



「次世代人工知能・ロボット中核技術開発」

(人工知能分野) 中間成果発表会

－人間と相互理解できる人工知能に向けて－

ネットワーク分析と言語処理の融合による 大規模文献データからの技術の 未来予測プラットフォームの研究開発

平成29年 3月29日

東京大学 工学系研究科

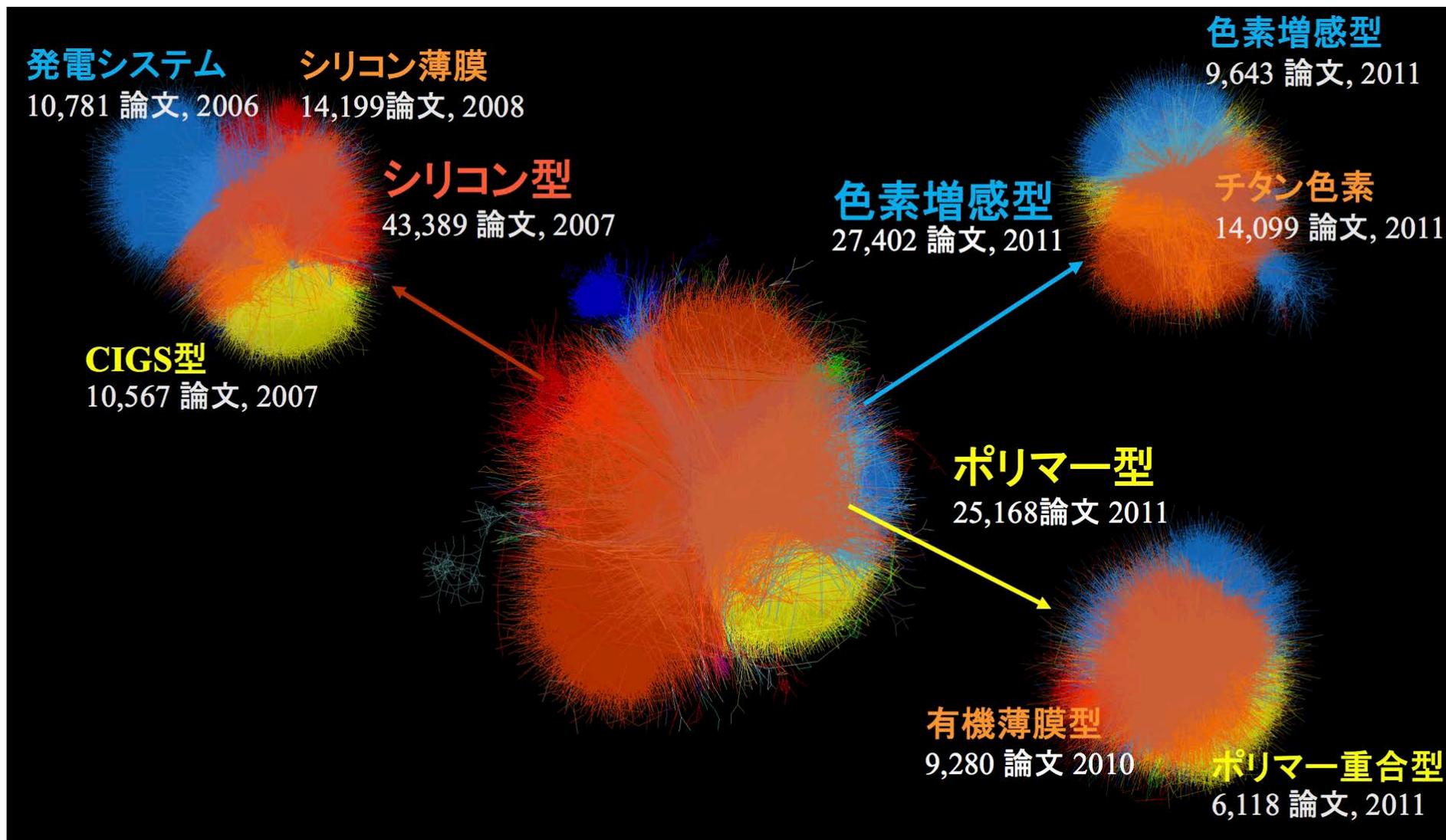
坂田一郎

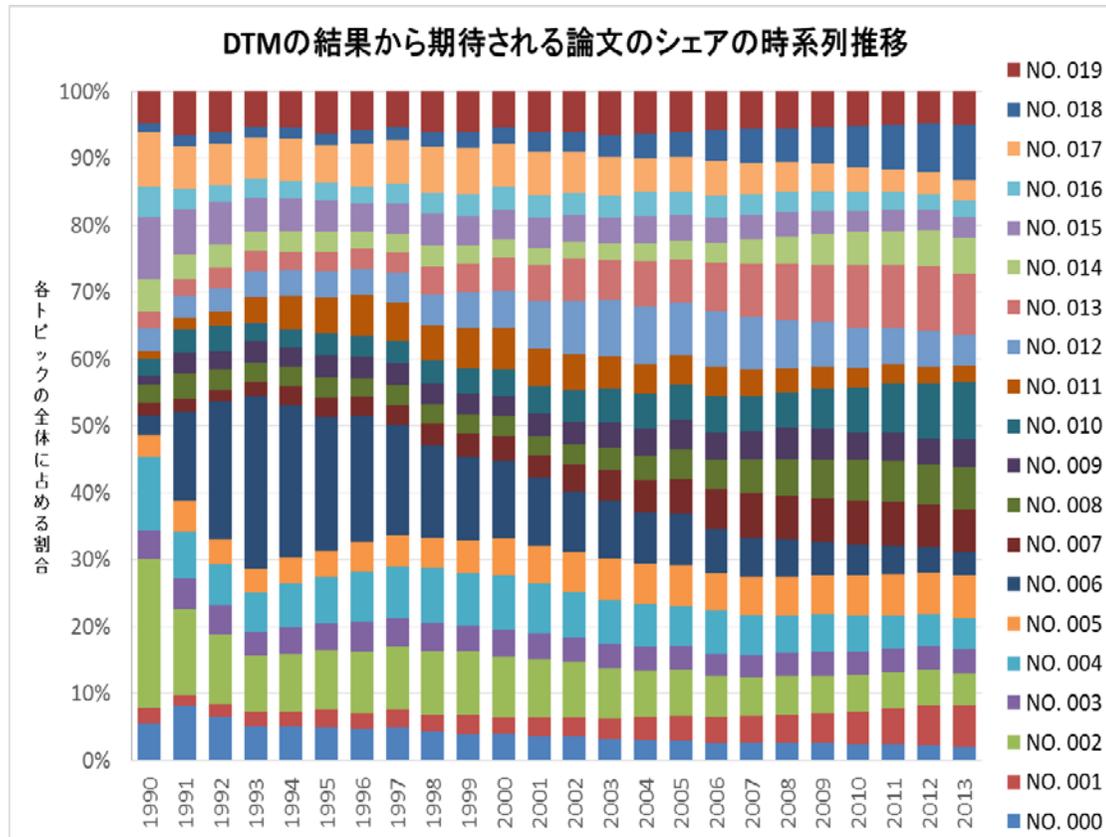
国立研究開発法人 産業技術総合研究所

国立研究開発法人 新エネルギー・産業技術総合開発機構

俯瞰分析：クラスタ情報（知の爆発）

ソーラーセル領域に属する10万件の論文情報に基づく学術俯瞰分析
(1900-2015.11, #1-6, TS=(Solar-cell or photovoltaic))



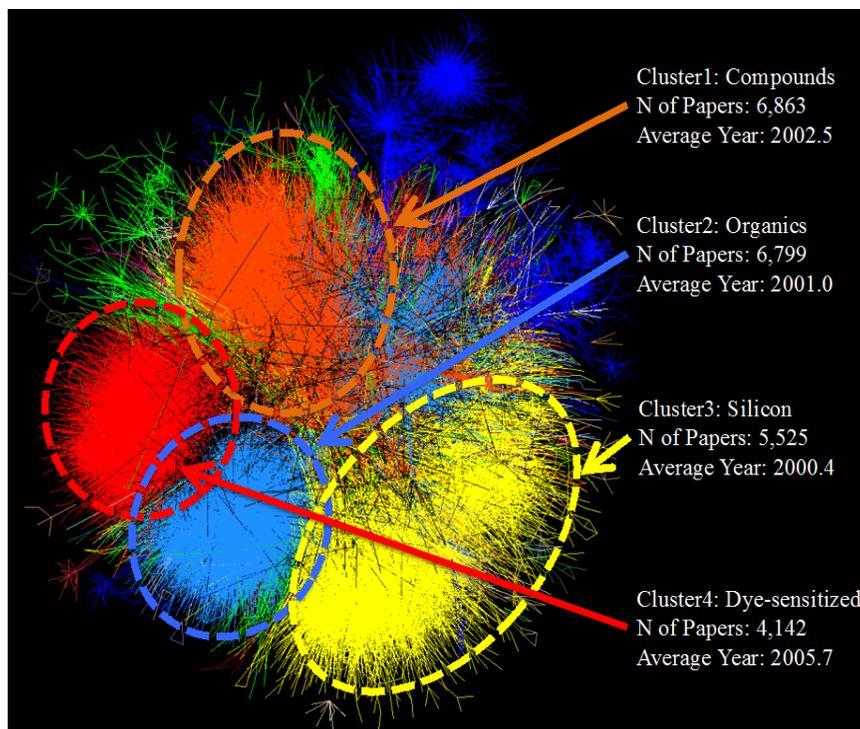


Topic No.	トピックラベル
No.0	バイオセンサ
No.1	リチウムイオン電池
No.2	環境影響評価(対生態系)
No.3	環境影響評価(成分分析)
No.4	Bio Medicine
No.5	吸着剤
No.6	フラーレン
No.7	ナノ複合材料
No.8	バイオセンサ
No.9	化学応用
No.10	ナノ粒子
No.11	フラーレンを用いた太陽光パネル
No.12	カーボン繊維の成長メカニズム
No.13	グラフェン
No.14	ドラッグデリバリー
No.15	ナノカーボンの作り方
No.16	基本物性
No.17	基本物性の評価
No.18	触媒と電池
No.19	CNTの合金

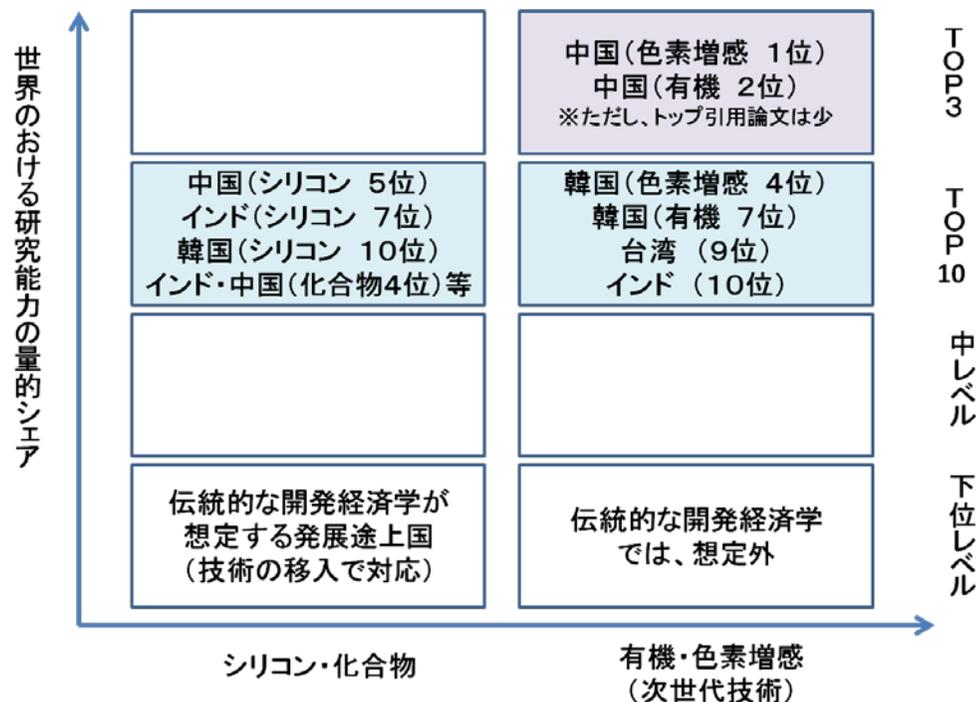
Web of Scienceを用いて収集したナノカーボン分野の約30万件の論文に対して、ダイナミックトピックモデルを用いて、上位20トピックを抽出した結果。

(Source) Y. Nakashio, et al. "Detecting structural changes in the nano carbon domain based on the time distribution of text information of academic papers" PICMET'2016

「科学面でのキャッチアップ (Scientific Catch-up)」の概念を提案し、太陽電池を事例に、その状況を実際に計測する手法を開発。中国、台湾において、シリコン電池よりも、次世代型電池において高いシェアを持つこと等、国毎の科学技術戦略の特徴や差異を明らかにした。これをもとに、APECの会議等でParallel-running-type growth modelを議論。



太陽電池のアカデミックランドスケープ



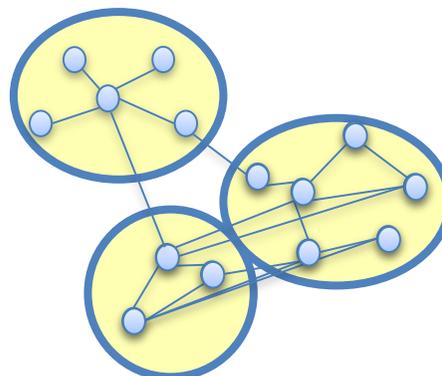
クラスタ別の国別論文数、引用数に基づくアジア諸国のマッピング



書誌情報
入手



引用ネットワーク
生成

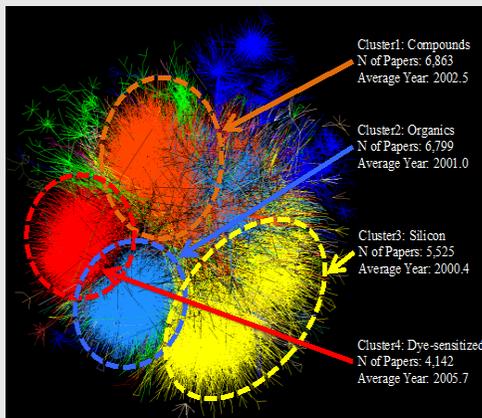


Thomson Reuters
Web of Science



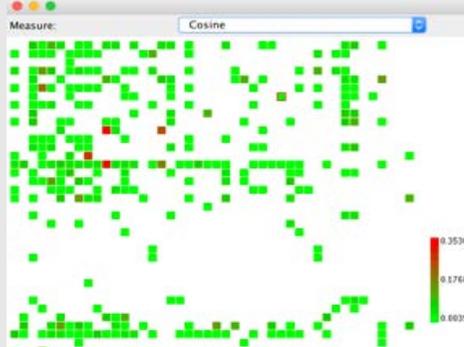
主な機能

④ 俯瞰分析



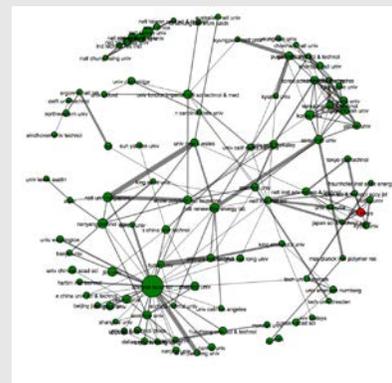
学術領域を俯瞰

⑤ ヒートマップ



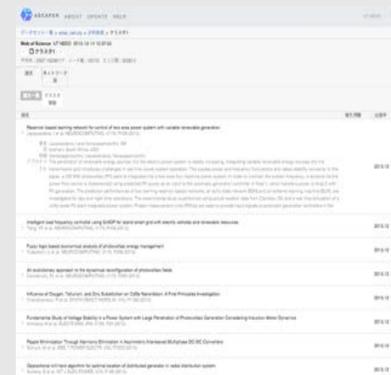
分野間の関係性を理解

⑥ 共著ネットワーク



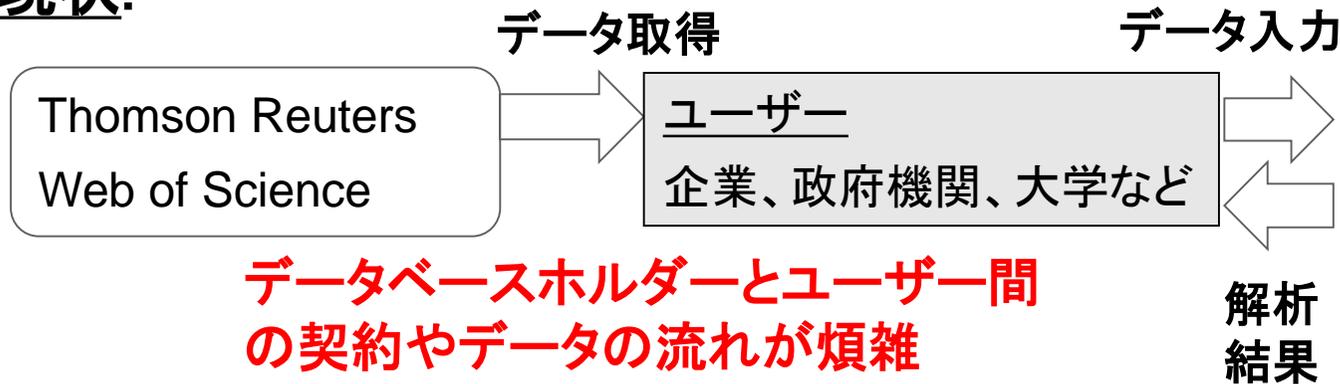
中心となる組織や人物を特定

⑦ 萌芽論文特定



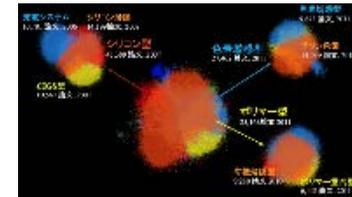
将来の高引用数文献を予測

現状:



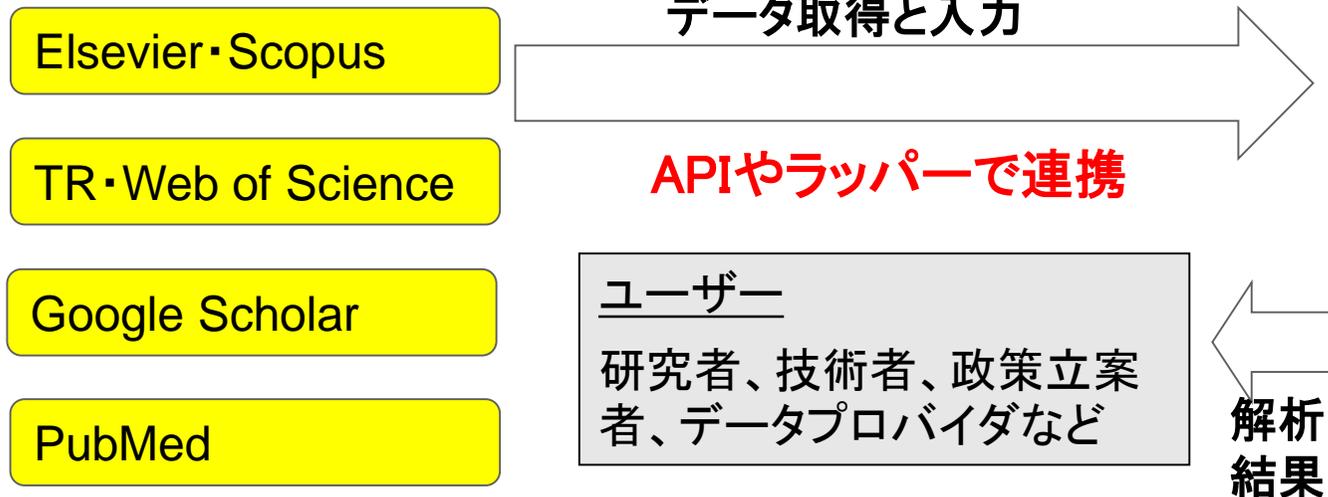
学術・産業技術俯瞰システム

学術領域の俯瞰、萌芽領域特定



今後: 任意のデータベースと統合したサービス化を検討
複数のステークホルダーへの試験的なサービス提供

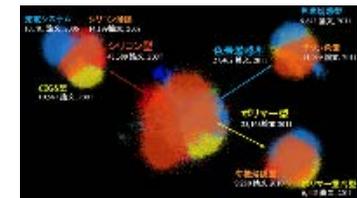
データ取得と入力



学術・産業技術俯瞰システム

学術文献データの解析

テキスト・ネットワーク分析



- 近年、文書・画像・ネットワーク上のノードなどの要素のベクトル表現を学習する表現学習手法の研究が進みつつある
- 表現学習を用いて獲得されたベクトル表現はコンピュータによる処理が容易なため、高精度な情報抽出やテキスト・ネットワークなどのマルチモーダルなデータからの情報抽出に利用可能

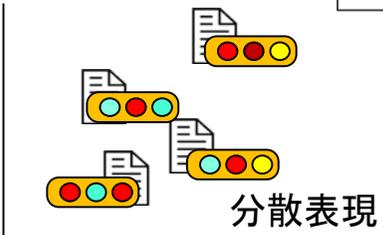
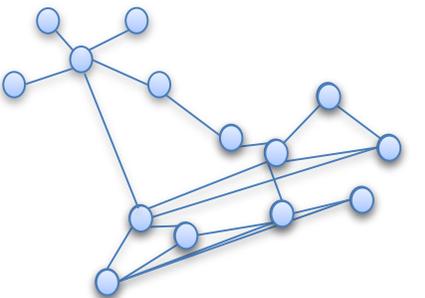
表現学習

引用ネットワーク

ネットワークの分散表現

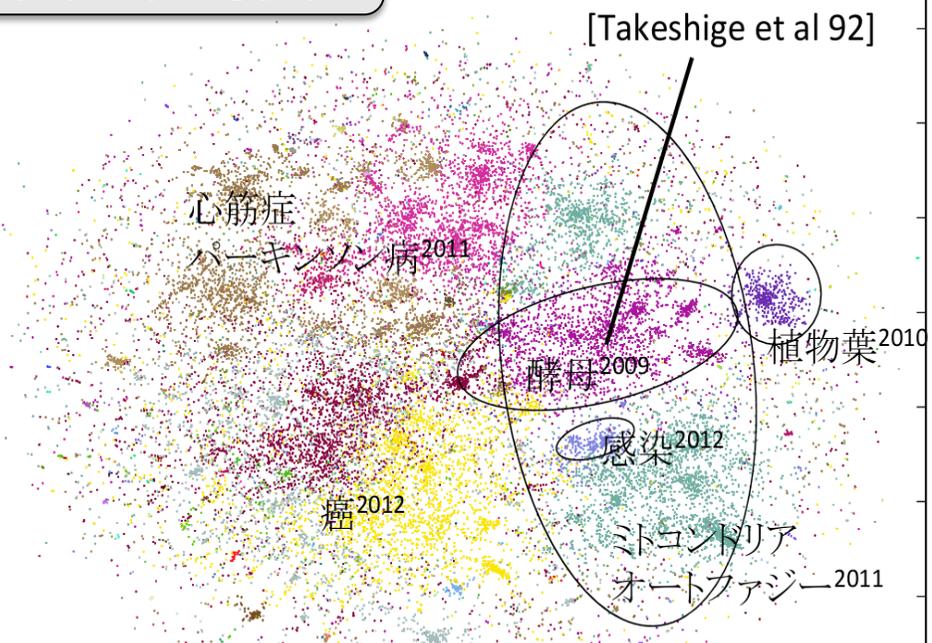
表現学習を用いた
学術領域の可視化

可視化



論文間のつながりの情報

各論文を高次元空間内のベクトル(分散表現)として表現



例: オートファジー分野

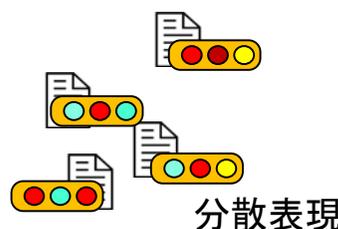
- 引用ネットワーク(既に研究蓄積あり)と本文・アブストラクトの分散表現の双方を用いて、分野の俯瞰や将来予測に有用な知見を抽出する

引用ネットワーク



ネットワークの分散表現

ネットワークの表現学習



分野のトレンドの抽出、論文被引用数の予測に有用な特徴量の開発

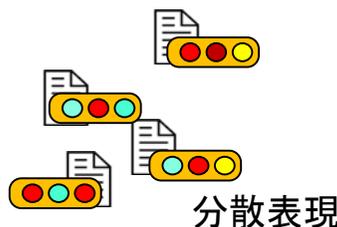
分野の成長の可視化

本文・アブスト



テキストの分散表現

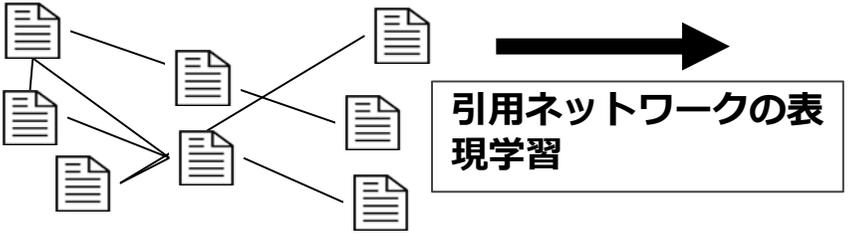
テキストの表現学習



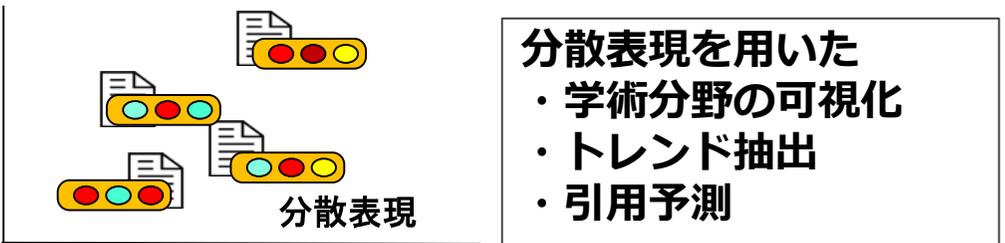
テキストとネットワークを融合した知識抽出
(実施予定)

- ・学術領域の発展するなかで分野の成長方向を定量化することは難しかった
- ・ネットワーク表現学習で得られた潜在空間内で成長方向を特定し、トレンドを抽出。全体トレンドの成長方向の中における各論文の位置づけを特徴量化し、萌芽予測に使用

引用ネットワーク

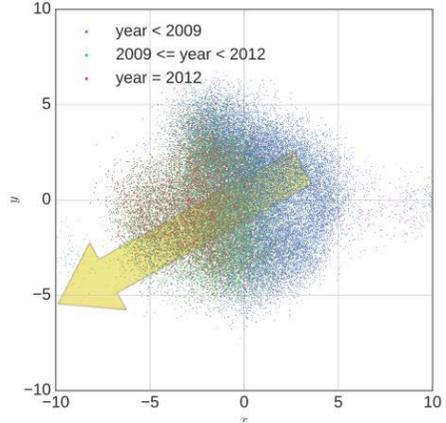


引用ネットワークにおける学術文献の分散表現



ネットワーク成長の方向の予測

ネットワークの分散表現を用いて、学術分野の可視化、トレンド抽出、引用予測などに応用
 学術領域の発展方向の検出
 成長方向上の位置を特徴量とした引用数予測結果



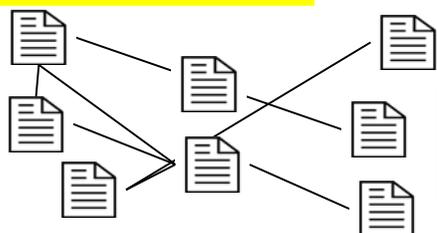
学術分野が空間内で特定方向に成長

	Web of Science			KDDCUP 2016			
	Nanocarbon	solar cell	complex network	Dopamine	Methane	Arsenic	
Total #Citations	IPY - 1st	0.3963	0.4123	0.1809	0.3906	0.3733	0.3034
	IPY - 2nd	0.36	0.2961	0.1464	0.1432	0.34	0.3193
	DA - 1st	0.1952	0.256	0.0875	0.261	0.1355	0.016
	DA - 2nd	0.2752	0.3297	0.1455	0.2421	0.1698	0.0549
	DC - 1st	0.0093	-0.007	-0.1014	-0.0448	-0.0547	-0.062
	DC - 2nd	0.1726	0.3226	0.057	0.2129	0.1535	0.0923
	Clustering Coefficient	-0.0194	-0.0023	0.0213	-0.0131	0.0159	-0.083
	Degree Centrality	0.2583	0.3072	0.1475	0.3179	0.1621	0.0152
	Closeness Centrality	0.3373	0.4077	0.1494	0.2846	0.2345	0.1805
	Pagerank	0.2293	0.2434	0.1441	0.2591	0.1651	-0.0352
	APYR	0.3242	0.2714	0.202	0.0912	0.1147	0.2606

既存手法よりよい予測精度を達成

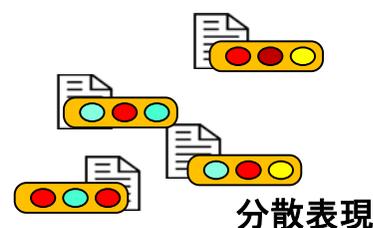
- ・学術領域の発展を可視化し、ノードを点として2次元に配置し、分野の成長、終焉、分岐、融合を観察
- ※既存手法ではクラスタや時間で集約した論文群の成長を可視化

引用ネットワーク



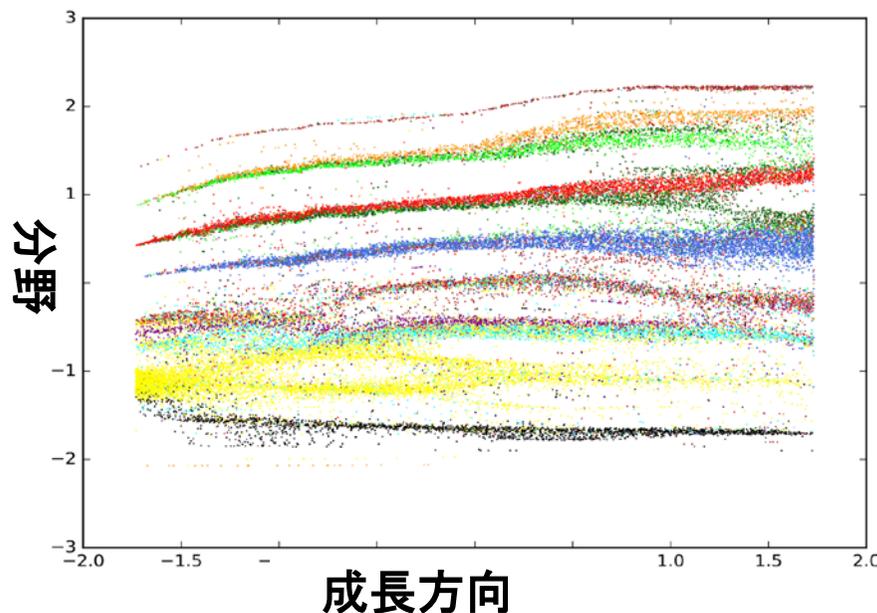
引用ネットワークの表現学習

引用ネットワークにおける学術文献の分散表現



- 分散表現を用いた
- ・学術分野の可視化
 - ・トレンド抽出
 - ・引用予測

ネットワーク成長の可視化



分野の成長、終焉、分岐、融合を可視化

結果の一例: グラフェン領域

黄色のナノチューブのクラスタは終焉の傾向があり、そこから、水色のフィルムや黒のレーザー関連が派生した。青のイオンやバッテリーのクラスタが近年発展の傾向。緑の光触媒関連のクラスタも融合しつつある。

研究計画

手法の高度化、他のデータベースでの検証、専門家へのヒアリング
ネットワーク情報とテキスト情報を組み合わせた情報抽出

研究成果の実現に向けた活動

① データベースホルダーとの連携

Elsevier社との連携によるシステムの評価予定
データベースホルダーとの連携によるサービスの提供が可能

② 企業、政府機関との連携による技法の検証

内閣府と連携した科学技術政策立案への本技法の活用の検討
化学系企業、メーカー研究所と共同研究を実施

③ アウトリーチ活動による成果の共有

「日仏シンポジウム 深層学習と人工知能」

主催：フランス大使館化学技術部等
本研究の取り組みに関して紹介

「新事業の創造に効く 学術 情報・ビッグデータの生かし方」

主催：日経テクノロジーオンライン
協賛：エルゼビア・ジャパン

俯瞰システムを用いた具体的な分析に関して共有
10社以上の会社から共同研究に関する問い合わせ

学術・産業技術俯瞰システム

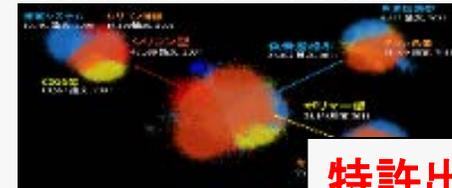
学術論文データ

Web of Science
Scopus(対応予定)

①

学術・産業技術俯瞰システム

学術領域の俯瞰、萌芽領域特定



特許出願済

②、③

想定ユーザー

企業、政府機関、大学等