



2019年6月21日

AIによる学術ビッグデータの処理とイノベーション経営への活用

坂田一郎

工学系研究科教授/

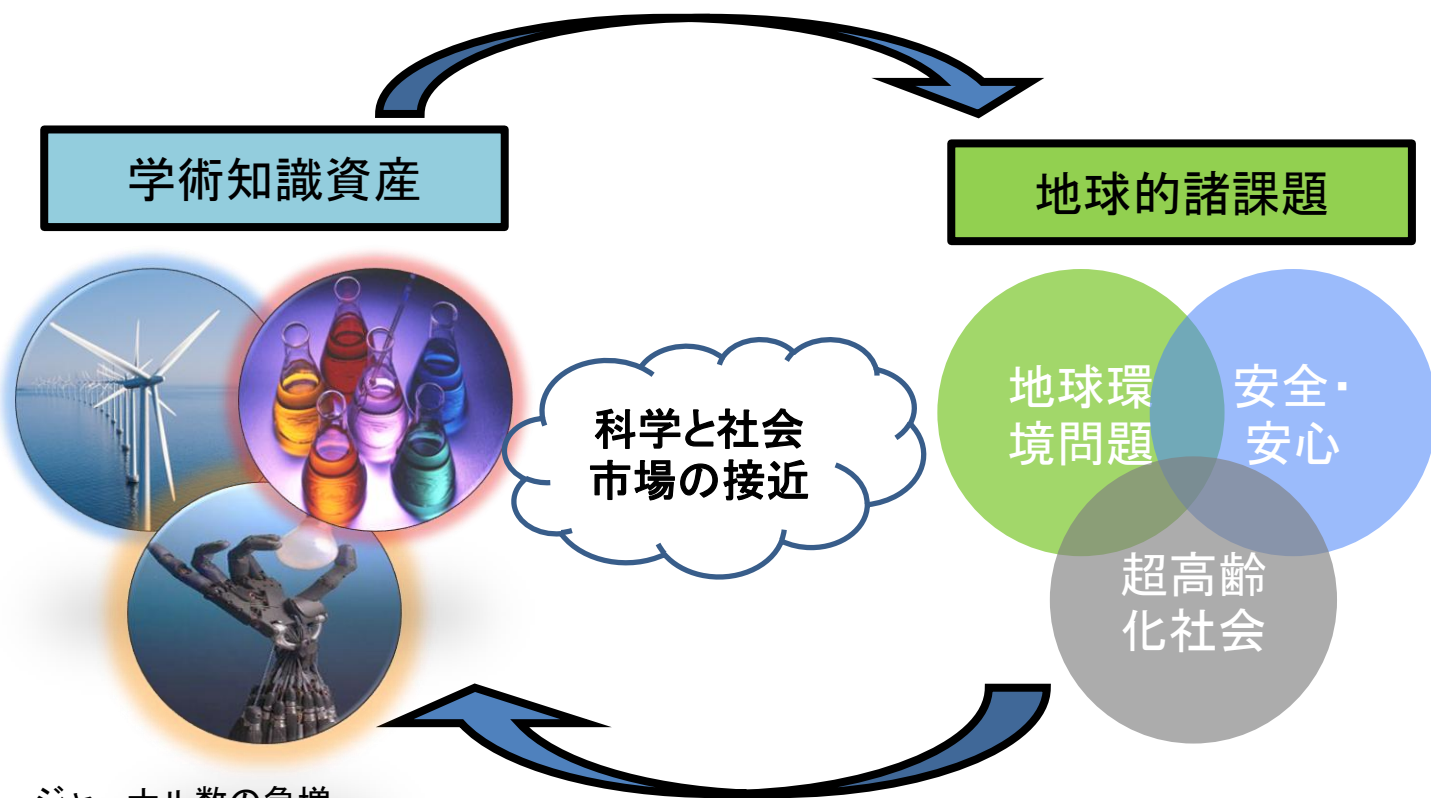
総合研究機構イノベーション政策研究センター長

The University of Tokyo

(備考)本発表は、NEDO/AIST「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」の成果

学術知識と課題解決との好循環

Analyze / Strengthen / Solve

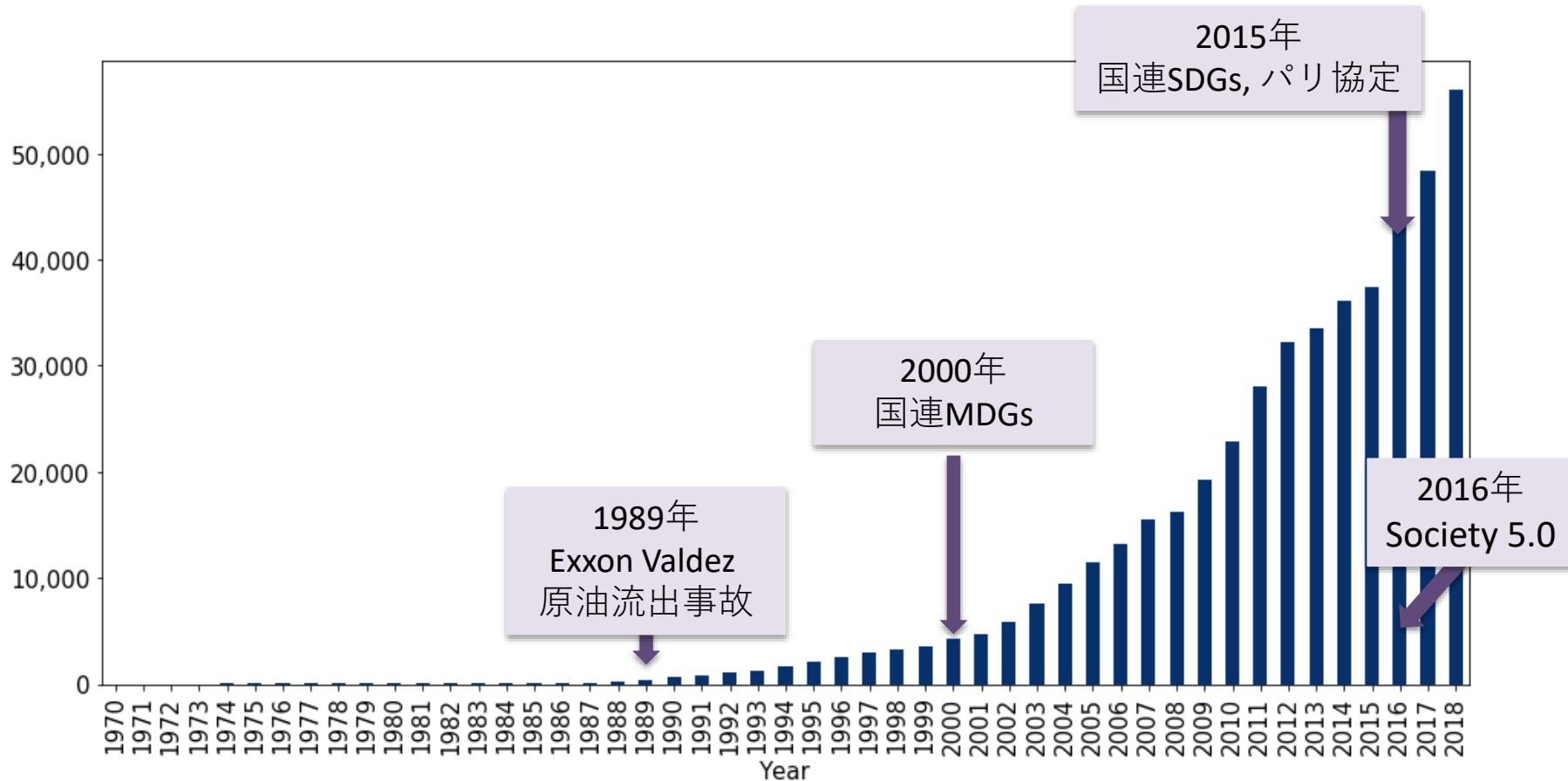


ジャーナル数の急増
メガ・オープンジャーナル
プレ・プリント/アーカイブ
知の生産の地政学的な変化

Stimulate / Demand / Feedback

学術知識の爆発と知の埋没

(サステナビリティ分野の例)



(備考) Scopusよりsustainability関連の論文(50万件弱)を取得し分析

知識の埋没と細分化

危険な投稿実験

(Douglas P. Peters & Stephen J. Ceci "Peer-review practices of psychological Journals: The fate of published articles, submitted again, The Behavioral and Brain Science (1982)5, 187-255による)

12本の異なる著名な心理学のジャーナルに一度、掲載された論文について、掲載の18-32カ月後に、再度投稿する(タイトルや組織名は変更)実験を実施。

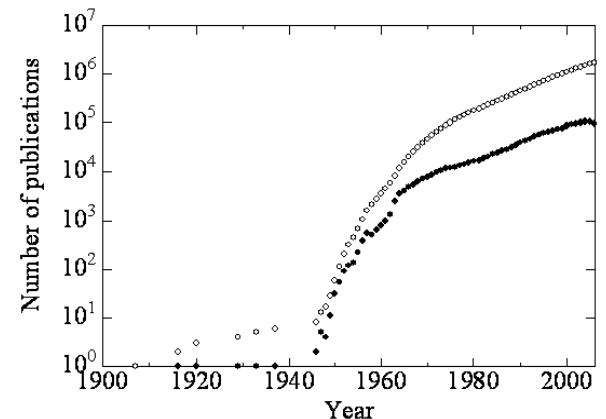
同内容の再投稿に気付いた編集者及び査読者 3/38

査読に回された論文 9/12

アクセプトされた論文 1/9

(rejectの理由 Methodological flaws)

最高レベルの専門家でも、実は、自分の周囲のごく限られた範囲の知識・情報しかみていない！



科学技術ビッグデータ活用 世界の動き

言語処理・機械学習アプローチ

[国外]

- IBMワトソン・診断支援
- DARPA ビッグメカニズム・癌治療
- 清華大学 Aminer・学術情報検索
- ワシントン大学 Semantic Scholar・学術情報抽出

[国内]

- 厚労省 医薬基盤研 (NIBIOHN)・新薬発見
- PRISM AI創薬
- 統数研・学術文献データ分析の新たな統計科学的アプローチ
- JST CREST・構造理解に基づく大規模文献情報からの知識発見

ネットワーク分析アプローチ

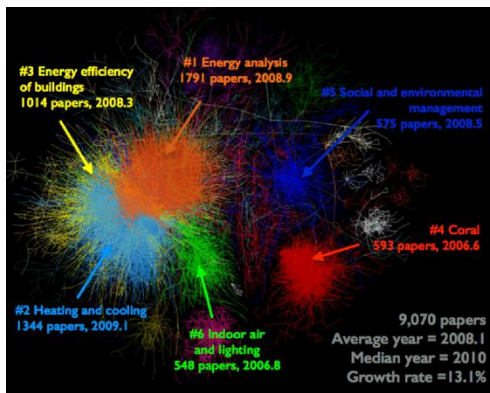
[国外]

- ライデン大学・研究機関評価
- 米国国立科学財団・政策のための科学

[国内]

- NISTEP・サイエンスマップ
- 東京大学イノベーション政策研究センター・「学術産業技術俯瞰システム」

「学術産業技術俯瞰システム」

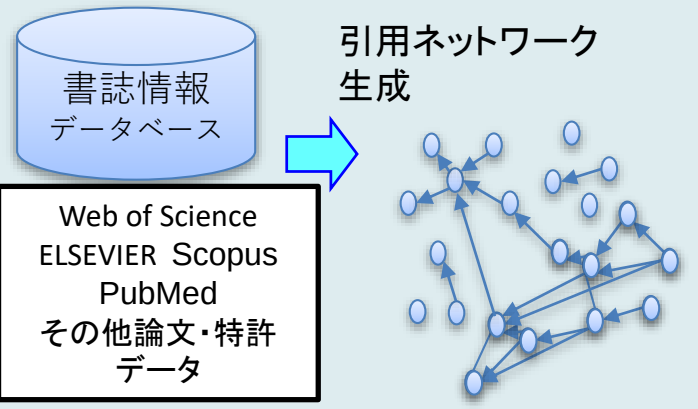


[国内外での関心]

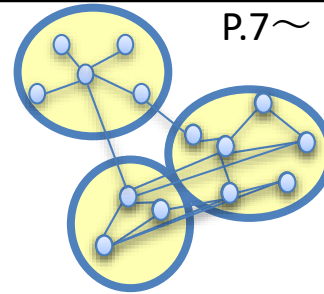
- Elsevier主催のResearch intelligence の会議で招待講演: (2017/06)
- M I T 原子力科学工学専攻のセミナーで招待講演()2018/09)
- S T M 国際出版社協会主催のGlobal voice of scholarly publishingで招待講演(2018/10)
- NHKスペシャルで特集「NHKスペシャル AIに聞いてみた どうすんのよ!? ニッポン第2回「働き方」(2018/03)で利用

開発中の技法群

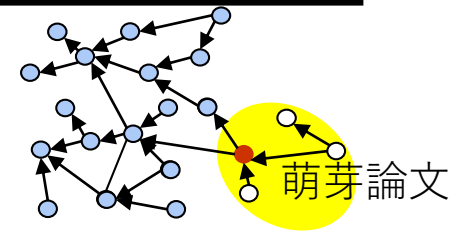
データの取得と前処理



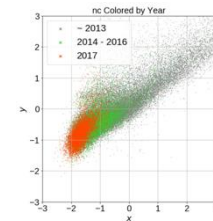
俯瞰分析(実装済み)



萌芽予測(実装済み) P.14~



トレンド検出(論文化済) 未来予測表現の抽出



P.18~

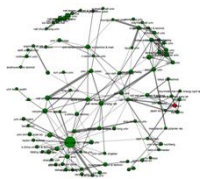
P.25~

キーワード抽出(開発中)

#Papers	Top 10 terms
0	uninspired dependency parsing technique tagging various random
6	cross-lingual conditional random random field conditional random field induction tagger
1	several statistical machine translation statistical machine learning alignment shared
2	shared task parallel phrase-based rule bilingual neural machine
3	Japanese tag type integrated adjacent integrated grammar
3	line-adjoining adjoining grammar tree-adjoining grammar deep quantifier dig
3	word disambiguation word sense lexicon word disambiguation word sense disambiguation
4	wordnet word sense lexicon word sense lexicon word sense lexicon
4	named entity named entity recognition named entity recognition biomedical
5	semantic relation extraction clinical knowledge line semantic conceptual
5	multitask multitask analysis neural network twitter detection topic
6	visual social polarity opinion analysis social analysis
2040	dialogue spoken dialogue system speech recognition spoken dialogue conversation
	multitask language generation act visual relevance expression referring

P.24~

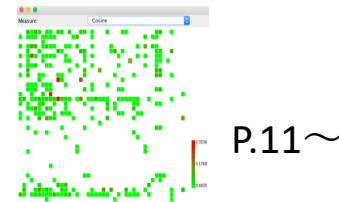
共著分析 (開発中)



研究機関のパ フォーマンス解析 (論文化済み)



ヒートマップ解析(実装済み)



分野間の影響度の推定 (開発中) P.30~

多角的な視点化から学術分野を俯瞰・分析・予測

先行して共同研究先との効果検証を実施中。⇒グローバル企業がシステムを使用。D社（空調メーカー）、F社（創薬化学化粧品）、J社（化学）、P社（創薬・化粧品）、T社（光学系）が俯瞰・萌芽技術をサイエンスロードマッピングに活用。システムの有用性を確認するとともに、萌芽予測機能などの改良を実施中。

多様なデータセットへの対応

- 多種の大規模論文データからの情報抽出

- Web of Science以外に対応した情報抽出のモジュール化
 - ✓ 引用データの抽出を実施済
 - DBLP, ACL, ACM,
 - ✓ 引用データの抽出を実施中
 - Scopus, PubMed
- 抽出された情報の学術俯瞰システムへの実装にむけたシステム開発
 - ✓ 学術俯瞰システムの汎用化に向けた設計を実施中

- 特許データセットも分析可能

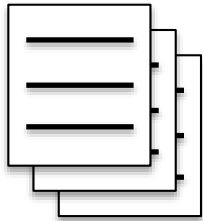
- Derwent Innovationデータによる特許データ分析・俯瞰

- 既存のテキスト分析ツールとの連携に関する設計

- マンチェスター大学のNaCTeMのテキスト分析ツールARGOと連携
- テキストとネットワークデータの統合分析基盤を設計

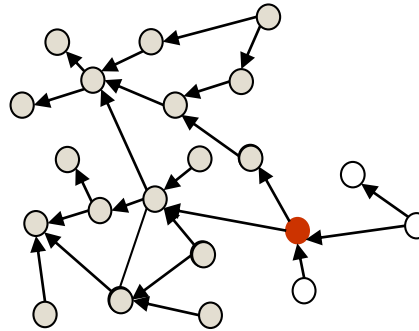
学術知識を構造化する手法

①書誌情報の取得



クエリに応じた情報を取得

②引用情報の抽出



学術書誌情報から引用を抽出

③分析の実行

ネットワーク解析

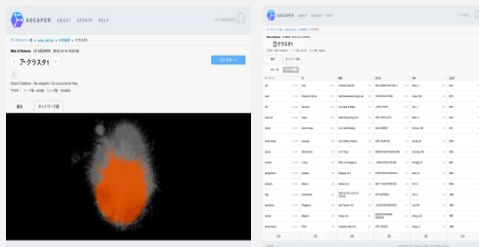
機械学習

自然言語処理

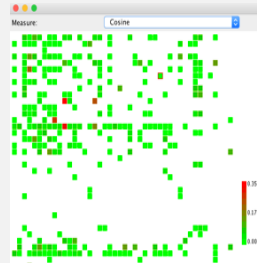
分析方法の選択
パラメータの入力

分析結果の表示

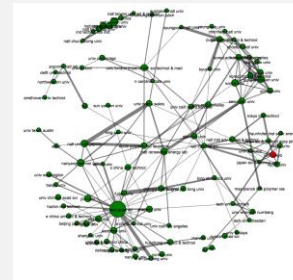
④俯瞰分析



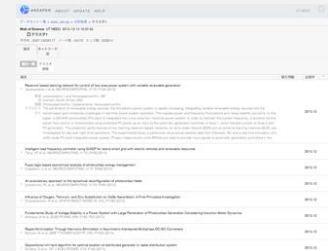
⑤ヒートマップ



⑥共著ネットワーク



⑦萌芽論文特定



サステナビリティに関する知の世界

20: Fiscal
sustainability

3: Bio/renewable energy

0: Soil, agriculture

17: Electrochemical

15: Water resource

11: Energy efficiency/emission

12: Water

1: Ecosystem

8: Fishery

13: Education

4: Urban/City

7: Energy/Emergy

14: Sustainable development

18: Mining

9: Tourism

21: Microfinance

2: Corporate

19: Waste management

16: Green computing

5: Building

10: Forest

6: Health care

Note: 0.3 Million Papers from Scopus (2019).
Source: 東京大学 坂田森研究室

Sustainabilityの分野の俯瞰(詳細)

Cluster	Year	#Nodes	Name	Keywords
0	2012.0	31,426	Soil, agriculture	soil, crop, agriculture, production, sustainable, system, management, agricultural, organic, plant, effect, land, nitrogen, food,
1	2012.0	29,118	Ecosystem	water, management, sustainable, ecosystem, change, sustainability, development, land, climate, river, forest, ecological, case, co
2	2012.4	28,902	Corporate	sustainability, sustainable, supply, corporate, chain, supply chain, management, environmental, social, performance, manufacturing
3	2011.1	27,032	Bio/renewable energy	production, sustainable, biomass, energy, synthesis, acid, based, waste, biodiesel, oil, using, catalyst, hydrogen, process, carbo
4	2012.4	23,570	Urban/City	sustainable, urban, sustainability, city, development, energy, transport, sustainable development, policy, case, food, planning, e
5	2013.6	18,010	Building	building, concrete, sustainable, construction, energy, sustainability, assessment, green, material, design, urban, performance, wa
6	2013.0	11,902	Health care	health, care, community, intervention, sustainability, program, based, implementation, study, sustainable, development, hiv, resea
7	2011.7	11,861	Energy/Energy	urban, china, ecological, analysis, water, footprint, sustainable, sustainability, energy, energy, city, development, case, based,
8	2014.2	9,173	Fishery	fishery, fish, aquaculture, marine, management, sustainable, sea, fishing, population, shrimp, sustainability, fishery management,
9	2013.6	8,827	Tourism	tourism, sustainable, sustainable tourism, development, ecotourism, case, sustainability, tourism development, destination, touris
10	2014.0	8,630	Forest	forest, management, forest management, sustainable, conservation, tree, de, specie, sustainable forest, forestry, timber, plant, h
11	2015.0	8,566	Energy efficiency/emission	energy, renewable, system, sustainable, power, renewable energy, development, solar, wind, electricity, grid, analysis, china, sus
12	2014.3	7,665	Water	water, wastewater, treatment, membrane, microbial, sustainable, microbial fuel, wastewater treatment, system, fuel cell, desalinat
13	2012.5	7,636	Education	education, sustainability, sustainable, development, sustainable development, learning, engineering, university, environmental, ed
14	2014.2	5,815	Sustainable development	sustainability, sustainable, assessment, development, environmental, sustainable development, indicator, economic, impact, policy,
15	2013.5	5,533	Water resource	groundwater, water, de, aquifer, irrigation, soil, management, china, basin, sustainable, recharge, resource, river, using, use, a
16	2012.7	5,057	Green computing	sustainability, sustainable, green, energy, information, design, software, system, data, technology, development, computing, enviro
17	2011.9	4,685	Electrochemical	carbon, hydrogen, doped, water, battery, hydrogen evolution, water splitting, metal, oxygen, ion, energy, efficient, photocatalyti
18	2015.3	4,343	Mining	mining, sustainable, mine, development, assessment, sustainability, metal, environmental, resource, remediation, waste, sustainabl
19	2013.2	3,601	Waste management	waste, solid waste, solid, waste management, management, municipal, life cycle, life, assessment, municipal solid, sustainable, cy
20	2012.8	1,864	Fiscal sustainability	fiscal, sustainability, debt, current account, fiscal sustainability, sustainable, debt sustainability, public debt, development,
21	2012.7	764	Microfinance	microfinance, microfinance institution, institution, development, financial, sustainability, microfinance institutions, sustainabl

赤字: 近年、活発に議論されている分野

エマージングなサブクラスタ

各々のクラスタ内の論文を再度クラスタリングすることで、数百～数千のサブクラスタを抽出し、近年活発に論文が出版されているクラスタを抽出。

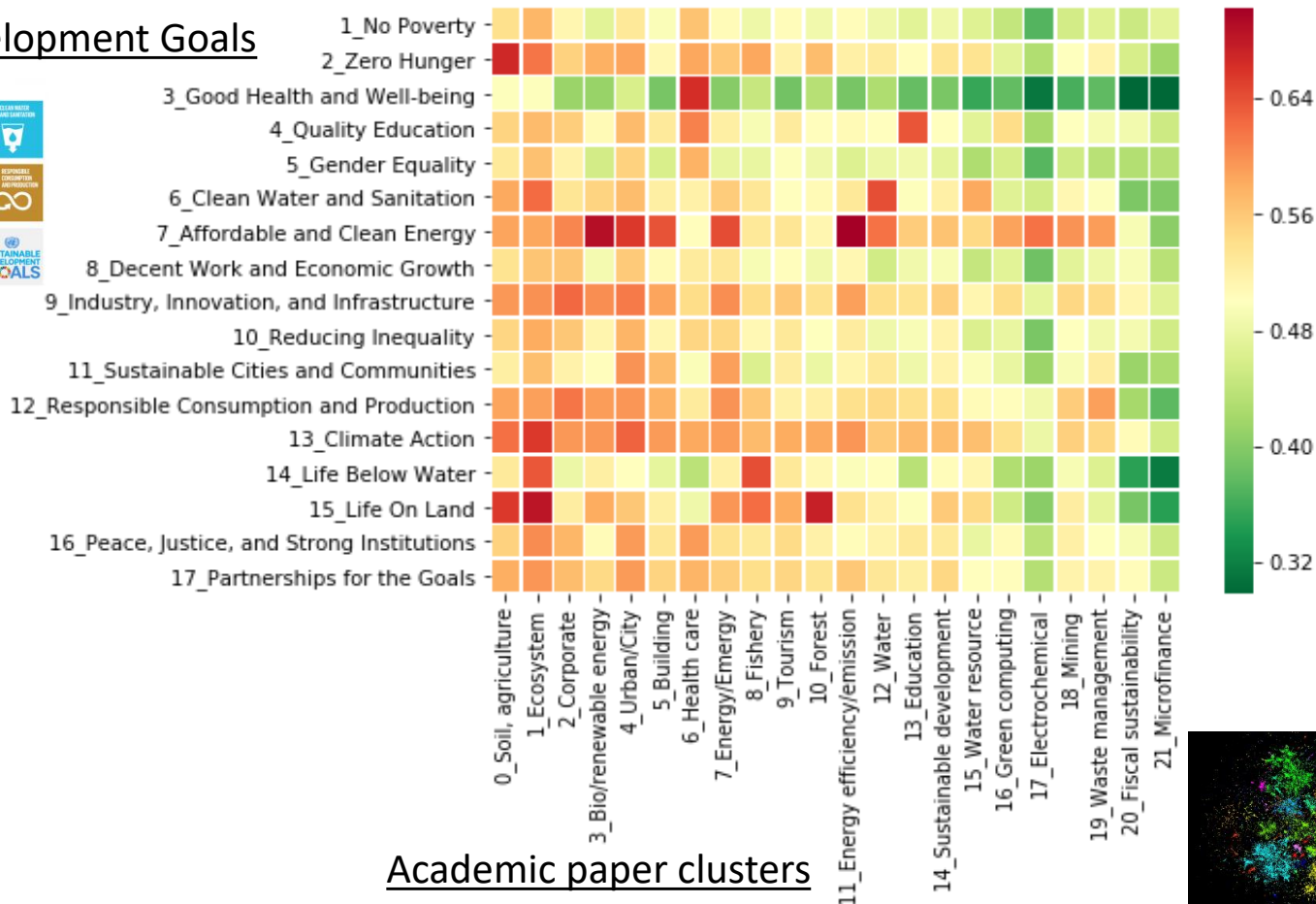
サステナビリティ 平均出版年トップ14サブクラスタ

Parent_cluster	Sub_cluster	Nodes	Year	Keywords	Description
Electrochemical	17_0	677	2016.688	hydrogen_evolution_reaction, oer, mv, oxygen_evolution_reaction, overpotential, electrocatalytic, electrocatalysts, mv_dec, overall_water_splitting, ni, water_splitting, low_overpotential, electrochemical, current_density, tafel, tafel_slope,	Electrocatalytic
Electrochemical	17_1	669	2016.288	porous_carbon, orr, porous, supercapacitors, hierarchical_porous, carbon_material, koh, mesoporous, high_specific, electrochemical, oxygen_reduction_reaction, hydrothermal_electrode, electrode_material, hydrothermal_carbonization, carbon,	Carbon material(super capacitors..)
Health care	6_1	942	2015.83	maternal_uhc, universal_health_coverage, livebirths, neonatal, u5mr, newborn, child, ncads, fap_sba, health, global_health, sustainable_development_goal, global, maternal_mortality_health_system, generic_medicine, middle_income_country,	Global health
Energy efficiency/emission	11_3	628	2015.346	economic_growth, ekc, economic, environmental_kuznets, financial_development, co2_emission, bin, energy_consumption, renewable_energy_consumption, gdp, renewable_fdi_trade_openness, emission, environmental_co2_arid,	Economic growth
Urban/City	4_6	1274	2015.307	smart_city, smart, smart_sustainable_city, urban, city, smart_sustainable, iot, sustainable, smartness, paper, big_data_analytics, citizen, smart_city_concept, ict, social_big_data, big_smart_city_project, sustainable_urban, smart_mobility,	Smart city
Bio/renewable energy	3_2	2593	2015.266	nanocellulose, pla, cellulose, lignin, ncncs, biodegradable, cellulose_nanocrystals, natural_fiber, composite, biocomposites, cnc, tensile, green_composite, mechanical, fiber_mechanical_property, polymer, natural_fibre, flax_cellulose_nanocrystal, cnf,	Nanocellulose
Bio/renewable energy	3_14	566	2015.233	des, deep_eutectic, deep_eutectic_solvent, nades, choline_chloride, extraction, natural_deep_eutectic, solvent, olive, antioxidant, bioactive, ultrasound, supercritical, ionic_liquid, natural_deep_eutectic_solvent, extraction_time, choline, green, phenolic,	Deep Eutectic solvent
Building	5_0	2882	2015.188	geopolymer, compressive, concrete, rca, recycled_aggregate, compressive_strength, uhpc, cement, mortar, fly_ash, geopolymer_concrete, geopolymers, recycled_aggregate_concrete, opc, cementitious_fine_aggregate,	Geopolymer
Bio/renewable energy	3_12	936	2015.143	catalytic, nanoparticles, magnetic, catalyst, catalytic_activity, kcc, synthesis, green, microwave, agnps, aqueous, reaction, pd, solvent_free, nm, gold_nanoparticles, magnetic_nanoparticles, sustainable, molecular_oxygen, organic, palladium,	Nanoparticles catalyst
Energy/Energy	7_7	652	2015.143	wf, water_footprint, virtual_water, blue, virtual, water_footprint_assessment, water, blue_water, virtual_water_content, water_resource, water_use, virtual_water_trade, total_wf, grey_water_footprint, china, total_water_footprint, green_agricultural,	Water footprint
Urban/City	4_9	774	2015.094	ev, electric, electric_vehicle, bevs, carsharing, car, collaborative_consumption, av, vehicle, car_sharing, hybrid_electric, bev, sustainable, ride, autonomous_vehicle, economy_paper, urban, peer_plugin, environmental_public, battery_electric_vehicle,	EV
Bio/renewable energy	3_11	1000	2015.07	epoxy, polymer, monomer, poly, cardanol, anticorrosive, polyester, polymerization, benzoxazine, copolymer, lubricant, epoxy_resin, nmr, polyurethane, bio, eugenol, thermoplastic, tribological, biodegradable, thermal, bisphenol, renewable, isosorbide,	Epoxy polymer
Energy/Energy	7_8	650	2014.971	dea, undesirable_output, china, dea_environmental_assessment, eco_efficiency, data_envelopment_analysis, eco, environmental, envelopment, non_radial, economic_undesirable, sustainable, managerial_disposability, dmus,	Environmental Assessment (DEA)
Bio/renewable energy	3_1	2846	2014.968	microalgae, microalgal, lipid, algal, algae, biodiesel, lipid_productivity, biodiesel_production, biofuels, lipid_production, scenedesmus, cyanobacteria, microalgal_biomass, microalgae_cultivation, mixotrophic, fatty_lipid_content,	Microalgae

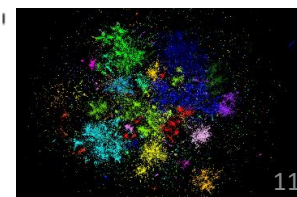
SDGs Goalsと学術研究の関連性

UNの17のSDGs Goalsの文章と各クラスタ内の論文のタイトル・アブストラクトとの言語的類似度を観察することで、SDGs Goalに関する議論が活発な学術領域を抽出。

UN Sustainable Development Goals



Academic paper clusters



俯瞰分析の活用例(太陽電池)

- ・ 詳細なクラスタリングによりエマージングな7分野を特定
- ・ その上で、主要な取り組みの立ち位置や競争環境を特定

第1 階層	第2 階層	第3 階層	論文数	平均出版年	国内論文数の世界ランク	論文数第一位(国)
シリコン型	シリコン薄膜	パッシベーション	5051	2008.2	4	ドイツ
		ナノワイヤ	3843	2012		アメリカ
		アモ不ファス	3580	2004.4	1	日本
		ZNO	1437	2010.4		韓国
	システム	GA AS	3992	2008.8	3	アメリカ
		Module	3932	2009.7		アメリカ
		Power inverter	3345	2009.6		アメリカ
	CIGS	CdTe太陽電池	4855	1999		アメリカ
		Cu In太陽電池	3791	2008	3	ドイツ
		CZTS太陽電池	2303	2012.3	4	中国
色素増感型	TiO2	NanoTube	5348	2011.7	4	中国
		Quantum dot	3475	2012.2	5	中国
		ZnO太陽電池	3210	2011.6	5	中国
	Dye	TiO2太陽電池	4747	2010.2	3	中国
		electrolyte	2423	2010	3	中国
		perovskite	2018	2012.5	5	中国
	Dye Sensitized		3262	2011.5	2	中国
ポリマー型	organic	organic	4087	2009.4	4	アメリカ
		polymer blend	3209	2012		アメリカ
		small molecule	1327	2011.9		中国
	film	Nanostructure	3076	2012.1	4	アメリカ
		hybrid-polymer	2568	2010.5		中国
		実装技術	2137	2012.3		中国
	polymer	fullerene	2381	2009.6	3	中国
		conjugated polymer	2074	2011.9	5	中国
		Copolymer / Organic	1291	2012.7	3	中国

ナノワイヤ太陽電池
文科省「革新的エネルギー
研究開発拠点形成事業」
“FUTURE-PV Innovation”

民間を中心に開発
(ソーラーフロンティア社等)

理論的には75%の変換効率
(東大先端研 岡田研究室)

ペロブスカイトを用いており、
容易に大型化可能
日本発 (桐蔭横浜大学 宮坂)
NIMS, NEDOなどが取り組み

ナノ結晶
2005年にNEDOが着目
(NEDO海外レポート NO.966)
阪大(Prof. Fukuzumi)

実用化に向けたポリマー型
太陽電池の安定性向上

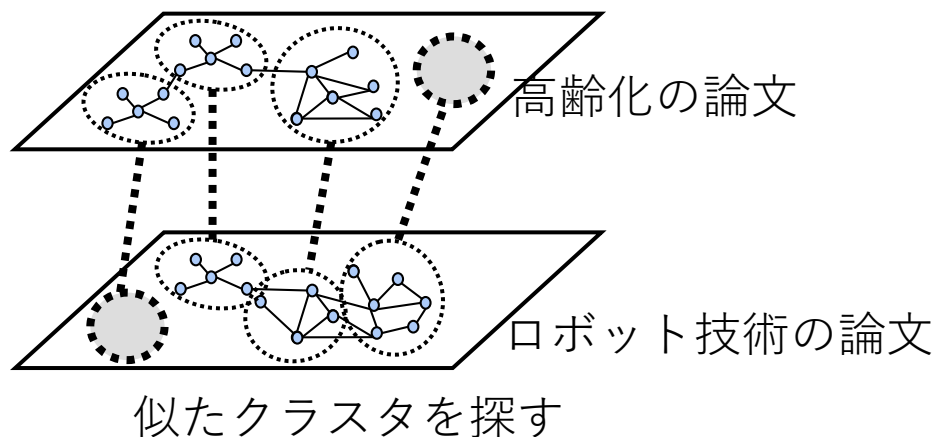
π 電子系骨格による高い変換効率
論文数4位 (理化学研究所 瀧宮)

※平均出版年2012年以降

ヒートマップ分析

異なるレイヤーの知識を接続する手法

検出の概念



検出の手法（テキスト類似度）

- ①各クラスタごとに特徴語をスコア化

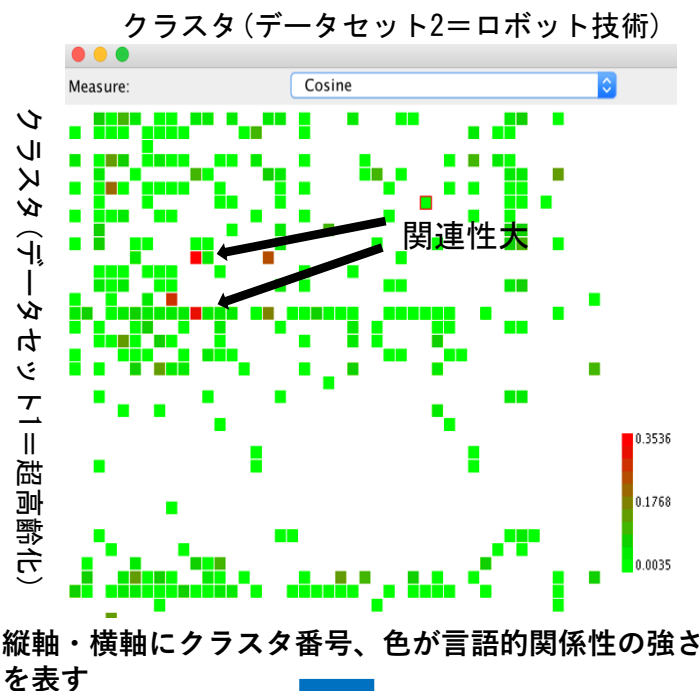
$$w(x, j) = tf(x, j) * \log\left(\frac{N}{df_x}\right)$$

Cluster x, term j

- ②2つのデータセットの各クラスタ間の類似度を計算

$$\text{similarity}(x, y) = \frac{\sum_{j=1}^n w(x, j)w(y, j)}{\sqrt{\sum_{j=1}^n w(x, j)^2} \sqrt{\sum_{j=1}^n w(y, j)^2}}$$

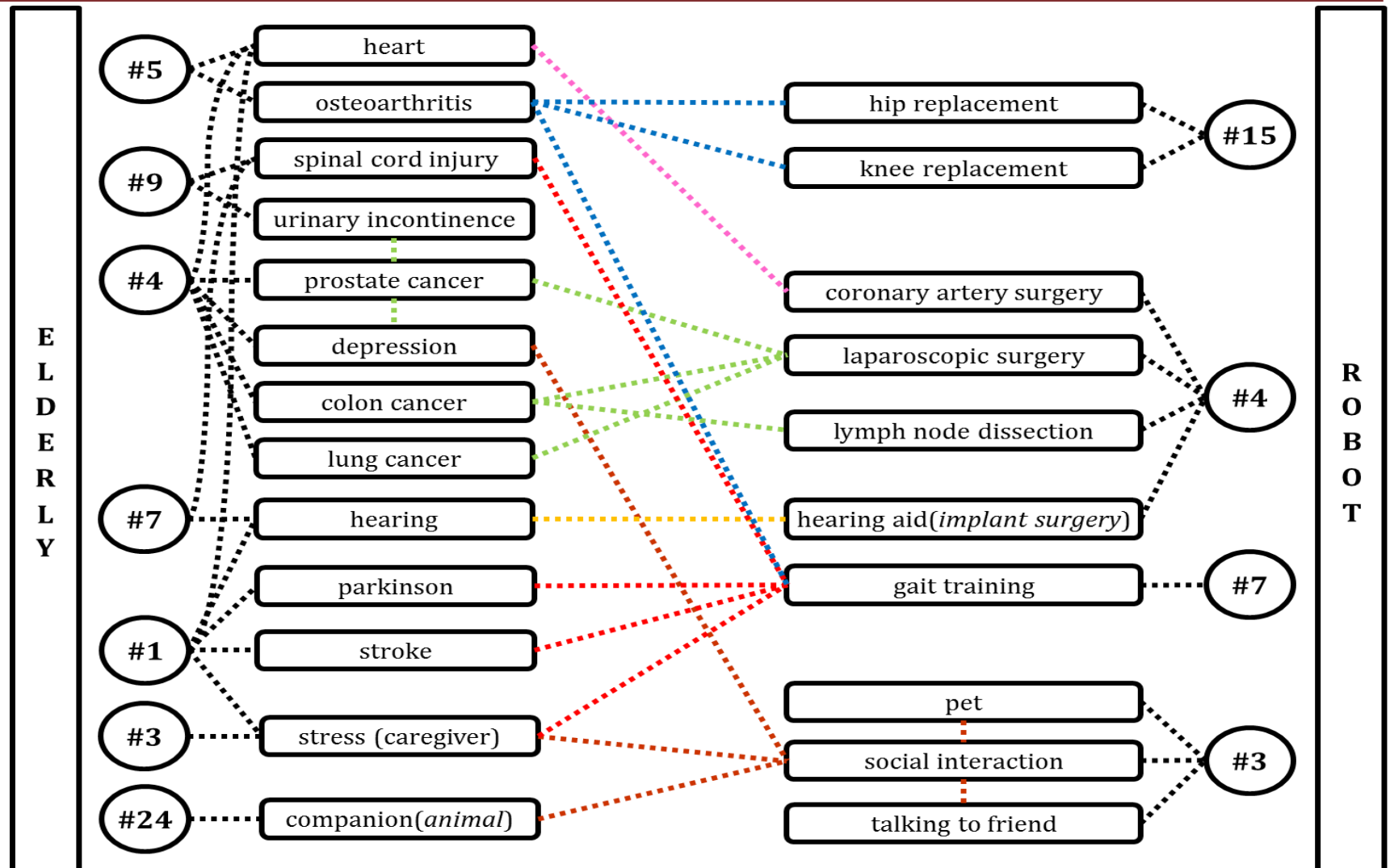
可視化、システムへの実装



関連度の高いクラスタのペアを抽出

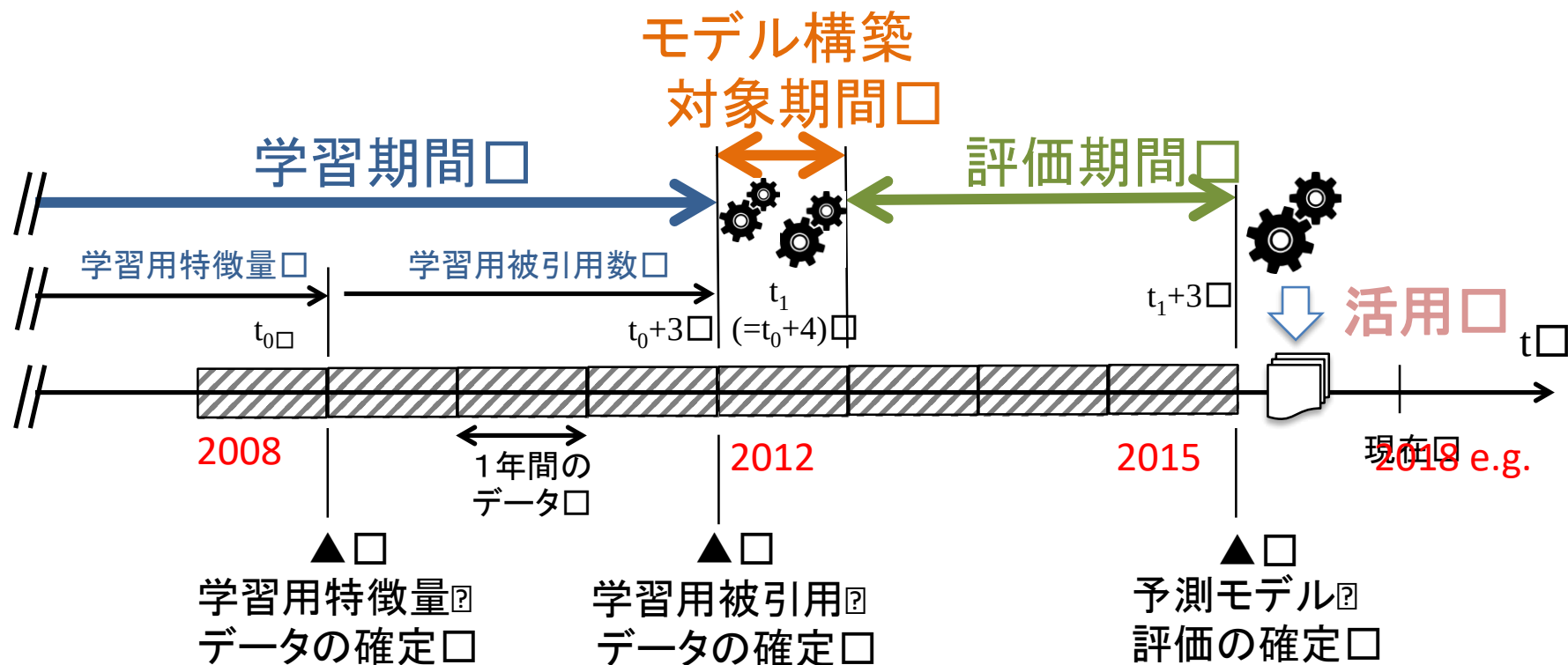
ヒートマップ分析結果

高齢社会の課題とロボット技術のつながり



(出典) Vitavin Ittipanuvate, Yuya kajikawa, Ichiro Sakata et al., Journal of Engineering and Technology Management vol.32 (2014), pp.160-184

萌芽分析：萌芽予測モデル



「モデル構築対象期間」を中心にした前後のデータ
「過去のデータによる学習」と「3年後※の実際の被引用数による検証」
によってモデルを構築。 ※:3年は可変パラメータ。目的や対象分野によって適切な数字を設計する。

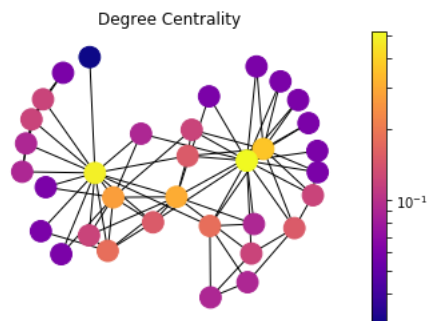
構築したモデルを現在手元にある最新の論文データに活用

学習用特徴量

ノード(論文)の中心性、ノードが属するクラスタの情報、近隣ノードの情報を表す指標群を網羅的に計算し特徴量として学習に用いる

1.中心性を表す指標群

ノードがネットワーク上の中心に位置しているかを示す



例: ページランク

$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$

次数中心性
媒介中心性
近接中心性
固有ベクトル中心性
クラスタリング係数
HITS
ページランク

2.ノードが所属するクラスタの指標群

ノードが属しているクラスタの大きさや内部密集度や独立性を示す

全ノード数
全エッジ数
モジュラリティ
(独立性を表す)
クラスタリング係数
(エッジの密集度)

3.ノードの周囲のノードの特徴量の平均/最大/最小



ノードの周辺の情報を集約し、ノードが属している近接分野の重要性を定式化

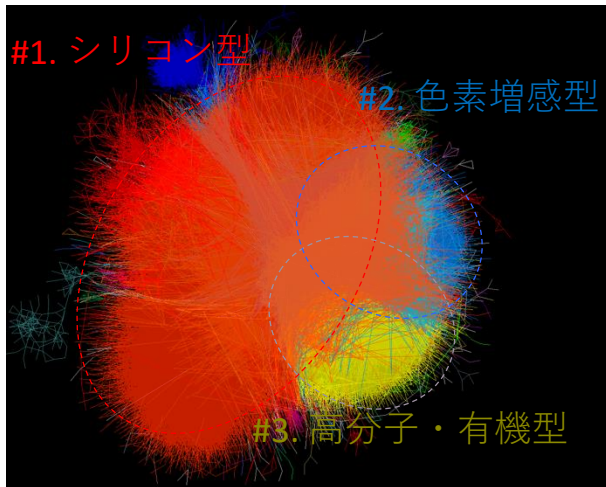


1,2,3合わせて63個の特徴量でノードの引用ネットワーク上での役割/位置を表現

萌芽分析結果

ソーラーセル領域(2012年発表論文)

高効率かつ安定したポリマー型太陽光発電における界面材料



色素増感系太陽光発電における、光物性、電気化学特性

最も研究が盛んなクラスター「#1シリコン型」から、「#2.色素増感型」や「#3高分子・有機型」へ注目が集まっていること示唆される。

クラスター	論文名	雑誌	現被引用数
3	Recent advances in solution-processed interfacial materials for efficient and stable polymer solar cells	ENERGY & ENVIRONMENTAL SCIENCE	320
3	Morphology characterization in organic and hybrid solar cells	ENERGY & ENVIRONMENTAL SCIENCE	180
3	Recent progress and future aspects of organic solar cells	PROGRESS IN PHOTOVOLTAICS	68
3	Solid-state dye-sensitized and bulk heterojunction solar cells using TiO ₂ and ZnO nanostructures: recent progress and new concepts at the borderline	POLYMER INTERNATIONAL	30
3	Rational Design of High Performance Conjugated Polymers for Organic Solar Cells	MACROMOLECULES	584
2	Photophysical and Electrochemical Properties, and Molecular Structures of Organic Dyes for Dye-Sensitized Solar Cells	CHEMPHYSICHEM	104
3	Small Molecule Organic Semiconductors on the Move: Promises for Future Solar Energy Technology	ANGEWANDTE CHEMIE-INTERNATIONAL EDITION	529
2	Characterisation of electron transport and charge recombination using temporally resolved and frequency-domain techniques for dye-sensitized solar cells	INTERNATIONAL REVIEWS IN PHYSICAL CHEMISTRY	35
3	Tandem polymer solar cells featuring a spectrally matched low-bandgap polymer	NATURE PHOTONICS	865
2	The Electrochemistry of Nanostructured Titanium Dioxide Electrodes	CHEMPHYSICHEM	

データセットにおける上位5%内論文(正例)

各対象分野における萌芽論文予測精度 (評価5セット [予測対象: '06~'10]の平均)

萌芽予測の予測結果

分野	適合率	再現率	F-scores	AUC
ナノカーボン	79.8%	88.8%	0.840 (0.021)	0.908 (0.018)
ソーラーフォト	84.5%	87.5%	0.859 (0.013)	0.919 (0.012)
ガリウムナイトライド*	75.3%	66.3%	0.706 (0.047)	0.784 (0.033)
複雑ネットワーク	73.6%	92.0%	0.818 (0.012)	0.888 (0.011)
ソーシャルシステム	76.4%	90.9%	0.828 (0.011)	0.892 (0.011)
リアルタイムロボット*	73.4%	78.9%	0.758 (0.034)	0.802 (0.041)
サイバーフィジカル	79.7%	75.6%	0.776 (0.026)	0.831 (0.018)

(括弧内は標準偏差)

● 各分野において平均で F-scores > 70, AUC > 0.7 を達成

(出典) 森純一郎, 原 忠義, 梶川裕矢, 坂田一郎 (2016). 大規模論文データからの異種ネットワーク組み合わせによる萌芽論文の推定, 人工知能学会全国大会.

萌芽論文の判定に寄与する特徴量

特徴量	寄与度
ネットワーク中心性(ページランク)	1.21
所属クラスタのノード(論文)数	0.99
雑誌名・著者名(「Science」など7件)	0.60~0.91
引用先論文の所属クラスタのノード(論文)数	0.58

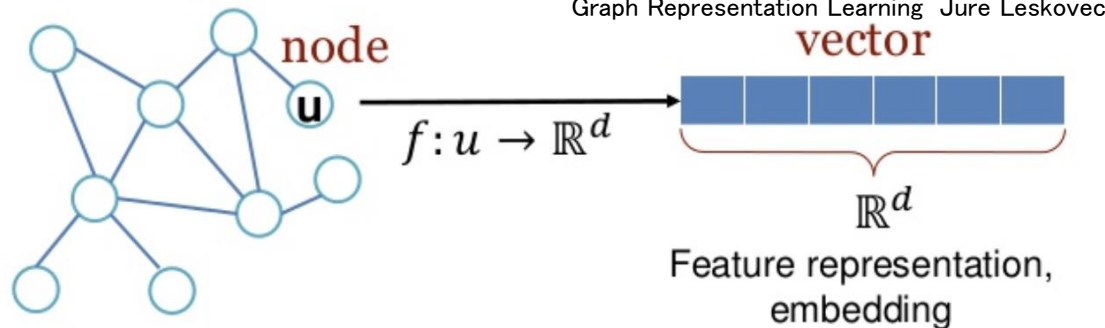
非萌芽論文の除外に寄与する特徴量

特徴量	寄与度
引用先論文の所属クラスタのモジュラリティ	-1.92
所属クラスタのモジュラリティ	-1.54
ネットワーク中心性(ネットワーク制約)	-0.71

ネットワークの表現学習の導入 (Network representation Learning)

ネットワーク構造よりノードの潜在的な表現を学習

Representation learning of large scale graphs in realistic time is proposed in DeepWalk (2014), and many subsequent methods have been proposed.



Network (sparse adjacency matrix)
コンピュータで処理するのが難しい

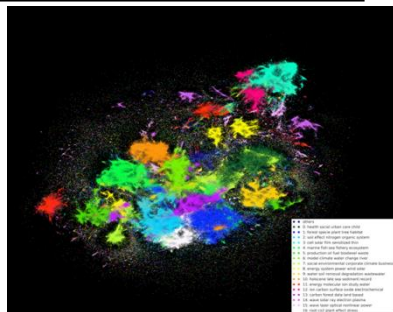
Vector representation

既存の機械学習のアルゴリズムにインプット可能



ネットワーク表現学習によるタスク

Large graph visualization



2.5 million papers by Largivis (2016)

クラスタリング・コミュニティ抽出

- ・ノードの分類やコミュニティをネットワーク情報のみから抽出

ラベル推定、教師あり学習

- ・ノードのラベルやノードの性質などに関して、一部のノードの正解データより全体を学習

リンク推定

- ・将来的にもしくは潜在的に存在するリンクを推定

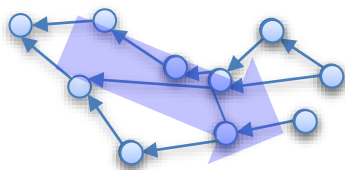
NPLを用いた新たなトレンド抽出法

- 大規模引用ネットワークからの学術領域トレンド抽出と、引用予測の精度向上を実現。

提案手法

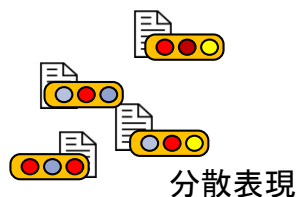
- 1) ネットワーク表現学習で得られた512次元の潜在空間上でネットワークの成長の方向性を特定
- 2) 各論文がカッティング・エッジ領域(その方向性の先端)にある度合いを指標化

引用ネットワーク



論文間のつながりの情報(潜在的に方向性の情報を含む)

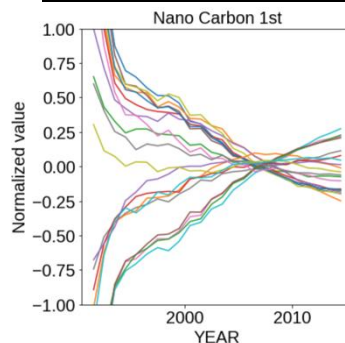
ネットワークの分散表現



各論文を高次元空間内のベクトル(分散表現)として表現

成長の方向性・カッティング・エッジ領域の検出

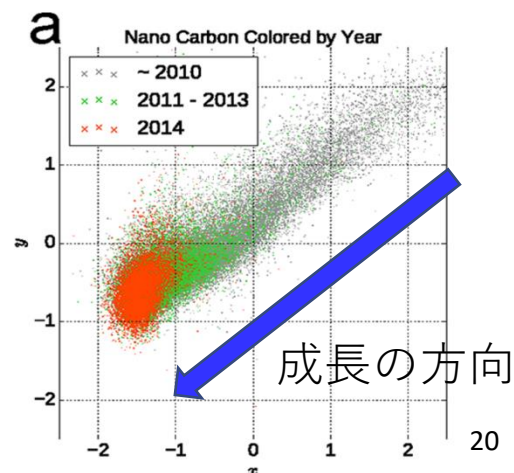
各次元における線形な論文群の移動



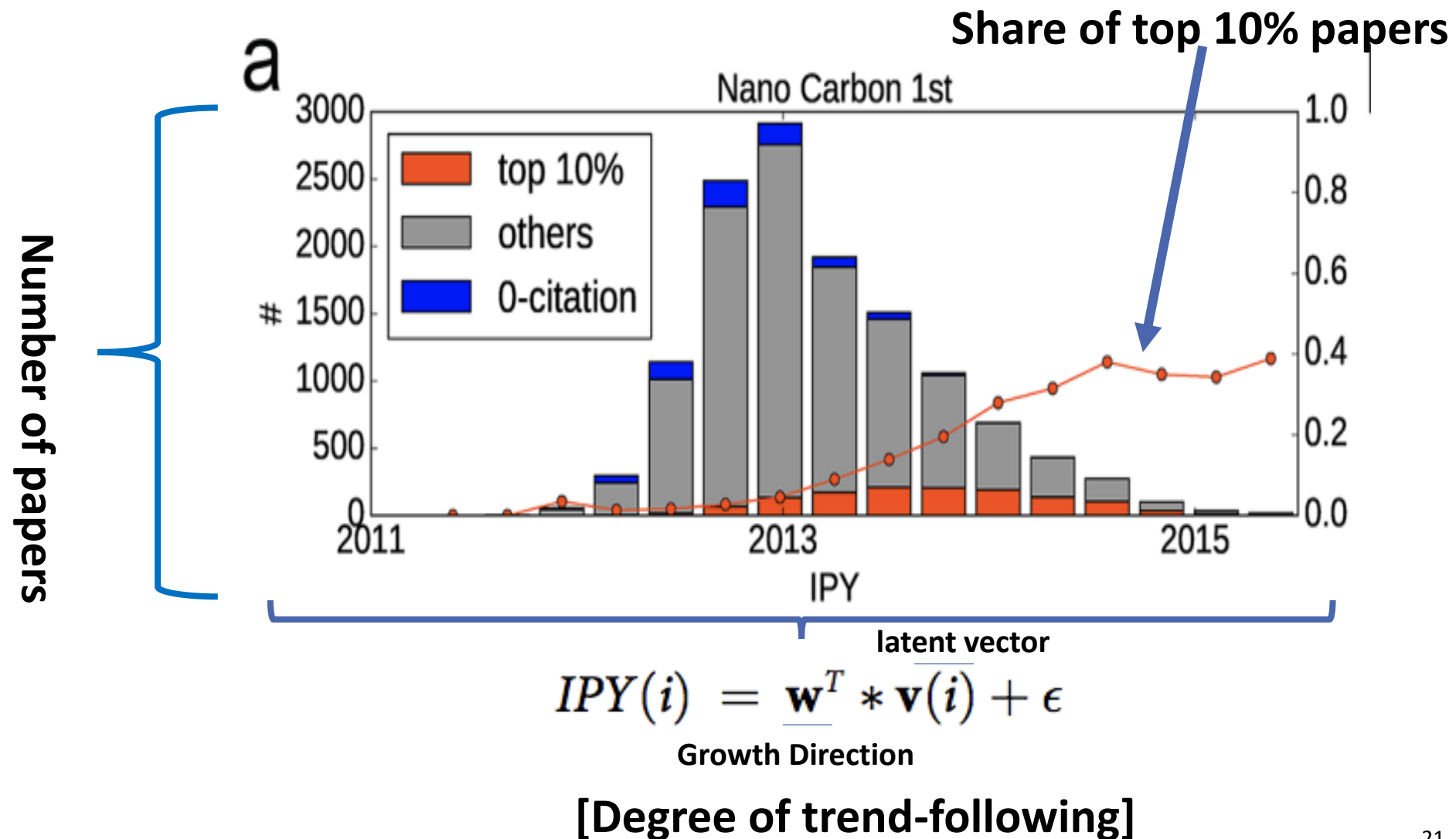
Linear regressionにより成長の方向性の推定

$$\mathbf{w}, \epsilon = \arg \min_{\mathbf{w}, \epsilon} \sum_{i \in \{Y(i) \geq C_y\}} (Y(i) - \mathbf{w}^T * \mathbf{v}(i) - \epsilon)^2$$

推定された成長の方向性 (2次元で可視化)



将来の引用数と IPY の関係



トレンドの先端にあるペーパーとトピック

Data: Paper of Nanocarbon published in October 2013 – December 2013

Predicted cutting edge paper and future citations

TITLE	JOURNAL	AUTHOR	NR	IPY
Nanocarbon Networks for Advanced Rechargeable Lithium Batteries	ACCOUNTS OF CHEMICAL RESEARCH	[Xin, S, Guo, YG, Wan, LJ]	299	2014.65
Semiconducting Monolayer Materials as a Tunable Platform for Excitonic Solar Cells	ACS NANO	[Bernardi, M, Palummo, M, Grossman, JC]	26	2014.33
Direct Optical Imaging of Graphene In Vitro by Nonlinear Femtosecond Laser Spectral Reshaping	NANO LETTERS	[Li, BL, Cheng, YW, Liu, J, Yi, CW, Brown, AS,	10	2014.27
Flexible graphene-based lithium ion batteries with ultrafast charge and discharge rates	PROCEEDINGS OF THE NATIONAL ACADEMY OF	[Li, N, Chen, ZP, Ren, WC, Li, F, Cheng,	315	2014.17
Crumpled Nitrogen-Doped Graphene Nanosheets with Ultrahigh Pore Volume for High-Performance Supercapacitor	ADVANCED MATERIALS	[Wen, ZH, Wang, XC, Mao, S, Bo, Z, Kim,	408	2014.17
Fingerprinting photoluminescence of functional groups in graphene oxide	JOURNAL OF MATERIALS CHEMISTRY	[Li, M, Cushing, SK, Zhou, XJ, Guo, SW, Wu,	72	2014.1
Quaternary Nanocomposites Consisting of Graphene, Fe ₃ O ₄ @Fe Core@Shell, and ZnO Nanoparticles: Synthesis and Excellent Electromagnetic Absorption Properties	ACS APPLIED MATERIALS & INTERFACES	[Ren, YL, Wu, HY, Lu, MM, Chen, YJ, Zhu,	150	2014.09
A high performance PZT ribbon-based nanogenerator using graphene transparent electrodes	ENERGY & ENVIRONMENTAL SCIENCE	[Kwon, J, Seung, W, Sharma, BK,	61	2014.03
Covalent Hybridization of Thiolated Graphene Sheet and Platinum Nanoparticles for Electrocatalytic Oxygen Reduction Reaction	JOURNAL OF NANOSCIENCE AND NANOTECHNOLOGY	[Ahmed, MS, Kim, D, Han, HS, Jeong, H,	32	2014.02
Nitrogen Doped Graphene-Rich Catalysts Derived from Heteroatom Polymers for Oxygen Reduction in Nonaqueous Lithium-O-2 Battery Cathodes	ACS NANO	[Wu, G, Mack, NH, Gao, W, Ma, SG, Zhong,	236	2013.89
One-pot, solvothermal synthesis of TiO ₂ -graphene composite nanosheets	JOURNAL OF COLLOID AND INTERFACE SCIENCE	[Yang, WS, Zou, XX, Xu, FG, Wang, XD,	15	2013.89
Synthesis of porous Ni@rGO nanocomposite and its synergetic effect on hydrogen sorption properties of MgH ₂	JOURNAL OF MATERIALS CHEMISTRY	[Liu, G, Wang, YJ, Qiu, FY, Li, L, Jiao, LF, Yuan, HT]	30	2013.89
Renewing Functionalized Graphene as Electrodes for High-Performance Supercapacitors	ADVANCED MATERIALS	[Jiang, Y, Luo, B, Jia, YY, Li, XL, Wang, B, Song, Q, Kang,	152	2013.88
Interfacial polymerized polyaniline/graphite oxide nanocomposites toward electrochemical energy storage	POLYMER	[Zhu, BT, Chen, MJ, Qu, HL, Zhang, X, Wei, HG, Luo, ZP,	88	2013.85
Graphene with Electrochemical Ionic-Liquid Assistance Synthesis to Explore the Metal Adsorption Ability and the High Electrical Conductivity in Graphene/Epoxy Composite	JAPANESE JOURNAL OF APPLIED PHYSICS	[Chang, CF, Chiang, YF, Chen, JR]	0	2013.82

Material and technology:

Nanocarbon network
 Graphene nano sheet
 Fingerprinting photoluminescence
 Graphene electrodes
 Graphene-Rich catalysts
 graphite oxide nanocomposites

Target and application

Lithium-ion batteries
 Supercapacitor
 Electromagnetic absorption
 Lithium-O-2 Battery Cathodes
 Energy storage

2013に予測されたトレンド上にある我々の研究

W. Gong, B. Fugetsu, Z. Wang, I. Sakata, Mingda Li et al., “Carbon nanotubes and manganese oxide hybrid nanostructures as high performance fiber supercapacitors”, *Communications Chemistry* 1 (2018): s42004-018-0017-Z

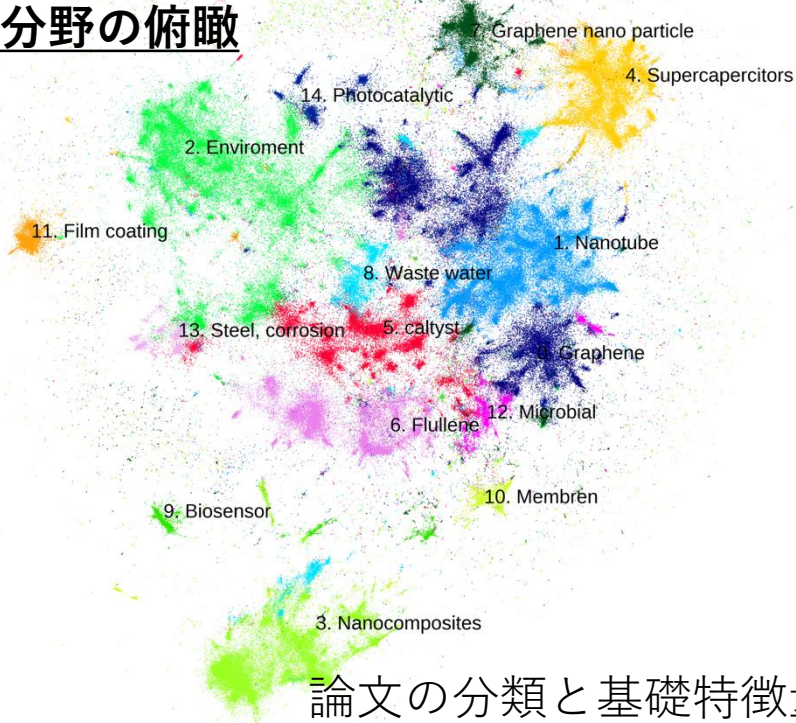


“Nanocarbon (CNT) Network” “Supercapacitor”

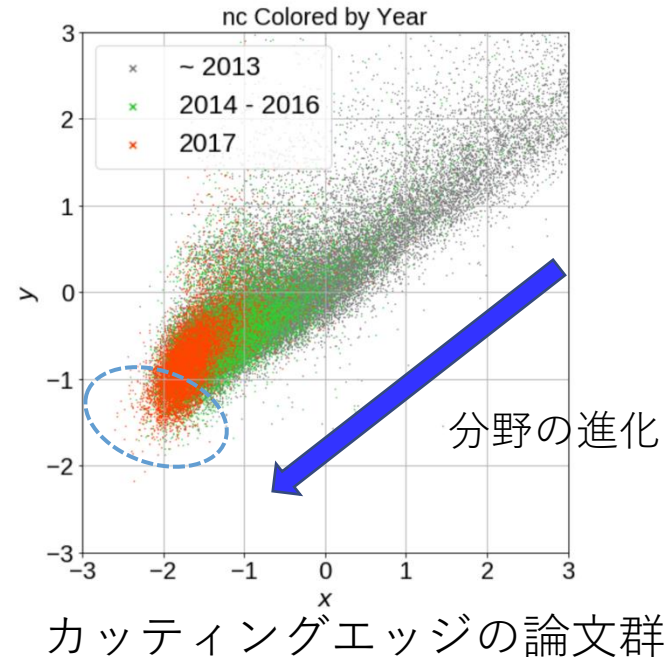


ホライズンスキヤニングへの応用 ナノカーボン領域 (2017年末)

1. 分野の俯瞰



2. 最新トレンドの把握



3. 各領域の最先端の論文の抽出

領域	論文タイトル	雑誌	概要
0 Graphene	Achieving high performance corrosion and wear resistant epoxy coatings via incorporation of noncovalent functionalized graphene	Carbon	耐腐食性・耐摩耗性の被覆
1 Nanotube	Transformation of the greenhouse gas CO2 by molten electrolysis into a wide controlled selection of carbon nanotubes	Sensors and Actuators B	温室効果ガスの選択と変換
2 Environment	Polysaccharide nanocrystals as fillers for PLA based nanocomposites	Cellulose	ナノ複合材料を用いた多糖ナノ結晶
3 Nanocomposites	Comparison of nitrogen removal and microbial properties in solid-phase denitrification systems for water purification with various	Bioresource Technology	水質浄化のための窒素除去
4 Supercapacitors	Exploring Advanced Sandwiched Arrays by Vertical Graphene and N-Doped Carbon for Enhanced Sodium Storage	Advanced Energy Materials	ナトリウム貯蔵強化のための垂直グラフェンとNドーピング炭素

開発中の技法群(一覧)

ID	内容	ユーザ	用途
1	学術俯瞰(ネットワーク分析)	広範	- 急増した知識の全体像の把握 - 知の構造変化の特定(時系列) - 知のホットスポットの特定
2	ヒートマップ(テキスト分析)	保有知識の活用先の探索	知識間の融合度合いや関連性を特定
3	未来予測モデル(ネットワーク分析&機械学習)	具体的な科学技術戦略、研究戦略担当者	ホライゾンスキャニング 分野内の有望な研究テーマ、研究チームの特定
4	トレンドの特定(表現学習)	具体的な科学技術戦略、研究戦略担当者	トレンドとカuttingエッジの特定
5	未来予測表現の抽出	具体的な科学技術戦略、研究戦略担当者	引用数(論文), H指数(著者), インパクトファクター(ジャーナル), Nature Index(組織)の未来予測
6	萌芽的タームの抽出	具体的な科学技術戦略、研究戦略担当者	かなり具体的なトレンドとカuttingエッジの特定
7	研究分野間の影響度検知(ネットワーク分析、因果推論)	研究組織マネジメント	組織内の研究資源の再配分
8	研究組織環境の評価手法(RPE)	大学等の評価者、先端技術スタートアップ等	- 研究組織の効率性の評価 - 組織改革の効果の測定 - 有望な研究チームが多く集まる組織の特定
9	マテリアル複合の予測(テキスト分析&機械学習)	企業の新材料開発、用途開発部門	マテリアルの複合化の予測
10	Start-up Readinessの推定(ネットワーク解析、多変量解析)	先端技術VC	スタートアップに加わりうる有望な研究者の早期特定

萌芽的タームの抽出

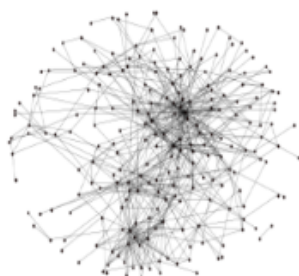
- 今後の学術分野で頻繁に使われるキーワードを予測する
- TF-IDFなどのテキスト情報のみを用いる手法の限界

1. 引用ネットワークを伝わるキーワード抽出

※Kuhnらの手法[1]を応用

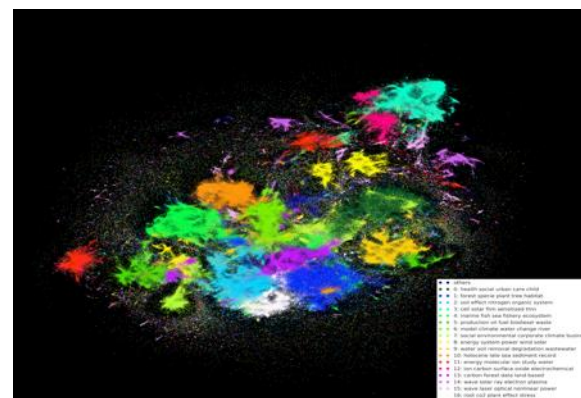
Better performance (240 papers)

Textual entailment (273 papers)



[1]Kuhn, T., Perc, M., & Helbing, D. (2014). Inheritance patterns in citation networks reveal scientific memes. *Physical Review X*, 4(4), 041036.

2. ネットワーク解析/表現学習で得られる各論文の空間的情報



効果的なキーワード抽出（クラスタごと）

	#Papers	Top 10 memes
c		
0	4255	unsupervised dependency_parsing treebank tagging german random cross-lingual conditional_random random_field induction tagger
1	3715	neural statistical_machine.translation statistical_machine bleu alignment shared shared_task parallel phrase-based smt bilingual neural_machine
2	3380	japanese cpg hpsg categorial adjoining categorial_grammar tree_adjoining adjoining_grammar tree_adjoining_grammar sign quantifier lfg
3	3219	sense disambiguation word_sense lexicon sense_disambiguation word_sense_disambiguation wordnet verb semantic_role argument sens role_labeling
4	2937	named_entity named web entity_recognition named_entity_recognition biomedical wikipedia relation_extraction clinical knowledge_base mention compound
5	2925	sentiment sentiment_analysis neural_network twitter detection topic tweet social polarity opinion medium social_medium
6	2840	dialogue spoken dialogue_system speech_recognition spoken_dialogue conversation multimodal language_generation act visual referring_expression referring

萌芽的なキーワードの抽出

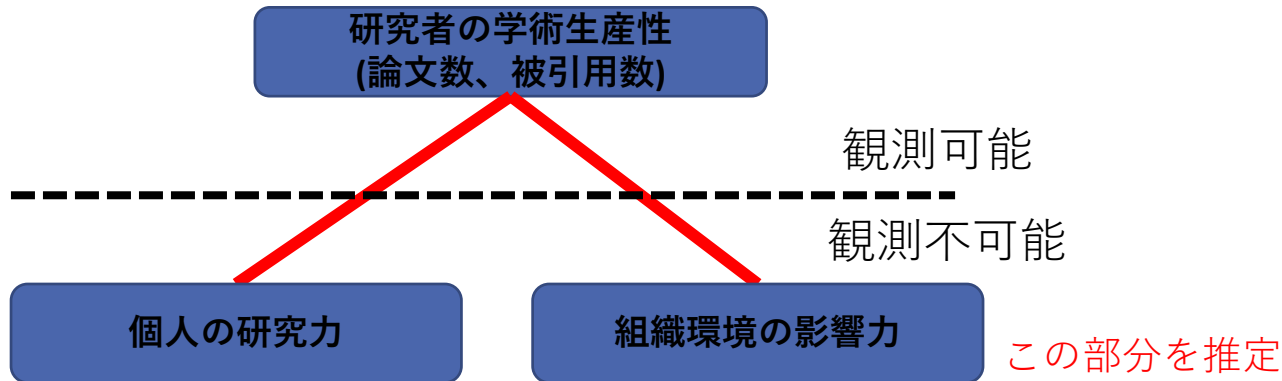
- 潜在空間内でのキーワードの伝播を観察することにより、萌芽的キーワードの伝播の特徴を確認
- ⇒萌芽的なキーワードの予測モデルを構築中
- 同時に、伝播の特性からキーワードの種類（手法/データ/目的）などを分類
- ⇒これらを合わせてシステムへ実装

ACL論文データセット内の重要キーワード抽出結果（クラスタ別）

※他に論文数1,000本以上のクラスタが10クラスタ

研究組織環境の定量的評価手法

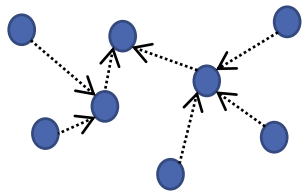
目的



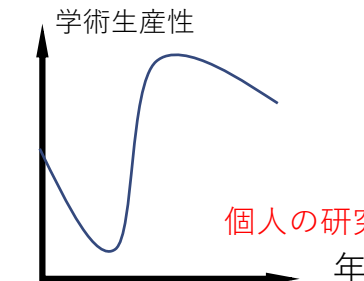
AERA 1991/5/28 朝日新聞社 p9

推定手法の概要

組織移動ネットワーク

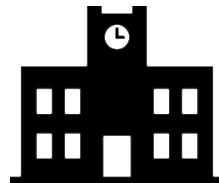


研究者の学術生産性変化



個人の研究力は一定と仮定

組織環境評価指標



→RPE(Research Productivity Enhancement)

組織のRPEが高い(組織環境が優れている)



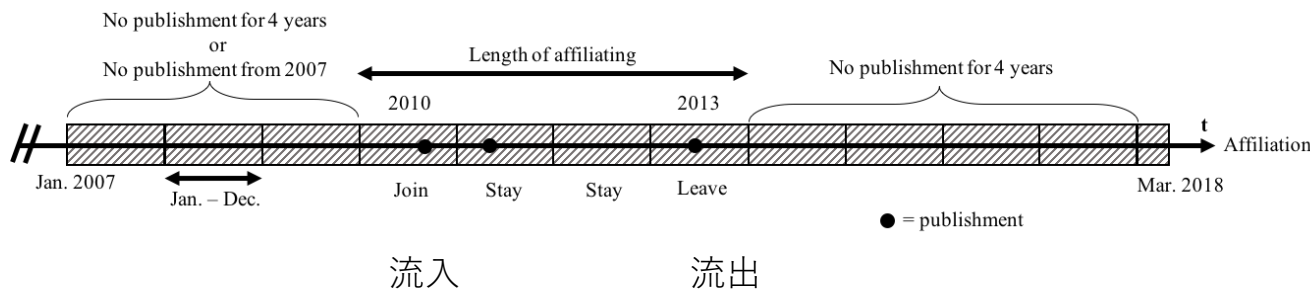
流入した研究者の学術生産性が上昇

or

流出した研究者の学術生産性が低下

提案手法: 組織移動ネットワークの作成

1. 投稿情報から研究者の流入流出を特定する。 → 組織の所属情報を取得



2. 研究者の学術生産性

$$I_{a,y,s} = \sum_{p \in P_{a,y}} (i_{p,s}) = \sum_{p \in P_{a,y}} \frac{w(p,s)}{(\alpha_p \times n_y \times (w_c(p,s) + w_y(p,s)))}$$

研究者の学術生産性 論文のインパクト 著者数 投稿年相対論文数 論文の2年内被引用数 分野平均被引用数 投稿年度平均被引用数

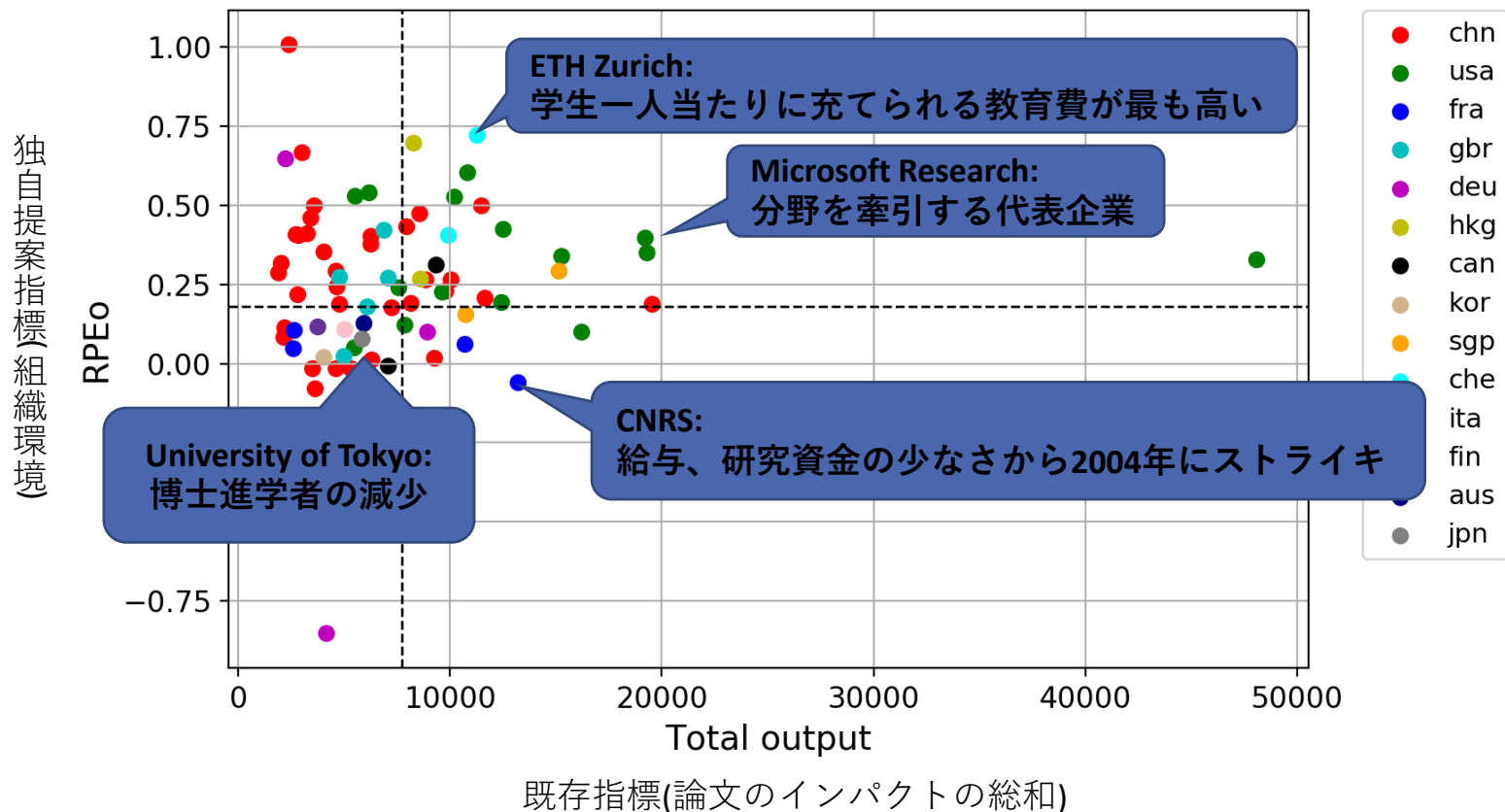
3. $g_{(\sigma,\tau),y}$: y 年に組織 σ から組織 τ に移動した研究者の学術生産性増分

$$RPE_{o,y} = g_{(*,o),y} - g_{(o,*),y}$$

$$= \log \left(\frac{(\sum_{a \in A_{y,\sigma,o}} \hat{I}_{a,y,s}^1) \times (\sum_{a \in A_{y,o,\tau}} \hat{I}_{a,y,s}^0)}{(\sum_{a \in A_{y,\sigma,o}} \hat{I}_{a,y,s}^0) \times (\sum_{a \in A_{y,o,\tau}} \hat{I}_{a,y,s}^1)} \right)$$

結果: RPEを用いた組織環境評価

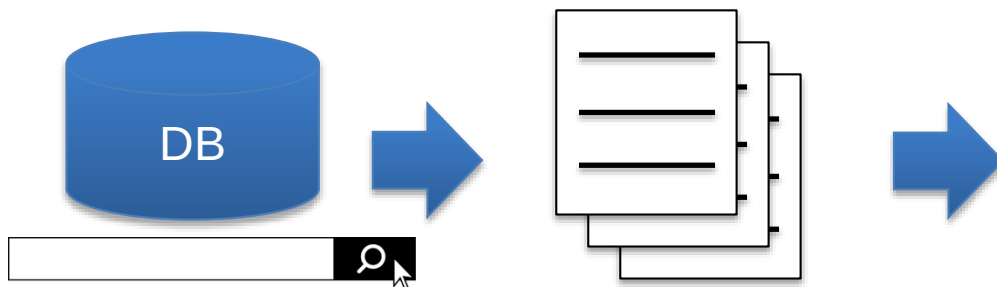
横軸は研究組織の総インパクト(既存指標) 縦軸が提案指標(全年平均)
点線は移動数が250以上の組織の平均線



組織の出す論文のインパクトと組織環境が相関せず各組織により特徴が異なり、各大学の状況を反映していると推測される。

ナノレベルでのマテリアルの複合

- Extracting frequently used material words



Extracting papers relating to composites by setting query “**nanocomposite materials**”

Rank	Word	Occurrence
1	material	1000
2	lead	900
3	silica	850
...	...	...

Frequent words in order of a number of occurrence



Rank	Word	Occurrence
1	material	1000
2	lead	900
3	silica	850
...	...	...

Searching for material words



Rank	Word	Occurrence
1	lead	900
2	silica	850
...	...	...
100	polybenzi midazole	10

Top 100 most frequently used material words

他の材料との複合研究が進むCNT、CNF

(トップ100マテリアルワードと複合数)

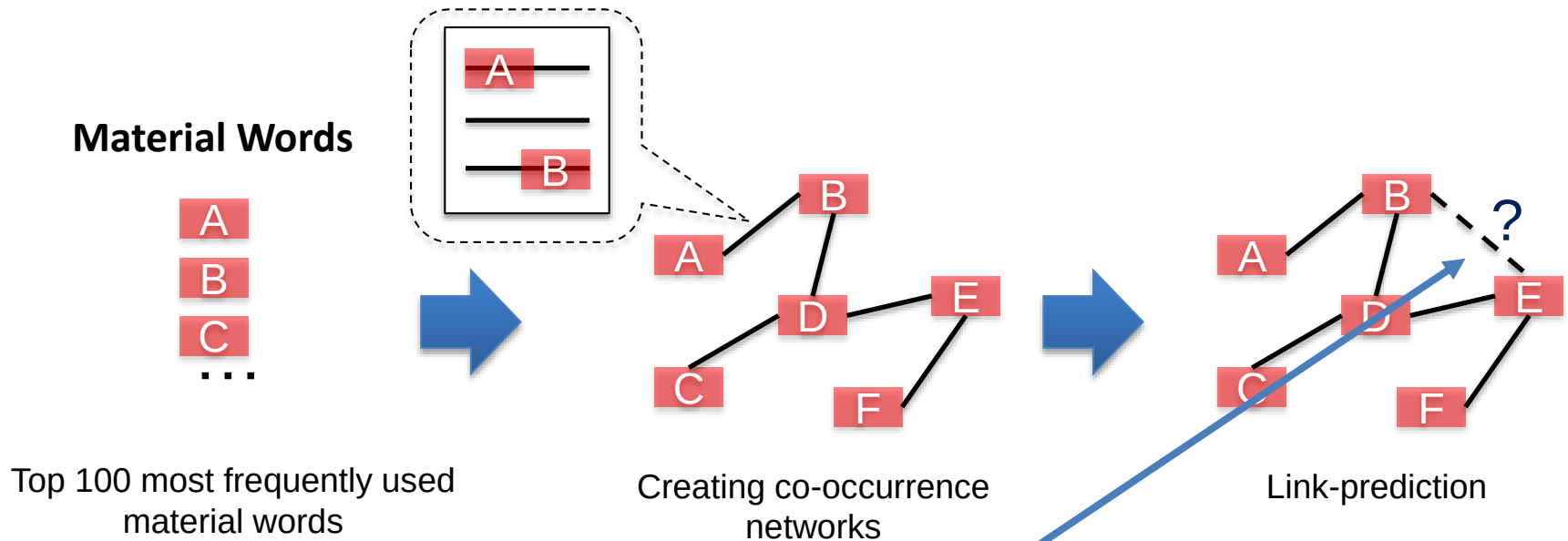
lead	97	Al ₂ O ₃	66	sodium dodecyl sulfate	41	melamine	29
silica	96	polyimide	62	CuO	40	MnO ₂	29
silver	94	aluminum	61	polydimethylsiloxane	40	MoS ₂	29
carbon nanotube	93	polypropylene	60	diamond	39	caprolactam	28
graphene	92	polyester	59	MgO	39	polydopamine	28
NiO	89	Fe ₂ O ₃	55	polymethylmethacrylate	38	polylactide	28
SiO ₂	87	calcium	54	polysaccharide	38	carboxymethyl cellulose	27
TiO ₂	87	glucose	52	rhodamine	38	collagen	27
polyaniline	84	latex	52	vinyl pyrrolidone	38	polybutadiene	27
silane	84	methylene blue	52	LiClO ₄	37	SnO ₂	27
polyethylene	82	potassium	52	polysulfone	37	vinylpyridine	27
polystyrene	81	silsesquioxane	52	polyvinylidene	37	CeO ₂	25
cellulose	78	zirconium	52	boron	36	chitin	24
Fe ₃ O ₄	76	nickel	50	polyacrylonitrile	36	polyolefin	24
graphite	76	FeCl ₃	49	cadmium	35	zirconium phosphate	24
chitosan	73	pyridine	49	polyethersulfone	35	CoFe ₂ O ₄	23
polypyrrole	72	polyelectrolyte	47	polyvinylidene fluoride	35	polyetherimide	23
epoxy	71	polyvinylpyrrolidone	46	polyacrylate	34	V ₂ O ₅	23
gold	70	polythiophene	44	polycarbonate	33	LiFePO ₄	21
polyurethane	70	AgNO ₃	43	fullerene	31	vanadium	21
ZnO	70	platinum	43	polysiloxane	31	carbon steel	20
pyrrolidone	69	polyacrylamide	43	vinylcarbazole	31	calcium phosphate	19
hydrogel	67	zeolite	43	chromium	30	MoO ₃	18
polyamide	67	BaTiO ₃	41	cyclodextrin	29	Co ₃ O ₄	14
polyvinyl alcohol	67	polycaprolactone	41	kaolinite	29	polybenzimidazole	10

複合がなされていない材料群

Materialwords /Nano-carbon	CNT		Graphene	
The words that have not co-occurred yet	Calcium phosphate	20	Calcium phosphate	20
	CeO ₂	25	CeO ₂	25
	Kaolinite	29	Chitin	24
	MoO ₃	18	Kaolinite	29
	Rhodamine	38	LiClO ₄	37
	Vanadium	21	LiFePO ₄	21
			Zirconium phosphate	24

マテリアルの複合の予測

- Predicting potential pairs of materials for composites



- Common neighbors

