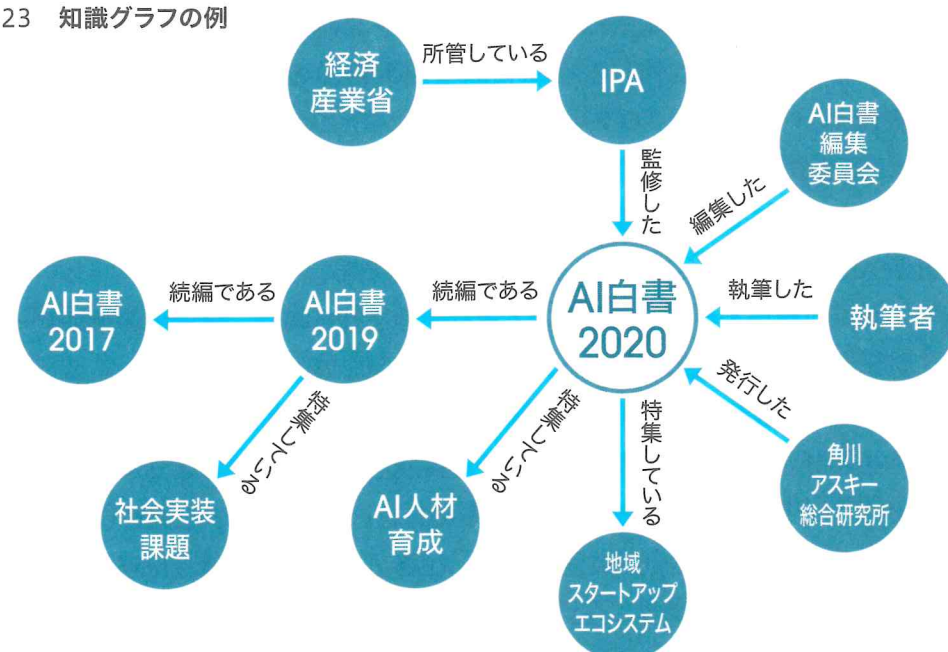


トリー予測(物語などのストーリーが途中まで与えられたとき、その後に起こることを予測する)などが課題として定義され、データセットやベンチマークの整備も進められている。

知識グラフ

「知識グラフ」とは、エンティティ間の「関係」を記述することで知識のネットワークをつくる手法である(図2-2-23)。主語、述語、目的語といった3つの情報の組(トリプル)を使って記述する。「皇居は、千代田区、にある」、「千代田区は、東京都に、ある」、「東京都は、日本の首都、である」といったトリプルを記述していくことで、グラフ構造(ネットワーク)を作成する。知識グラフに関しては、Google (Knowledge Graph)、Microsoft (Concept Graph)、Facebook (Entities Graph) などがインターネットなどから得られるテキスト情報などを基に構築した例がある。

■図2-2-23 知識グラフの例



(3) 最新技術動向

知識を使った推論(推論チャレンジ)

「知識グラフは、獲得した知識を、オントロジーを使って統合し、推論を適用することで新しい知識を導き出すものである」とも定義される^{※54}が、この「推論」部分に焦点を当てた技術開発も盛んに行われている。例えば人工知能学会で開催される「ナレッジグラフ推論チャレンジ」^{※55}では、推理小説「まだらの紐」に記載される事実(ファクト)を、推理小説の内容を記述するオントロジーを使ってRDF化し、この知識を使って推論をして、犯人を突き止める(新しい知識を導き出す)。

知識グラフを作成する際の問題点として、知識の資源自身が不完全である問題、新しい情報を追加する問題、誤った知識の混入問題などがあり、知識グラフ補完、オントロジーマッチング技術、誤り検知・修正技術が研究されている。

※54 Ehrlingerほか“Towards a Definition of Knowledge Graphs”,2016.

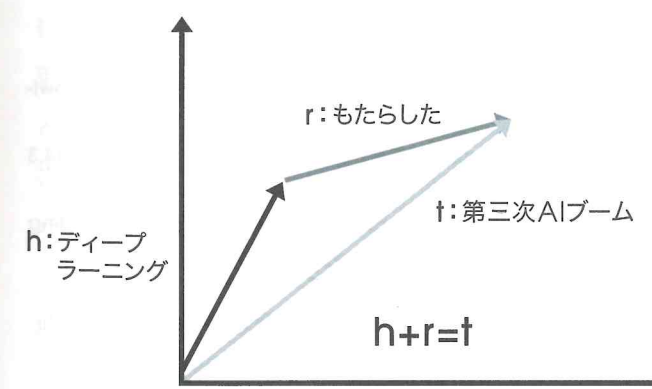
※55 <https://challenge.knowledge-graph.jp/2019/>

知識グラフの埋込みと補完

知識グラフ自体は、大量のエンティティと関係を含むため、このままでは扱いが難しい。また全ての情報が網羅されるわけでもない。既存の知識を活用して足りない知識を補完する技術が必要となる。これには知識グラフ補完(Knowledge Graph Completion)という方法が知られている。知識グラフ補完では、記述されている既知の情報を学習データとして用いて、グラフ構造に適したディープラーニングのアルゴリズムを活用することで、未知の関係の予測(Link Prediction)を行う。例えば、ある人物に対して、国籍の記述がなかったとしても、生まれた場所の記述があれば、その情報を元にして、国籍の予測ができる。言語における分散表現と同様に、知識グラフを特徴空間に「埋め込む」ことで、補完などを実現する方法が考えられた。例えば、主語、述語、目的語の関係をベクトル化し、主語と述語を与えたときに目的語(とその確率)を得られるようにすれば、知識グラフが持つ知識の利用が簡単になる。こうした知識グラフ埋込みモデル(knowledge graph embedding model)には、大きく「translation-basedモデル、bilinearモデル、neural network-basedモデルの3つ」^[5]がある。

translation-basedモデルとして代表的なTransE^[6]は、主語(エンティティ) h と述語(関係;リレーション) r 、及び目的語 t に対し h と t がリレーション r で関連していることを示すtriplet= $(h; r; t)$ が知識グラフに存在している場合、エンティティとリレーションを $h+r=t$ という式を満たすようにベクトル化する(ユークリッド空間に埋め込むという)(図2-2-24)。ここで、 h と r と t はそれぞれ、埋め込まれた h と r と t を表す^{※56}。例えば、知識の更新時にデータが十分でないとき、未知のエンティティを補完する技術として活用される。

■図2-2-24 TransEによるベクトル化の例



※56 蛭子・市瀬、知識グラフの補完におけるTranslation-based Modelsの発展と課題、SIG-SWO-044-03