスの内側の点の集合がCQの応答エンティティの集合に対応するという条件に従ってCQを埋め込むことを目的とする。それぞれのオントロジールールごとにルールの左手側及び右手側をクエリに変換することができるので、オントロジーをKGEモデルに注入することは、それぞれのクエリに対応するボックスの相互の包含を保証することに等しい。

【0048】

本方法を、図2を参照してさらに説明する。

【0049】

10

20

30

40

50

本方法への入力は、KG200及びオントロジー202である。任意選択肢として、以下に説明するようにクエリ深さ204が提供される。

【0050】

訓練は、正のサンプル206-1及び負のサンプル206-2に基づいている。訓練においては、KGEモデル208が訓練される。

【0051】

本方法は、ステップ1を含む。

【0052】

ステップ1は、正のサンプル及び負のサンプルを決定することを含む。

【0053】

正のサンプルは、訓練クエリと、その応答とを含む。訓練クエリは、以下に説明される訓練戦略に従って決定される。負のサンプルは、正のサンプルと同様に構造化されており、特に、KGからランダムにサンプリングされる。

【0054】

本例においては、複数の正のサンプル及び複数の負のサンプルが決定される。

【0055】

その後、ステップ2が実施される。

【0056】

ステップ2は、正のサンプル及び負のサンプルを用いてKGEモデル208を訓練することを含む。KGEモデル208は、本例においては、複数の正のサンプル及び複数の負のサンプルを用いて訓練される。

【0057】

訓練目的、特にKGEモデル208を訓練するための損失関数を、以下に説明する。

【0058】

ステップ2の結果として計算されたKGEモデルを使用して、CQに応答することができる。

【0059】

ステップ1において活用される訓練戦略は、KG200のシグネチャ_G内の述語及び定数を使用して形成することができる可能な単項のCQの集合Q_Gを考慮している。好ましくは、全ての可能な単項のCQが考慮される。

【0060】

即ち、本方法は、知識グラフ200に出現するエンティティの集合及び関係の集合を使用して、オントロジーによって、即ち、オントロジーに従って形成することができる可能な単項の連続的なクエリの集合Q_Gから、訓練クエリを決定することを含み得る。

【0061】

以下に、オントロジーによってガイドされるように訓練クエリを決定するための3つの例を説明する。

【0062】

特定の応答に基づくサンプリング：

特定の応答に基づくサンプリングは、クエリをランダムにサンプリングすることと、それらのクエリを、訓練において標準の応答ではなくそれらのクエリの特定の応答とともに使用することとを含む。

【0063】

オントロジーが、クエリ応答を効率的に実施することができる言語での公理を含む場合には、訓練クエリを、それら訓練クエリの特定の応答とともに生成することが、実際に実現可能である。特定の応答に基づくサンプリングのための例示的な言語は、Artale, A., Calvanese, D., Kontchakov, R., Zakharyaschev, M.: The dl-lite family and relations. CoRR abs/1401.3487 (2014)に開示されている。

【0064】

例として、KG200が、以下のファクト：

10

20

30

40

50

```

hasAlumnus(ul, pete);
worksFor(pete, ibm);
hasAlumnus(ul, john);
managesAt(john, bosch)

```

を格納していて、オントロジーオが、以下のルール：

```
managesAt(X, Y) worksAt(X, Y)
```

と、特にランダムに選択された所与のクエリ：

```
q1(X) hasAlumnus(ul, X) ^ worksFor(X, Y)
```

とを含み、そのクエリの特定の応答が、

```
{john, pete}
```

10

であると仮定する。

【0065】

本例においては、このクエリと、そのクエリの特定の応答とが、訓練ステップ2のための正のサンプルを定義する。

【0066】

クエリ書き換えに基づくサンプリング：

クエリ書き換えに基づくサンプリングは、ランダムにサンプリングされたクエリの集合に、それらのクエリの特化 (specialization) 及び汎化 (generalization) も追加する。クエリの特化は、オントロジーオを考慮して得られる。クエリの汎化は、オントロジーオを考慮して得られる。

20

【0067】

所与のクエリ q の特化は、以下においては $\text{Spec}(q)$ として表される。特化は、クエリ q に対する応答を構築するために有用である可能性のある情報を組み込んでいる。所与のクエリ q の汎化は、 $\text{Gen}(q)$ として表される。汎化は、もっともらしい欠落した応答である可能性のある追加的な関連するエンティティを組み込んでいる。

【0068】

クエリの汎化

【数8】

g

\rightsquigarrow

30

及び特化

【数9】

s

\rightsquigarrow

を取得するための例示的なルールは、

【数10】

40

50

(R1) $\text{type}(X, A) \rightarrow \text{type}(X, B) \in O$ ならば $\alpha \wedge \text{type}(T, B) =^s > \alpha \wedge \text{type}(T, A)$

$\alpha \wedge \text{type}(T, A) =^g > \alpha \wedge \text{type}(T, B)$

(R2) $p(X, Y) \rightarrow \text{type}(X, A) \in O$ ならば $\alpha \wedge \text{type}(T, A) \xrightarrow{S} \alpha \wedge p(T, Z)$

$\alpha \wedge p(T_1, T_2) \xrightarrow{g} \alpha \wedge \text{type}(T_1, A)$

(R3) $p(X, Y) \rightarrow \text{type}(Y, A) \in O$ ならば $\alpha \wedge \text{type}(T, A) \xrightarrow{S} \alpha \wedge p(Z, T)$

$\alpha \wedge p(T_1, T_2) \xrightarrow{g} \alpha \wedge \text{type}(T_2, A)$

(R4) $p(X, Y) \rightarrow s(X, Y) \in O$ ならば $\alpha \wedge s(T_1, T_2) \xrightarrow{S} \alpha \wedge p(T_1, T_2)$

$\alpha \wedge p(T_1, T_2) \xrightarrow{g} \alpha \wedge s(T_1, T_2)$

(R5) $s(X, Y) \rightarrow p(Y, X) \in O$ ならば $\alpha \wedge p(T_1, T_2) \xrightarrow{S} \alpha \wedge s(T_2, T_1)$

$\alpha \wedge s(T_1, T_2) \xrightarrow{g} \alpha \wedge p(T_2, T_1)$

(R6) $\theta \rightarrow \text{vars}(q) \cup E$ ならば $\alpha \wedge p(T_1, T_2) \wedge p(T'_1, T'_2) \in q \xrightarrow{S} \alpha \theta \wedge p(T_1, T_2) \theta$

s.t. $\theta(T_1) = \theta(T'_1)$ ならば $\alpha \wedge p(T_1, T_2) \text{ s.t. } T_1 \text{ または } T_2 \in E \xrightarrow{g} \alpha \wedge p(Z, T_2) \text{ または } \alpha \wedge p(T_1, Z)$

10

20

30

40

50

である。

【 0 0 6 9 】

例として、以下の公理：

`type(X, assistant_prof) type(X, professor);
teacherAt(X, Y) worksAt(X, Y)`

と、以下のクエリ：

`q(X) type(X, assistant_prof) ^ worksFor(X, Y)`
とが与えられた場合、第1のルールR1は、クエリqの汎化として

`q'(X) type(X, professor) ^ worksFor(X, Y)`
をもたらし、第3のルールR3は、クエリq'の特化として

`q''(X) type(X, assistant_prof) ^ teacherAt(X, Y)`

をもたらす。

【 0 0 7 0 】

本方法は、複数のクエリを、それらのクエリの汎化及び特化とともにランダムにサンプリングすることを含み得るものであり、複数のクエリと、それらのクエリの汎化及び特化とを使用して、訓練サンプルを構築することができる。

【 0 0 7 1 】

計算労力を軽減するために、本方法は、汎化深さ k_g 及び / 又は特化深さ k_s を提供することを含み得るものであり、この汎化深さ k_g 及び / 又は特化深さ k_s まで、訓練クエリが生成される。本例においては、本方法に入力されるクエリ深さ 204 が、汎化深さ k_g 及び / 又は特化深さ k_s を定義する。

【 0 0 7 2 】

ランダムなクエリに汎化及び特化を追加することによって、オントロジーの背景知識のいくつかの部分を捕捉することが可能となる。

【 0 0 7 3 】

オントロジーに基づく戦略的な訓練：

オントロジーに基づく戦略的な訓練は、オントロジーオに基づいて関連するクエリを見つけることを目的とする。

【 0 0 7 4 】

本方法は、オントロジーオに依存することによって訓練クエリを生成することを含み得る。

【 0 0 7 5 】

目標クエリの集合が、有向非巡回グラフ(N, E)によって定式化され、ここで、 N はノードの集合であり、 E は、有向エッジの集合である。このような有向非巡回グラフ DAG は、クエリ形状を捕捉する。この形状を、関係 R 及び定数の集合によってシグネチャ G からインスタンス化することができる。シグネチャ G は、まださらに決定されるべき関係 R 又はエンティティ

【数11】

 ε

のためのシンボル、例えば、関係シンボル又は変数をさらに含み得る。次いで、シグネチャ G 内のシンボルを KG のノード及びエッジに割り当てるためのラベル付け関数 f を適用することにより、目標クエリの集合が得られる。

10

【0076】

本例においては、クエリ形状 S は、タプル(N, E, A, n)であり、従って、(N, E)は、 DAG であり、 $n \in N$ は、 S の識別ノードであり、 $A \subseteq N$ は、 S におけるアンカーノードの集合を表す。シグネチャ G からの関係及び定数の所与の集合について、ラベル付け関数 f は、 $N \times E$ から $G \times V$ へのマッピングであり、ここで、 V は、変数の集合であり、従って、それぞれのアンカーノードは、定数にマッピングされ、それぞれの非アンカーノードは、変数又は定数のいずれかにマッピングされ、それぞれのエッジは、シグネチャ G 内の関係シンボルにマッピングされる。

20

【0077】

シグネチャ G 及びクエリ形状 S が与えられた場合、 CQ の集合は、

【数12】

$$Q_S^{\Sigma_G}$$

であり、ただし、

【数13】

$$q(f(n)) \leftarrow \exists f(n_1), \dots, f(n_k) \bigwedge_{e_i=(n,n') \in E} f(e_i)(f(n), f(n'))$$

30

である。

【0078】

本例においては、ラベル付け関数 f は、以下の集合：

【数14】

$$inv(p) = \{p' \mid p(X, Y) \rightarrow p'(X, Y) \in O\}$$

$$dom(p) = \{A \mid p'(X, Y) \rightarrow type(X, A') \in O \text{ s.t., } p(X, Y) \xrightarrow{G} * p'(X, Y) \text{かつ} type(X, A) \xrightarrow{G} * type(X, A')\}$$

$$range(p) = \{A \mid p'(X, Y) \rightarrow type(Y, A') \in O \text{ s.t., } p(X, Y) \xrightarrow{G} * p'(X, Y)\}$$

40

$$\text{かつ} type(X, A) \xrightarrow{G} * type(X, A')$$

$$follows(p) = \{p' \mid range(p) \cap dom(p') \neq \emptyset\}$$

$$inter_r(p) = \{p' \mid range(p) \cap range(p') \neq \emptyset \text{ または } p_1 \in inv(p') \text{かつ} dom(p_1) \cap dom(p_2) \neq \emptyset\}$$

$$inter_d(p) = \{p' \mid dom(p) \cap dom(p') \neq \emptyset \text{ または } p_1 \in inv(p), p_2 \in inv(p') \text{かつ} range(p_1) \cap range(p_2) \neq \emptyset\}$$

によって決定され、ここで、所与の関係 p について：

$inv(p)$ は、 p の全ての逆関係を含む集合である。

$dom(p)$ は、 p の複数のドメインタイプを含む集合である。 $dom(p)$ は、好みしくは全てのドメインタイプを含む。

50

`range(p)` は、`p` の複数の範囲タイプを含む集合である。`range(p)` は、好ましくは全ての範囲タイプを含む。

`follows(p)` は、`p` に続くことができる複数の関係 `p'` を含む集合である。`follows(p)` は、好ましくは全てのそのような関係 `p'` を含む。

`interr(p)` は、範囲内で `p` と交差することができる複数の関係 `p'` を含む集合である。`interr(p)` は、好ましくは全てのこののような関係 `p'` を含む。

`interd(p)` は、ドメイン位置において `p` と交差することができる複数の関係 `p'` を含む集合である。`interd(p)` は、好ましくは全てのこののような関係 `p'` を含む。

【0079】

10

本例においては、クエリ形状 `S` 及びオントロジー `O` について、エッジのそれぞれのペア `e = (n_1, n_2)`, `e' = (n_2, n_3)` について、

`f(e')` `follows(f(e))` である場合、又は、

`f(e) = type`, `f(n_2) = A`かつ`A dom(f(e'))` である場合、又は、

`p inv(f(e'))`かつ`A range(p)` である場合、又は、

`f(e') type`, `f(n_3) = A`かつ`A range(f(e))` である場合、又は、

`p inv(f(e))`かつ`A dom(p)` である場合には、

ラベル付け関数 `f` は、`O` について `S` に対して有効である。

20

【0080】

20

本例においては、クエリ形状 `S` 及びオントロジー `O` について、エッジのそれぞれのペア `e = (n_1, n_2)`, `e' = (n_3, n_2)` について、

`f(e')` `interr(f(e))` である場合、又は、

`f(e) = type`, `f(n_2) = A`かつ`A dom(f(e'))` である場合、又は、

`p inv(f(e'))`かつ`A range(p)` である場合、又は、

`f(e) = f(e') = type`, `f(n_1) = A_1`, `f(n_3) = A_2` であって、

かつ、

【数15】

30

$$A_1 \sqsubseteq^* A, A_2 \sqsubseteq^*$$

となるような概念 `A` が存在する場合には、

ラベル付け関数 `f` は、`O` について `S` に対して有効である。

【0081】

30

本例においては、クエリ形状 `S` 及びオントロジー `O` について、エッジのそれぞれのペア `e = (n_1, n_2)`, `e' = (n_1, n_3)` について、

`f(e')` `interd(f(e))` である場合、又は、

`f(e) = type`, `f(n_2) = A`かつ`A dom(f(e'))` である場合、又は、

`p inv(f(e'))`かつ`A range(p)` である場合、又は、

`f(e) = f(e') = type`, `f(n_2) = A_1`, `f(n_3) = A_2` であって、か

つ、

【数16】

40

$$type(X, A_1) \xrightarrow{g} * type(X, A) \text{かつ} type(X, A_2) \xrightarrow{g} * type(X, A)$$

となるようないくつかのエンティティ `A` が存在する場合には、

ラベル付け関数 `f` は、`O` について `S` に対して有効である。

50

【 0 0 8 2 】

上記の * 記号は、例えば、 $\text{type}(X, A) \quad \text{type}(X, A_1) \quad \text{type}(X, A'_1) \quad \text{type}(X, A'_1') \cdots \text{type}(X, A)$ のように、複数の公理を介して汎化を実施することができるることを反映している。

【 0 0 8 3 】

このようにして、意味論的に意味のあるクエリが作成される。

【 0 0 8 4 】

例として、クエリ形状

$S = (\{n_1 n_2 ; n_3\}, \{e_1 = (n_1 n_2), e_2 = (n_2, n_3) g\}, n_1)$

10

と、
 $f_1(e_1) = \text{worksAt}$ 、及び、
 $f_2(e_2) = \text{type}$ 、及び、
 $f_3(e_3) = \text{worksAt}$

をマッピングするラベル付け関数 f_1 とが与えられた場合、ラベル付け関数 $f_1(n_3) = \text{company}$ は、オントロジーオに關して有効であるが、その一方で、ラベル付け関数 $f_2(e_1) = \text{worksAt}$ 及び $f_2(e_2) = \text{teachesAt}$ は、 worksAt の範囲と、 teachesAt のドメインとが交差していないので、オントロジーオに關して有効ではない。

【 0 0 8 5 】

複数の訓練クエリを含む訓練集合は、例えば、それぞれのクエリ形状に対するそれぞれの有効なラベル付け関数を計算し、オントロジーオによって捕捉されないデータパターンを追加することによって構築される。この文脈における、オントロジーオによって捕捉されないデータパターンとは、その汎化又は特化であるものとしてよい。

20

【 0 0 8 6 】

本例においては、KGに対する応答を有している所与の形状 S の全てのラベル付けされたクエリは、単項の CQ の集合 Q_G である。オントロジーオを使用して、それぞれのクエリの汎化が、例えば、上記のルール R1 乃至 R6 によって決定される。取得された複数の訓練クエリは、本例においては、オントロジーオと、データパターンの集合とが与えられた場合に構成することができる全てのクエリを含む。

30

【 0 0 8 7 】

本方法は、それぞれのアンカーノードに対して、特にランダムに、有効なエンティティの一部を選択することを含み得る。即ち、アンカーとして選択されたそれぞれのエンティティに対して、結果として生じるクエリにより、KG200に対する特定の応答が生成される。

【 0 0 8 8 】

Ren, H., Leskovec, J.: Beta embedding for multi-hop logical reasoning in knowledge graphs. In: Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual (2020) は、目的関数を用いて知識グラフ埋め込みモデルを訓練するための例示的な訓練方法を開示している。

40

【 0 0 8 9 】

KGE モデル 208 は、この訓練方法に基づいている。この方法とは対照的に、訓練のための目的関数は、オントロジーオ公理も考慮する。

【 0 0 9 0 】

本例における方法は、オントロジーオに依存して、クエリの汎化及び / 又は特化を含む、クエリの表現を学習することを目的とする。

【 0 0 9 1 】

本例における方法は、クエリの埋め込みと、応答、即ち、正のサンプルを表現する埋め込みとの間の距離を減少させること、好ましくは最小化することを目的とし、その一方で

50

、クエリの埋め込みと、非応答、即ち、負のサンプルを表現する埋め込みとの間の距離を増加させること、特に最大化することを目的とする。

【0092】

本例においては、距離は、 box_q と v の間の L_1 距離として、クエリボックス

【数17】

$$box_q \in \mathbb{R}^d$$

と、エンティティベクトル

【数18】

$$v \in \mathbb{R}^d$$

10

との間の距離であり、即ち、

$$d(v, q) = \max(v - q_{\max}, 0) + \max(q_{\min} - v, 0)$$

であり、ここで、

$$q_{\max} = Cen_q + Off_q \text{かつ } q_{\min} = Cen_q - Off_q$$

である。

【0093】

本例においては、関数

20

【数19】

$$p(d; \gamma, \sigma) = e^{-(d+\gamma)^2/\sigma^2}$$

が、距離を(0, 1)の間隔に変換し、ここで、 γ は、マージンであり、 σ は、ペナルティの強さを制御する。

【0094】

本方法は、特定の訓練クエリ q が、オントロジー O に依存している他のクエリ q' の汎化である場合に、訓練クエリ q のためのボックス box_q が、この他の訓練クエリ q' のためのボックス $box_{q'}$ を含むことを保証することを目的とする。

30

【0095】

より一般的には、 a が訓練クエリ q に対する応答エンティティである場合、本方法は、応答 a と訓練クエリ q との間の距離だけでなく、応答 a と訓練クエリ q の特化との間の距離も最小化することを目的とする。

【0096】

一例においては、クエリの所与の訓練集合が、クエリの応答とともに提供され、同様にして、クエリの複数の汎化又はクエリの全ての汎化も提供される。

【0097】

例えば、クエリ q の全ての汎化の集合 $Gen(q) = \{q_1, \dots, q_n\}$ が、オントロジー O に基づいて決定される。次いで、オントロジー O に関して、訓練クエリ q と、この訓練クエリ q の特定の応答 v [G, O] とが与えられた場合、目的関数は、 v のための損失関数であるものとしてよい。損失関数の2つの例、即ち、第1の損失関数及び第2の損失関数を以下に提示する。

40

【0098】

第1の損失関数は、負の対数尤度：

【数20】

50

$$L = -\log(p(d(v, q); \gamma_1, \sigma_1)) - \beta \sum_{i=1}^n \log(p(d(v, q_i); \gamma_1, \sigma_1)) \\ - \sum_{j=1}^k \log(1 - p(d(v'_j, q); \gamma_2, \sigma_2))$$

であり、ここで、

【数21】

$v'_j \notin q[G, O]$

10

は、負のサンプリングによって得られたランダムなエンティティであり、 O は、固定されたスカラーである。

【0099】

第2の損失関数は、

【数22】

$$L = -\log \sigma(\gamma - d^*(v, q)) - \sum_{j=1}^k \log \sigma(d(v'_j, q) - \gamma)$$

20

であり、ここで、 $v \in q[G, O]$ は、 O に関する G に対する q の特定の応答、即ち、

【数23】

$v'_j \notin q[G, O]$

であり、 γ は、マージンであり、 $d^*(v, q) = q_i \text{Gen}(q) + d(v, q_i)$ 及び $0 \leq q_i \leq 1$ は、固定されたスカラーである。

【0100】

30

例として、オントロジー O は、以下のルール

$O = \{ \text{teachesAt}(X, Y), \text{worksAt}(X, Y); \text{type}(X, \text{assistant_prof}), \text{type}(X, \text{professor}) \}$

を含み得る。

【0101】

次に、 q の汎化の集合 $\text{Gen}(q) = \{q_1, q_2, q_3\}$ は、連言肢 $\text{teachesAt}(X, Y)$ を $\text{worksAt}(X, Y)$ に置き換えることによって q から得られた q_1 と、 $\text{type}(X, \text{assistant_prof})$ の代わりに $\text{type}(X, \text{professor})$ を有するクエリ q である q_2 と、 q_1, q_2, q_3 及びそれぞれの場合と同様である第1、第2及び第3の連言肢を有する q_3 と、を含む。

40

【0102】

オントロジーオによるKG G に対するクエリ q への特定の応答が $q[G, O] = \{p\}$ であるとすると、訓練目的は、関数 p の埋め込み v_p と、クエリ q の埋め込み $b_{o \times q}$ との間の距離も、埋め込み v_p と、クエリ q の汎化 q_1, q_2, q_3 に対応する埋め込み空間内のボックスとの間の距離も最小化することである。

【0103】

上で定義された2つの目的関数の一方と、上で説明された方法のうちの1つを使用して生成された正のクエリサンプル及び負のクエリサンプルとに依存して、知識グラフ埋め込みモデル208が訓練される。

【0104】

50

取得された知識グラフ埋め込みモデル208を使用して、オントロジーを備えた不完全なKGに対する連言クエリに応答することができる。

【0105】

知識グラフ200、オントロジー202及び／又は埋め込みモデル208は、機械状態、デジタル画像内のオブジェクトのプロパティ、又は、質問に対する応答に関することができる。

【0106】

知識グラフ200は、機械のステータスマッセージを機械状態にマッピングすることに関する知識を表現することができる。本方法は、ステータスマッセージを受信すること、ステータスマッセージに依存して機械状態を出力することとを含み得る。ステータスを表現する主語エンティティと、機械状態を表現する目的語エンティティとを含むトリプルが存在するかどうかを知識グラフ埋め込みモデル208を用いて予測することによって、状態を決定することができる。本方法は、機械状態を出力することとを含み得る。

10

【0107】

デジタル画像処理の場合、知識グラフ200は、画像のためのオブジェクト認識において認識されるオブジェクトの記述であるものとしてよい。知識グラフ200におけるエンティティは、オブジェクト及び／又はそのプロパティを表現することができる。本方法は、オブジェクトを受信することと、オブジェクトに依存して記述を出力することとを含み得る。

20

【0108】

ストリートビューにおいては、オブジェクトは、自動車、人間、家、又は、インフラストラクチャの他の部分であるものとしてよい。ストリートビューにおいては、知識グラフ200、オントロジー202及び／又は埋め込みモデル208は、オブジェクト、及び／又は、特にデジタル画像におけるオブジェクトと他のオブジェクトとの関係を記述することができる。本方法は、オブジェクトを受信することと、オブジェクトに依存して記述を出力することとを含み得る。

【0109】

本方法は、オントロジーによって強化される不完全なKGに対する複雑なクエリに応答するために使用可能である。本方法は、例えば製造ドメインにおけるデジタルツインの文脈において適用可能である。

30

【0110】

図3には、オントロジー202によって強化される知識グラフ200の知識グラフ埋め込みモデル208を訓練するための装置300の少なくとも一部が概略的に図示されている。装置300は、本方法における各ステップを実施するように構成されている。

【0111】

装置300は、少なくとも1つのストレージと、少なくとも1つのプロセッサとを含む。

【0112】

本例においては、ストレージ302は、KG200と、KGEモデル208と、オントロジー202と、正のサンプル206-1と、負のサンプル206-2とを格納するよう構成されている。

40

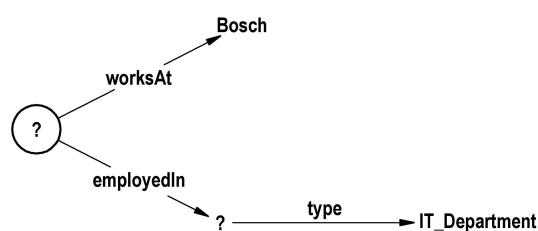
【0113】

本例においては、プロセッサ304は、上記の方法を実施するように構成されている。ストレージ302は、プロセッサ304によって実行された場合に、本方法をプロセッサ304に実行させるためのコンピュータ可読命令を格納することができる。プロセッサ304は、例えば、ストレージ302又はインタフェース（図示せず）からクエリ深さ204を受信するように構成可能である。

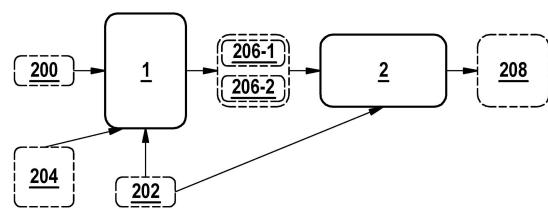
50

【図面】

【図1】

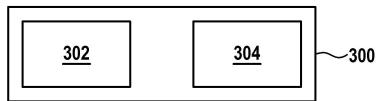


【図2】



10

【図3】



【外国語明細書】

2023004969000028.pdf

20

30

40

50

フロントページの続き

弁理士 永島 秀郎

(74)代理人 100162880

弁理士 上島 類

(72)発明者 チャバ ドモコス

ドイツ連邦共和国 ズイモツハイム ヘルダリンシュトラーセ 43 / 1

(72)発明者 ダリア ステパノワ

ドイツ連邦共和国 レオンベルク ヒンデンブルクシュトラーセ 68

(72)発明者 メディナ アンドレーゼル

オーストリア国 ウィーン シュプリンガーガッセ 27 / 11

(72)発明者 チュン キエン チャン

ドイツ連邦共和国 シュトゥットガルト ベーレンゼーシュトラーセ 9ベー

F ターム（参考） 5B175 DA01 EA01