概要

元論文

Predicting research trends with semantic and neural networks with an application in quantum physics セマンティックネットワークとニューラルネットワークによる研究動向の予測、量子物理学への応用 https://www.pnas.org/doi/10.1073/pnas.1914370116

元論文の引用論文

Machine Learning for Quantum Matter

量子力学のための機械学習

https://www.tandfonline.com/doi/full/10.1080/23746149.2020.1797528

量子力学的な性質を持つ物質の相を研究する「量子力学」は、ハード物性物理学、材料科学、統計力学、量子情報、量子重力、大規模数値シミュレーションなど、さまざまな分野から派生した研究分野である。近年、量子物質や強相関量子系に興味を持つ研究者が、それぞれの分野の発展に向け、最新の機械学習の基礎となるアルゴリズムに着目している。ここでは、合成実験データから従来の物質状態やトポロジー状態を認識するアルゴリズムから、ニューラルネットワークによる量子状態の表現とその量子系のシミュレーションや制御への応用まで、最近の機械学習の発展や量子物質研究の進展のための適応について簡単にレビューする。機械学習と量子多体系物理学の接点にある分野の今後の展開の見通しについて述べる。

Data-driven materials research enabled by natural language processing and information extraction

自然言語処理と情報抽出が可能にするデータ駆動型材料研究 https://aip.scitation.org/doi/abs/10.1063/5.0021106

データサイエンスや機械学習が社会のあらゆる場面で活用されるようになり、特に科学分野ではデータを取得することの重要性が高まっています。材料科学におけるデータは特に異質であり、探索される材料クラスや注目される材料特性の多様性に基づいています。このため、データは何桁もの幅があり、これらのデータは数値的なテキストや画像ベースの情報として現れることがあり、定量的な解釈を必要とします。自然言語処理の技術を応用して、領域を超えた科学文献を自動的に収集し、コード化する能力は、データサイエンスや機械学習に必要な豊富なデータセットを生成する上で非常に大きな可能性を持っている。このレビューでは、材料科学文献の自然言語処理とテキストマイニングの進展と実践に焦点を当て、論文の図や表に含まれるテキスト以外の追加情報を抽出する機会を明らかにする。データ編集、仮説の開発、分野内および分野間の傾向の理解など、材料に関する自然言語処理を追求するいくつかの理由について議論し、その例を挙げる。現在および将来の自然言語処理手法と、それらの材料科学への応用について詳述する。そして、今後の方向性が重要となる材料科学分野における自然言語処理とデータの課題について議論する。

Analysis and best parameters selection for person recognition based on gait model using CNN algorithm and image augmentation

CNNアルゴリズムと画像補強を用いた歩行モデルに基づく人物認識のための解析と最適なパラメータ選択 https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00387-6

歩行モデル(PRGM)と動作特徴に基づく人物認識は、その用途と、人物の姿勢の変化、人体の隠蔽、カメラの視点の変化などの重大な問題から、実に困難で新しいタスクである。このプロジェクトでは、深層畳み込みニューラルネットワーク(CNN)を修正し、歩行特徴に応じて画像補強(IA)技術を用いた人物認識のために適応させました。適応は、最適なCNNモデルを得るために、CNNのパラメータに最適な値を得ることを目的としている。CNNのパラメータに加え、CNNモデル自体の設計も最適化し、CNNの種類、層数、層間の正規化に影響を与える設計を行った。最適なパラメータと設計を選択した後、ディープラーニングアルゴリズムの学習に使用する異なる画像の数を増やすために、画像のコピーを多数使用して学習データセットのサイズを大きくするImage augmentationが使用されました。テストは、既知

のデータセット(Market dataset)を使って実現しました。このデータセットには、異なる歩行状態の人物の連続写真が含まれている。CNNモデルで行列として扱われる画像は、畳み込みによって多数の画像または行列に抽出されるため、データセットサイズは100倍にもなり、ビッグデータ問題となる。本プロジェクトでは、歩行モデルを用いて人物認識を行った結果、適応を行わないモデルに比べ、適応により認識精度が向上することを確認しました。また、データセットには、人が物を運んでいる画像も含まれています。IA技術により、画像の寸法(品質と解像度)、回転、持ち物などのバリエーションに対してロバストにモデルを改善することができました。200人認識の結果、検証精度はIA無しで約82%、IA有りでは96.23%であった。800人認識では、IAを用いない場合の検証精度は93.62%であった。

Developing a mathematical model of the co-author recommender system using graph mining techniques and big data applications

グラフマイニング手法とビッグデータ活用による共著者推薦システムの数理モデルの構築 https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00432-y

科学分野において効果的な研究を行うためには、最適な共著者を見つけることが最も重要な方法の1つです。データサ イエンスは、この可能性の実現に大きく貢献しています。本研究では、グラフマイニング技術とビッグデータ活用によ り、バイオインフォマティクスにおける共著者推薦システムの数理モデルを設計することを目的としています。本研究 では、応用開発研究法と混合研究法を採用した。研究対象はPubMedデータベースに登録されたバイオインフォマティ クスの全科学論文である。研究目的を達成するために、共著者を選択する際に最も適切で効果的な特徴を専門家が選択 し、優先順位をつけ、重み付けを行った。次に、グラフマイニング技術とビッグデータ活用により重み付けを行った。 最後に、バイオインフォマティクスにおける数学的共著者推薦システムモデルを提示した。データ分析手段としては、 ビッグデータサーバーにExpert Choice、Excel、Spark、Scala、Pythonの各プログラミング言語が用いられた。研究は 4つのステップで行われた。(1)AHPを用いた共著者選定に有効な基準の特定と優先順位付け、(2)アルゴリズムとビッグ データアプリケーションを用いてステップ1で得られた基準に基づく論文の相関度を決定、(3)数学的共著者推薦システ ムモデルの開発、(4)開発した数学モデルの評価、である。その結果、数学的共著者推薦システムモデルにおいて、雑誌 のタイトルと引用の基準が最も高い重みとなり、抄録が最も低い重みとなることが示された。また、提案モデルの精度 は72.26であった、このことから、コンテンツに基づく特徴量と専門家の意見を用いることで、最適な共著者を推薦で きる可能性が高いと結論づけられる、提案する共著者推薦システムモデルは、科学情報の様々なコンテキストにおい て、様々な分野の共著者を選択するための適切な推薦を行うことができると期待されます。このモデルの最も重要な革 新性は、専門家の意見とシステム的な重みを組み合わせることで、共著者の発見を加速し、その結果、時間を節約し、 科学製品の品質をより高くすることができることである。

Combining deep neural network and bibliometric indicator for emerging research topic prediction

ディープニューラルネットワークとビブリオメトリック指標の組み合わせによる新興研究トピック予測 https://www.sciencedirect.com/science/article/abs/pii/S0306457321001072

新たな研究テーマを予測することは、研究者や政策立案者にとって重要である。本研究では、新興トピック予測の問題 に対して、2段階の解決策を提案する。第一のステップは、トピック候補の将来の人気スコア(インパクトと成長性を 反映する新規指標)を時系列に予測する。第二のステップでは、第一のステップで予測された人気候補から新規トピッ クを選択する。トピック候補には、ドメイン特性を持つ単語を使用します。ディープニューラルネットワーク、具体的 にはLSTMとNNARをトピックの9つの特徴量を用いて適用し、人気度スコアを予測する。2つのデータセットでモデル と5つのベースラインを、(1)正しい指標値の予測能力、(2)最適なランキング順の再構築能力、という2つの観点から評価 した。また、全てのトピックを用いてモデルを学習するグローバル戦略と、異なるトピックのグループを用いて別々の モデルを学習するローカル戦略の2種類の学習戦略を比較した。その結果、LSTMとNNARはMAEとRMSEで測定される 人気度スコアの値の予測において他のモデルを上回り、LightGBMはNDCG@20の観点からトピックのランキングにおい て競争力のあるベースラインであることが示された。グローバル戦略とローカル戦略の性能差は大きくない。我々の手 法で予測された新興トピックと他の手法で予測されたトピックを比較した。機械学習により予測されたトピックは、ル ールベースモデルにより予測されたトピックと比較して、より類似していることが示唆された。また、予備的な文献分 析により、いくつかの重要なトピックがノミネートされた。本研究では、機械学習と計量書誌学的手法の両方の長所を 生かし、新興トピックの予測を行った。ディープニューラルネットワークは、最適化目標を定義し測定することが可能 である。ビブリオメトリック指標は、候補から新規トピックを選択する効率的な方法を提供する。このハイブリッドア プローチは、新興トピックの様々な特性を考慮して予測を行うことに期待が持てる。

Embedding technique and network analysis of scientific innovations emergence in an arXiv-based concept network

arXivベースの概念ネットワークにおける科学技術イノベーション創発の埋め込み手法とネットワーク分析 https://ieeexplore.ieee.org/abstract/document/9204220

イノベーションと発見には、新しさがつきものです。このようなプロセスは、新しいアイデアの出現、あるいは既存のアイデア間の非定型的な接続の出現と考えることができる。このような接続の重要性は、アイデアの空間におけるネットワークまたはグラフ表現によってイノベーションを調査するためのヒントとなる。このような表現では、グラフのノードが関連する概念(アイデア)に対応し、2つのノード間のエッジは対応する概念が共通のコンテキストで使用されていることを意味する。本研究では、イノベーションを生み出す可能性のある既存概念間のエッジを特定する可能性についての問題に取り組む。この目的のために、2007年4月から2019年9月までの120万件のarXiv.org原稿の科学的知識景観を使用する。我々は、ScienceWISE.infoプラットフォームを使用して、それらに関連する概念を抽出します。複雑ネットワーク科学とグラフ埋め込みで開発されたアプローチを組み合わせて、イノベーションが現れる可能性のある科学的知識ランドスケープ上のエッジ(リンク)の予測可能性を議論する。

Potential index: Revealing the future impact of research topics based on current knowledge networks

ポテンシャルインデックス 現在の知識ネットワークから、研究テーマの将来的な影響力の明確化 https://www.sciencedirect.com/science/article/abs/pii/S1751157721000365

科学論文の量が急速に増加している現在、科学者や研究機関にとって顕著な研究動向を把握することは重要である。これまでにも多くの研究者が科学研究の現状を把握することを試みてきたが、潜在的な影響力を持つ研究テーマを明らかにするための努力はもっと必要である。本研究では、ある研究トピックの科学的影響力とその知識ネットワークの構造との関係を調査する。構造情報に基づいてトピックの影響度をモデル化するために、ポテンシャル指数という新しい指標を提案する。この指標は、知識の新規性と多様性という2つの要素からなる即時的な指標であり、その運用には、betweenness centralityとnetwork entropyの概念を用いている。実証実験の結果、ポテンシャル指数は高いR2と正の相関を持ち、将来のトピックインパクトの良好な予測因子として機能することが示された。その優位性は回帰モデルの入力特徴量として用いた場合にも維持される。さらに、提案する指標はより良い結果を達成し、モデルの複雑性が増すにつれて他の特徴量との差が顕著になる。また、提案指標の変化を説明するために、トピックの進化過程に関する定量的・定性的な分析も行っている。本研究は、トピックのインパクトと知識構造の間に明示的な関係を確立することにより、科学的インパクトモデリングの研究に貢献し、研究トピックの潜在的インパクトを予測する上で有用である。

Network of scientific concepts: empirical analysis and modeling

科学的概念のネットワーク: 実証的分析とモデル化

https://arxiv.org/abs/2108.03962

科学のある領域における概念は、知識の構造を反映した本質的なつながりで結ばれている。この構造の定性的な洞察と定量的な記述を得るために、物理学の領域における科学的概念のネットワークの実証分析とモデリングを行う。この目的のために、我々はe-printリポジトリarXivに投稿された原稿のコレクションと、このhttp URLプラットフォームを介して収集された科学的概念の語彙を使用し、出版物におけるそれらの共起に基づいて科学的概念のネットワークを構築します。その結果得られた複雑なネットワークは、一般的に用いられるいくつかのネットワークモデルによる単純な成長の結果としては理解できない多くの特異な特徴(高いノード密度、異相性、構造的相関、歪んだノード次数分布)を持っている。我々は、ブロックによる成長と優先的選択という2つの要因の同時説明に基づくモデルが、経験的に観察される概念ネットワークの特性を説明することを示す。

Embedding-based Detection and Extraction of Research Topics from Academic Documents Using Deep Clustering

深層クラスタリングを用いた埋め込み型学術文書からの研究トピックの検出と抽出 https://sciendo.com/es/article/10.2478/jdis-2021-0024

研究分野やトピックを発見し、そのダイナミクスを理解することは、科学者コミュニティが科学分野の確立を決定する際に役立ちます。また、政府や企業とのより良い協力関係を築くのにも役立つ。本研究は、研究分野の時間的な発展を

調査し、それをトピック検出問題に変換することを目的とする。この目的を達成するために、我々は学術文書の抄録とタイトルから研究動向を検出するための修正ディープクラスタリング手法を提案する。文書をベクトルベースの表現に変換するために、文書埋め込みのアプローチを利用する。提案手法は、ベンチマークデータセットに対して、異なる埋め込みとクラスタリングアプローチの組み合わせや、古典的なトピックモデリングアルゴリズム(LDAなど)と比較することで評価される。また、ケーススタディとして、人工知能(AI)の進化を調査し、関連するAI出版物における研究トピックやサブフィールドを検出することも行う。本研究では、トピック抽出のために、Doc2Vec表現と結合した修正・調整された深層埋め込みクラスタリングを紹介する。また、本研究では、ラベリングアプローチとして概念抽出法を用いる。本手法の有効性は、過去30年間のAIトピックを分析したAI論文のケーススタディで評価されている。

A mechanism for evolution of the physical concepts network

物理概念ネットワークの進化メカニズム

https://arxiv.org/abs/2106.01022

我々は、論文中の共起に基づいて異なる科学的概念間のつながりを反映する複雑なネットワークである、概念のネットワークの成長を支配する根本的なメカニズムを提案する。この目的のために、我々はarXiv.orgに投稿された物理学のプレプリントに基づく概念のネットワークの実証分析を行う。我々は、ネットワークの特性を計算し、それらが一般的に使用されるいくつかの単純なネットワーク成長モデルの結果として従うことができないことを示す。さらに、ブロックによる成長と優先的選択という2つの要因を同時に説明することで、経験的に観測されている概念ネットワークの特性を説明できることを示唆する。さらに、観測された構造はこれら二つの要因の相乗効果として現れ、それぞれ単独では満足のいく図式を導き出せない。

Are AI ethics conferences different and more diverse compared to traditional computer science conferences?

AI倫理学会は従来のコンピュータサイエンス学会と異なり、多様性があるのか? https://osf.io/75gfs

コンピュータサイエンス(CS)には歴史的に性別や人種の表現が欠けていましたが、そのAI研究は最終的にはすべての人に影響を及ぼします。 CS会議に部分的に根ざしているため、FAccTやAIESなどの「AI倫理」(AIE)会議は、AIの社会的影響について話し合い、解決策を提案する別個の場所になりました。ただし、これらの会議が従来のCS会場の歴史的な表現の問題を改善するかどうかはほとんど不明です。この作業では、AIE会議の進化を調査し、人口統計学的特性、出版物の内容、および引用パターン全体でそれらを比較します。 AIE会議は、内部のトピックの多様性と他のCS会議への影響を増大させていることがわかります。重要なのは、AIE会議は非常に差別化可能であり、他の場所では表現されていないトピックをカバーしていることです。しかし、おそらくこの分野の願望に反して、白人の作家がより一般的であり、年功序列と黒人の研究者はCS会場と同様に表されます。私たちの結果は、AIE会議は、特にCSにかなりのルーツがあることを考えると、より多様な著者を引き付けるための取り組みを増やす可能性があることを示唆しています。

Earth, wind, (water), and fire: Measuring epistemic boundaries in climate change research

地球、風、(水)、および火: 気候変動研究における認識論的境界の測定 https://www.sciencedirect.com/science/article/abs/pii/S0304422X21000577

科学研究は、厳格な懲戒規範と象徴的な境界によって支配されています。この高度に構造化されたコンテキストは、どのような調査が行われる可能性が高いかを決定する可能性のある空間です。学際性は、独創性を促進するために分野間の認識論的ギャップを埋めることによって、これらの分野の規範を混乱させる可能性があります。しかし、学際性は実際に時間の経過とともに可能性のある空間を拡大することができますか?この論文では、2000年から2019年までの30,228の科学的要約に基づいて、気候変動研究の学際的分野の進化を分析します。パースのセミオティックの観点から解釈された新しい単語埋め込み技術を使用して、分野の学際的ダイナミクスとエピステミック境界を調べます。時間とともに。過去20年間で学際性が高まっているにもかかわらず、認識論的境界はより定着し、研究はより予測可能になります。学際性は境界を侵食するのではなく、高度に構造化され制約された可能性のある空間内で機能すると結論付けます。

Research trend prediction in computer science publications: a deep neural network approach

コンピュータサイエンスの出版物における研究動向予測: ディープニューラルネットワークアプローチ https://link.springer.com/article/10.1007/s11192-021-04240-2

毎日何千もの研究論文が発表されており、これらすべての研究の中で最も急速に成長している分野の1つはコンピュータサイエンス(CS)です。したがって、この特定の研究分野でどの研究分野がトレンドになっているのかを知ることは、かなりの数の学者、研究機関、および資金提供組織にとって有利です。現在のCSの傾向を分析し、結果としてさまざまな視点から将来の傾向を予測することに焦点を当てた多くの科学計量研究が行われてきました。この膨大な数のCS出版物からの大規模なデータセットと、このようなビッグデータの問題における深層学習手法の力にもかかわらず、深層ニューラルネットワークはまだこの分野で十分に活用されていません。したがって、この論文の目的は、長期短期記憶ニューラルネットワークを使用して、今後数年間のCSトレンドを予測することです。したがって、1940年のCS論文と、マイクロソフトアカデミックグラフデータセットの対応する研究分野が、この研究傾向予測の問題を解決するために活用されてきました。次に、提案された方法の予測精度は、RMSEおよび決定係数(R2)メトリックを使用して評価されます。評価は、提案された方法が、考慮されたすべての期間で予測精度の点でベースラインアプローチよりも優れていることを示しています。続いて、提案手法の予測を採用し、コンピュータサイエンス研究の今後のトレンド分野をさまざまな観点から調査します。

Joint Content-Context Analysis of Scientific Publications: Identifying Opportunities for Collaboration in Cognitive Science

科学出版物の共同コンテンツ-コンテキスト分析: 認知科学におけるコラボレーションの機会の特定 https://openreview.net/forum?id=yzVECygEpF

この作品は、認知科学の分野で出版物を研究し、数学的手法を利用して、論文の内容(要約)の分析を文脈(引用、ジャーナル)に結び付けます。 抄録に階層的トピックモデリングを適用し、引用ネットワークのコミュニティ検出アルゴリズムを適用し、コンテンツとコンテキストの不一致を測定して、類似したトピックを研究しているが、互いに引用したり、同じ場所で公開したりしない学術分野を見つけます。 これらの結果は、認知科学や機械学習などの高度に学際的な分野での科学的コラボレーションの機会を特定するための有望で体系的なフレームワークを示しています。

A Method to Predict Semantic Relations on Artificial Intelligence Papers

人工知能論文の意味関係を予測する方法

https://ieeexplore.ieee.org/abstract/document/9671315

大規模に進化するネットワークでのリンクの出現を予測することは、多くの実用的なアプリケーションでは困難な作業です。最近、Science4castコンテストは、64.000のAIコンセプトのネットワークを提示し、参加者に将来一緒に研究されるトピックを予測するように求めるこの課題を示しています。この論文では、ディープラーニングアプローチの新しいファミリーであるグラフニューラルネットワークに基づいてこの問題の解決策を提示します。チャレンジの結果は、計算効率を上げるために厳しい制限を課す必要がある場合でも、ソリューションが競争力があることを示しています。倹約的なモデル:グラフの固有のダイナミクスを無視し、ターゲットリンクを囲むノードの小さなサブセットのみを使用します。この論文で提示された予備実験は、モデルが2つの関連する、しかし異なるパターンを学習していることを示唆しています。サブグラフによるノードの吸収と、より密度の高いサブグラフの結合です。このモデルは、最初のタイプのパターンの認識に優れているようです。

Dynamic Embedding-based Methods for Link Prediction in Machine Learning Semantic Network

機械学習セマンティックネットワークにおけるリンク予測のための動的埋め込みベースの方法 https://ieeexplore.ieee.org/abstract/document/9672040

この論文は、意味ネットワークにおけるリンク予測を研究することにより、科学的発見を加速することを目的としています。 ノードは機械学習では未確認の概念であり、タイムスタンプ付きのエッジは科学論文での共起を示しています。この時間情報を利用して、1994年から2017年まで毎年グラフにノード埋め込みを実行し、2つの方法を適用してノードペアの特徴を見つけます。1つはトランスフォーマーを使用し、もう1つは既知のリンクと組み合わせた距離メトリックを使用します。 予測機能。 3層多層パーセプトロンを使用した後者の特徴抽出手法は、2020年のグラフのエッジの予測で0.902のAUCを達成しました。 結果として得られる特徴を調べると、モデルは実際に特徴の動的な性質に注意を払っ

ていることがわかります。たとえば、埋め込みスペースのノードペアの距離が何年にもわたってどのように変化するかなどです。

Determining Research Priorities for Astronomy Using Machine Learning

機械学習を使用した天文学の研究優先順位の決定

https://iopscience.iop.org/article/10.3847/2515-5172/ac4990/meta

機械学習技術が科学の戦略的計画を強化できるかどうかについての最初の探索的調査を要約します。 影響力の大きい天文学ジャーナルから抜粋した要約を使用した潜在的ディリクレ配分に基づくアプローチは、研究トピックへの将来の関心の主要な指標となる可能性があることがわかりました。 2010年の全米アカデミーズの天文学および天体物理学の10年間の調査で特定された、優先度の高い研究分野とよく相関する2つのトピック指標を示します。 1つのメトリックは、すべての科学論文(「カウント」)による各トピックへの部分的な貢献の合計に基づいており、もう1つは、カウントの複合年間成長率です。 これらの同じメトリックは、同じ10年間の調査に提出されたホワイトペーパーと同じ程度の相関関係も示しています。 私たちの結果は、10年調査が急成長している研究を過小評価している可能性があることを示唆しています。 私たちの仕事の予備バージョンは、Thronsonらによって発表されました。

Predicting Research Trends in Artificial Intelligence with Gradient Boosting Decision Trees and Time-aware Graph Neural Networks

勾配ブースティング決定木と時間認識グラフニューラルネットワークを使用した人工知能の研究動向の予測 https://ieeexplore.ieee.org/abstract/document/9671505

Science4cast 2021コンテストは、進化するセマンティックネットワークの将来のエッジを予測することに焦点を当てています。各頂点は人工知能の概念を表し、頂点のペア間のエッジは、2つの概念が科学論文で一緒に調査されたことを示します。 このホワイトペーパーでは、この競争に対するソリューションについて説明します。 ツリーベースの勾配ブースティングアプローチとディープラーニングアプローチの2つの異なるアプローチを提示し、両方のアプローチが競争力のあるパフォーマンスを達成することを示します。 2つのアプローチを組み合わせた最終的なソリューションは、参加したすべてのチームの中で1位を獲得しました。 このペーパーのソースコードは、

https://github.com/YichaoLu/Science4cast2021で入手できます。

Prediction of Research Trends using LDA based Topic Modeling

LDAベースのトピックモデリングを使用した研究動向の予測

https://www.sciencedirect.com/science/article/pii/S2666285X22000206

変化は唯一の定数です。 多くの分野で、ますます急速になっている変化が目撃されています。 これにより、多くの新しいイノベーションの可能性がもたらされます。 これには、傾向、将来の発展、およびそれらの結果に関する十分な根拠のあるデータが必要です。 この研究は、LDAとBoWのようなトピックモデリング技術を通じて発見される各トピックの関連性を備えた予測因子としての新しい方向性、パラダイムを捉えることを目指しています。 このために、Journal of Applied Intelligenceの3269の研究記事について実証分析が行われ、30年間にわたって収集されました。 次に、推測されたトピックは、予測分析を実行するのに適した方法に構造化されました。 これは、将来どのテクノロジーに遭遇するか、そして物事を革新して発見する人間の能力がこの世界にどれだけつながるかを予測するのに役立つという意味で重要です。 TF-IDFスコアを使用した最終モデルは、ベースラインモデルを41%上回っています

Dynamics of senses of new physics discourse: Co-keywords analysis

新しい物理学の談話の感覚のダイナミクス: 共同キーワード分析 https://www.sciencedirect.com/science/article/pii/S1751157721001164

この論文は、新しい物理キーワードの共起パターンの進化の縦断的分析を示しています。 そのために、1989年から 2018年までINSPIREデータベースで索引付けされたドキュメントを調査します。私たちの目的は、急成長している新しい物理学のサブフィールドの知識構造を定量化することです。 キーワード共起分析への新しいアプローチの開発は、こ

の論文の要点です。 従来の共同キーワードネットワーク分析とは対照的に、物理学の概念を異なるドキュメントに統合し、異なるドキュメントを同じ物理学の概念にバインドする構造を調査します。 概念間の関係を明らかにする構造を位相幾何学的と見なし、それらを「物理感覚」と呼びます。 軌道相互情報量の概念に基づいて、この論文は物理感覚のクラスタリングを提供し、それらの寿命を決定し、感覚の「権威」の分類を構築します。

Neural Networks Representation For Semantic Networks

セマンティックネットワークのニューラルネットワーク表現

https://ieeexplore.ieee.org/abstract/document/9311964

セマンティックネットワークは、非常によく知られている知識表現アプローチの1つです。 さまざまな場合に知識の不正確さや不完全さがあり、そのような問題は機械学習によって克服することができます。 この論文では、機械学習プロセスの準備ができるように、セマンティックネットワークをニューラルネットワークに変換するアプローチを紹介します。 変換プロセスでは、セマンティックネットワークで考えられるすべての概念と関係が、すでに存在するかどうかに関係なく考慮されます。 提案されたニューラルネットワーク構造には、概念層、関係層、概念および関係層など、さまざまないくつかの層があります。 学習プロセスでは、いくつかのニューラルネットワークに対して独立したトレーニングを行うことが提案されています。

Towards Study of Research Topics Evolution in Artificial Intelligence based on Topic Embedding

トピック埋め込みに基づく人工知能の研究トピック進化の研究に向けて https://ieeexplore.ieee.org/abstract/document/9721503

人工知能(AI)は、コンピューターサイエンスの最も注目されている研究トピックの1つです。 それは急速に成長している領域であるため、さまざまな期間の変化を分析することは、研究者が将来起こりうる変化に応じてキャリアパスを開発するためにその進化を分析するのに役立ちます。 この論文では、AIサブフィールドが互いにどのように近づいたり遠ざかったりしているのかを調査します。 この研究はまた、研究トピックの関係と相互作用を経時的に研究するための一般的な方法論を提供します。 この点で、AIのさまざまなサブフィールドが時間の経過とともにどのように進化するかを調べるために、トピック埋め込みアプローチを採用しています。 トピックの埋め込みは、トピックを意味のある数学モデルに変換して、潜在的な関係を維持する方法です。 この方法は、計算効率が高いだけでなく、サブドメイン間の関係に関する重要なパターンを明らかにする能力を証明しています。

Identifying Academic Creative Concept Topics Based on Future Work of Scientific Papers

科学論文の将来の研究に基づいた学術的創造的概念トピックの特定 https://manu44.magtech.com.cn/Jwk infotech wk3/EN/abstract/abstract5059.shtml

【目的】本論文は、科学論文から今後の研究文を分析し、学術革新のアイデアを自動生成することを目的としています。 【方法】まず、ルールマッチングとBERTを組み合わせて、論文から将来の作品の文章を抽出しました。 次に、関連分野の論文の拡大計算を行い、今後の方向性に関するキーワードや論文を特定しました。 最後に、これらの革新的な原材料は、革新的な概念のトピックを作成するためにUniLMベースのモデルに供給されました。 【結果】生成された結果の平均イノベーションスコアは6.04ポイント、平均関心レベルスコアは6.01ポイントです。 【制限事項】トピック生成モデルには、事前の意味知識が含まれておらず、実験に大規模なデータを使用していないため、生成されたトピックの品質を向上させる必要があります。 【結論】提案手法は、技術革新を拡大するための新しいアイデアを提供します。

Improving random walk rankings with feature selection and imputation Science4cast competition, team Hash Brown

特徴選択と代入Science4cast競争、チームハッシュブラウンによるランダムウォークランキングの改善https://ieeexplore.ieee.org/abstract/document/9671785

Science4castコンペティションは、セマンティックネットワーク内の新しいリンクを予測することで構成され、各ノードは概念を表し、各エッジは2つの概念に関連する論文によって提案されたリンクを表します。 このネットワークには、1994年から2017年までの情報が含まれており、日数が離散化されています(基礎となる論文の発行日を表します)。 チームハッシュブラウンの最終提出物であるee5aは、テストセットで0.92738のスコアを達成しました。 私たちのチームのスコアは2位で、勝者のスコアより0.01低くなっています。 このホワイトペーパーでは、モデル、その直感、およびテストセットでのバリエーションのパフォーマンスについて詳しく説明します。

Link Prediction of Artificial Intelligence Concepts using Low Computational Power

低計算力を使用した人工知能概念のリンク予測

https://ieeexplore.ieee.org/abstract/document/9671719

このペーパーでは、人工知能の先端研究所が主催する「Science4cast」2021コンテストに提案されたアプローチを紹介します。このコンテストの主な目的は、セマンティックネットワークにおける機械学習の概念間の将来の関連の可能性を予測することでした。 開発された方法論は、低次のトポロジー特徴の抽出とノード間の将来の接続の程度を推定するための最適化された分類器への組み込みを活用して、低計算能力のみの可用性のシナリオのソリューションに対応します。 開発された方法論を動機付けた理由、およびいくつかの結果、制限、および改善の提案について説明します。

Improving random walk rankings with feature selection and imputation

特徴選択と代入によるランダムウォークランキングの改善

https://arxiv.org/abs/2111.15635

Science4castコンペティションは、セマンティックネットワーク内の新しいリンクを予測することで構成され、各ノードは概念を表し、各エッジは2つの概念に関連する論文によって提案されたリンクを表します。 このネットワークには、1994年から2017年までの情報が含まれており、日数が離散化されています(基礎となる論文の発行日を表します)。 チームハッシュブラウンの最終提出物であるee5aは、テストセットで0.92738のスコアを達成しました。 私たちのチームのスコアは2位で、勝者のスコアより0.01低くなっています。 このホワイトペーパーでは、モデル、その直感、およびテストセットでのバリエーションのパフォーマンスについて詳しく説明します。

On scientific understanding with artificial intelligence

人工知能による科学的理解について

https://arxiv.org/abs/2204.01467

すべての素粒子物理学実験の結果、すべての化学反応の生成物、またはすべてのタンパク質の機能を正しく予測する神 託を想像してみてください。そのような神託は、私たちが知っているように、科学技術に革命をもたらすでしょう。しかし、科学者として、私たちはオラクル自体に満足することはありません。もっと欲しい。オラクルがこれらの予測を どのように考えたかを理解したいと思います。科学的理解として示されるこの偉業は、科学の本質的な目的として頻繁 に認識されてきました。現在、コンピューターと人工知能の成長し続ける力は、1つの究極の問題を提起します。高度 な人エシステムは、科学的理解にどのように貢献するか、それを自律的に達成することができるでしょうか。 これは単なる技術的な問題ではなく、科学の中核にあると私たちは確信しています。したがって、ここで私たちは私たちがどこにいるのか、そして私たちがここからどこに行くことができるのかを答えることに着手しました。私たちはまず、科学 的理解を理解するために科学哲学からアドバイスを求めます。次に、文学と、コンピューターの助けを借りて新しい概念の理解をどのように獲得したかについての科学者からの数十の逸話を収集することによって、現在の最先端技術をレビューします。これらの組み合わされた洞察は、Androidが支援する科学的理解の3つの次元を定義するのに役立ちます。AndroidはI)計算顕微鏡、II)インスピレーションのリソース、そして究極の、まだ存在しないIII)理解のエージェントです。それぞれの側面について、現状を超えて、科学の中心的な目的に対する人工知能の貢献の全力を解き放つための新しい道を説明します。私たちの視点が、新しい科学的理解を獲得し、最終的に私たちを真の人工科学者に近づけるアンドロイドに向けた研究に刺激を与え、焦点を当てることを願っています。

The Dawn of Metamaterial Engineering Predicted via Hyperdimensional Keyword Pool and Memory Learning

超次元キーワードプールと記憶学習によって予測されるメタマテリアル工学の夜明け https://onlinelibrary.wiley.com/doi/full/10.1002/adom.202102444

メタマテリアルの研究は20年以上継続されており、多くの公的および科学的関心を集めています。特に2010年の「知識ツリー」の代わりに、メタマテリアルの将来を予測する多くの専門家がいます。10年後、自然言語処理などの自動化されたコンピューターツールを使用してこれらの主張を再検討することが提案されています(NLP)、出版物の非構造化テキストから処理および分析するための研究情報を抽出します。この研究では、Scopus Search API(アプリケーションプログラミングインターフェイス)を使用して2000年から2021年の間に公開されたメタマテリアルに関連する43678の抄録の完全に自動生成されたデータベースが構築されます。単語の埋め込みを適用して、各キーワードは超次元ベク

トル空間とクラスターで研究され、研究テーマの人気と傾向を評価するためにそれらの関係を視覚化できます。エンコーダーデコーダーの長短期記憶(LSTM)アーキテクチャに基づいて開発されたニューラルネットワークモデルは、選択されたトピックの今後4年間の将来の方向性とテーマの進化を予測するように最終的にトレーニングされます。この研究は、メタマテリアル研究の影響に関する重要な情報を提供するだけでなく、Gartnerの誇大宣伝サイクルの形で将来のメタマテリアル研究ロードマップを開発するための確固たる基盤を築きます。

Efficiently Predicting Scientific Trends Using Node Centrality Measures of a Science Semantic Network

科学セマンティックネットワークのノード中心性測定を使用して科学的傾向を効率的に予測する https://ieeexplore.ieee.org/abstract/document/9671671

科学のすべての分野が拡大しており、これらの変化はいくつかの予測可能なパターンに従う傾向があります。 新しい科学の大部分は、他の分野を統合した結果です。 グラフベースのセマンティックネットワークモデルを使用して、この論文は、リンク予測問題で組み立てられたマージブランチを予測する方法を示しています。 65000を超えるノードと数百万のエッジを持つグラフを分析する作業は困難です。 この課題を克服するために、この問題の最も重要なグラフ機能を特定し、それらを効率的に計算する方法を示します。 データについてLSTMをトレーニングし、データ機能とモデルハイパーパラメーターを最適化し、結果をベースラインモデルと比較します。

Information silos distort biomedical research

情報サイロは生物医学研究を歪める

https://www.biorxiv.org/content/10.1101/2021.07.26.453749v1.abstract

情報サイロは、科学者自身のコミュニティ内で生み出される知識への偏見を導入するための科学研究のしばしば悪意のある機能です。情報検索とテキストマイニングの最近の改善にもかかわらず、科学文献の広大さは一般にこの現象のせいにされてきました。しかし、科学の進歩に対するその実際の悪影響は、定量化されたことはありません。この分析は、特に病気、遺伝子、化合物の間の関係の発見において、生物医学的発見への影響を調査することによってそうすることを試みます。結果は、2つの科学的事実が新しい事実の発見を可能にする確率は、これら2つの事実が科学的状況の中でどれだけ離れて公開されたかに依存することを示しています。特に、確率は引用距離とともに指数関数的に減少します。したがって、科学の進歩の方向は、各科学的事実が公開されている場所に基づいて歪められ、元々近くにあった発見が将来の発見のシーケンスを推進する経路依存のバイアスを表しています。このバイアスに対抗するために、科学者は最新の計算アプローチで科学的研究の範囲を広げる必要があります。

The speed of information propagation in the scientific network distorts biomedical research

科学ネットワークにおける情報伝播の速度は、生物医学研究を歪めます https://peerj.com/articles/12764/

科学コミュニティ全体での科学的発見の伝播の遅れは、科学者の最も近いコミュニティ内で生み出される知識への偏見を導入するための科学研究のしばしば悪意のある特徴でした。情報検索とテキストマイニングの最近の改善にもかかわらず、科学文献の広大さは一般にこの現象のせいにされてきました。しかし、科学の進歩に対するその実際の悪影響は、定量化されたことはありません。この分析は、特に病気、遺伝子、化合物の間の関係の発見において、生物医学的発見への影響を調査することによってそうすることを試みます。結果は、2つの科学的事実が新しい事実の発見を可能にする確率は、これら2つの事実が元々科学的状況の中でどれだけ離れていたかに依存することを示しています。特に、確率は引用距離とともに指数関数的に減少します。したがって、科学の進歩の方向は、各科学的事実が公開されている場所に基づいて歪められ、元々近くにあった発見が将来の発見のシーケンスを推進する経路依存のバイアスを表しています。このバイアスに対抗するために、科学者は最新の情報検索および抽出アプローチで科学的研究の範囲を広げる必要があります。