

Starrydata: 論文中的 材料科学データのグラフを集めた オープンデータベース

(国研) 物質・材料研究機構
マテリアル基盤研究センター
材料設計分野 材料モデリンググループ
主任研究員 桂 ゆかり

2024年6月18日

従来技術とその問題点

Materials Informatics (MI):

材料科学へのデータ科学（機械学習）の応用

1. 機械学習には通常、大規模データが必要
2. 実験データの大規模データが入手できない
3. 計算データや自動実験データに頼る必要

現場の材料開発実験の経験や勘を学習できない

新技術の特徴・従来技術との比較

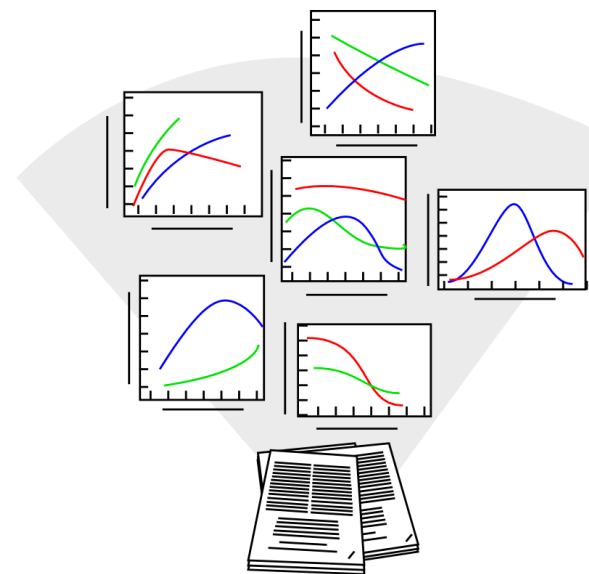
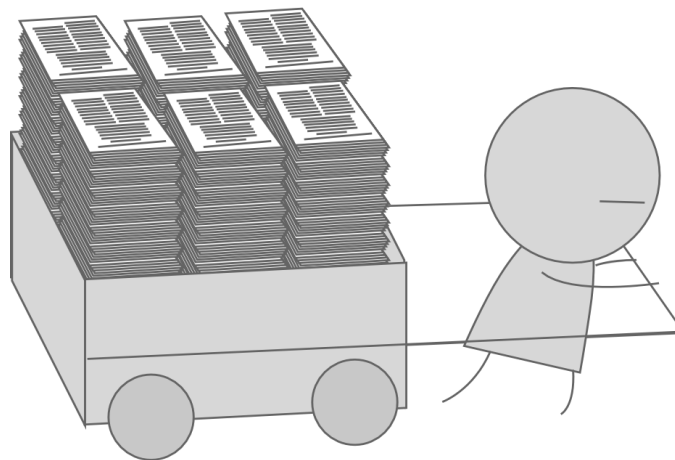
- 論文中的**グラフ**の実験データを丸ごとデータベース化
- 材料科学の実験データを試料情報とともに効率的に収録
- 全ユーザーのデータをオープンデータベースとして共有
(収集後一定期間はデータを非公開とすることも可能)

新技術の特徴・従来技術との比較

- 論文中的**グラフ**の実験データを丸ごとデータベース化



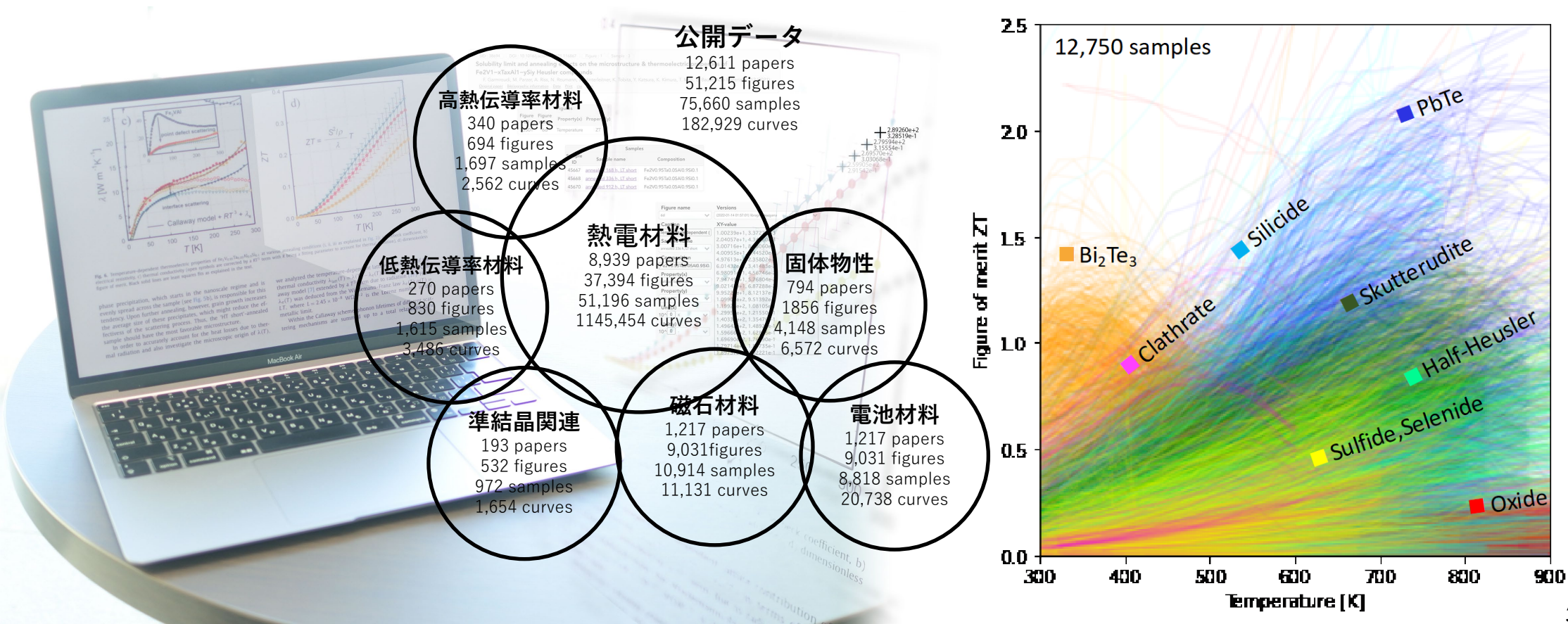
過去に行われた実験の論文の
全てに目を通すことは不可能



共同研究費で雇用した専属データ収集者が
グラフ画像から実験データを丸ごと抽出して
対象分野の大規模実験値データベースを構築

新技術の特徴・従来技術との比較

- 論文中の**グラフ**の実験データを丸ごとデータベース化



新技術の特徴・従来技術との比較

- 材料科学の実験データを
試料情報とともに効率的に収録

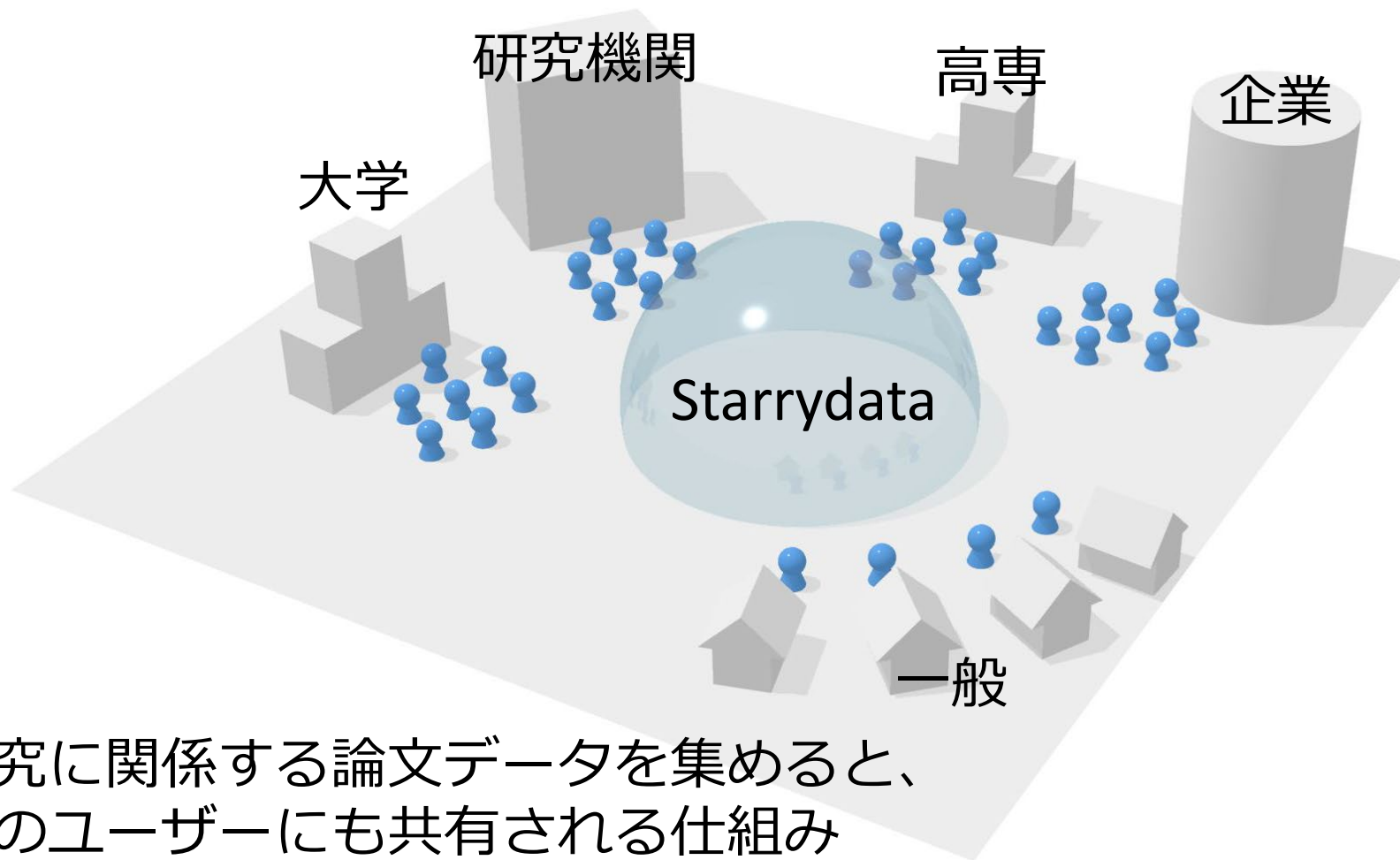
熱電材料試料の分類方法

Form		RelativeDensity	FabricationProcess	最後に行われたプロセスを優先	例
SingleCrystal	単結晶	$\geq 95\%$	MeltGrowth	溶融体からの結晶成長	ブリッジマン法、チョクラルスキー法、溶融凝固、FZ法、方向凝固法
Bulk	バルク(多結晶など)	$\geq 90\%$	FluxGrowth	(組成の違う)液体中での結晶成長	フラックス法、自己フラックス法
OrientedBulk	多結晶配向体	$\geq 80\%$	VaporGrowth	気相中での結晶成長	Vapor Transport法など
EpitaxialFilm	エピタキシャル薄膜	$\geq 70\%$	FilmGrowth	薄膜成長	スパッタリング、PLD、MBEなど
Film	薄膜	$\geq 60\%$	Melting	溶融法	アークメルト法、高周波加熱炉など
Ribbon	リボン	$\geq 50\%$	MeltQuench	液体急冷法	
Wire	ワイヤー/チューブ	$\geq 40\%$	HighPressureSynthesis	高圧合成 (GPaレベル)	マルチアンビル、ダイヤモンドアンビル
Device	デバイス(ナノテク系)	$\geq 30\%$	PressureSintering	プレス焼結 (MPaレベル)	SPS, ホットプレス, 静水圧プレス(静水圧は等方的圧力という意味なので、水熱合成とは違う)
Module	モジュール	$< 30\%$	Sintering	焼結(常圧・融点以下での熱処理)	電気炉など
Other	その他	unknown	MeltQuench	液体急冷法	
unknown	不明	StructureType	MechanicalAlloying	メカニカルアロイング	常温のボールミルだけで加熱せずに合成が完了するもの
Purity		Bi ₂ Te ₃	SoftChemical	溶液法(水溶液を加熱焼結)	錯体重合法など
SinglePhase	単相	PbTe	Hydrothermal	水熱合成・ソルボサーマル	高温・高圧の水中での合成
Impurities	不純物が存在	Skutterudite	Organic	有機合成(有機物)	
Precipitates	粒内に析出物が存在	Clathrate	Other	その他	
MetalComposite	金属との複合材料	d-silicide	unknown	不明	
Composite	金属以外との複合材料	Silicide	ElectricalMeasurement		
Other	その他	Si	SteadyState+4P	S: 定常法, σ : 四端子法	通常の熱電特性測定装置(ULVAC/アドバンス理工のZEM-3)など
unknown	不明	Perovskite	SteadyState+2P	S: 定常法, σ : 二端子法	ディスク法など
Length/Width/Thickness/GrainSize		Chalcogenide	SteadyState+vdP	S: 定常法, σ : van der Pauw法	S: 単独測定, σ : ホール係数との同時測定など
≥ 1 cm	1cm以上	SnSe	Harman	ハーマン法(S, σ , κ の同時測定)	PPMS (Quantum Design社 Physical Property Measurement System) のThermal Transport Optionなど
≥ 1 mm	1mm以上	Cobaltite	MicroProbe	マイクロプローブ法(スキャン)	熱電特性マッピングなど
≥ 100 micron	100 μ m以上	Antimonide	Other	その他	
≥ 10 micron	10 μ m以上	Organic	unknown	不明	
≥ 1 micron	1 μ m以上	ZnO	Thermal Measurement		
≥ 100 nm	100 nm以上	Heavy-Fermion	LaserFlash+DSC	レーザーフラッシュ(比熱はDSC)	
≥ 10 nm	10 nm以上	HalfHeusler	LaserFlash	レーザーフラッシュ	
< 10 nm	10 nm未満	Heusler	LaserFlash+3R	レーザーフラッシュ(比熱はDulong-Petit則)	
Other	その他	Alloy	LampFlash+DSC	ランプフラッシュ(比熱はDSC)	Xe Flashなど
unknown	不明	Sulfide	LampFlash	ランプフラッシュ	

AIにより
収集が加速

新技術の特徴・従来技術との比較

- 全ユーザーのデータをオープンデータベースとして共有



自分の研究に関する論文データを集めると、それが他のユーザーにも共有される仕組み

想定される用途

- マウスを乗せると情報が出る
目標物性の散布図
- 過去に合成された試料の化学組成の
分布とそれらの物性値の分布
- 機械学習による目標物性の予測

想定される用途

- マウスを乗せると情報が出る
目標物性の散布図

Select x-axis

- ☒ Seebeck coefficient
- ☐ log10(Electrical conductivity)
- ☐ Thermal conductivity
- ☐ Lattice thermal conductivity
- ☐ Power factor
- ☐ Z
- ☐ ZT
- ☐ Weighted mobility

Maximum

1.0e-3

-

+

Minimum

-1.0e-3

-

+

Select y-axis

- ☐ Seebeck coefficient
- ☒ log10(Electrical conductivity)
- ☐ Thermal conductivity
- ☐ Lattice thermal conductivity
- ☐ Power factor
- ☐ Z
- ☐ ZT
- ☐ Weighted mobility

Maximum

8.0e+0

-

+

Minimum

-2.0e+0

-

+

Select Temperature

- ☐ 100 K
- ☐ 200 K
- ☐ 300 K
- ☒ 400 K
- ☐ 500 K
- ☐ 600 K
- ☐ 700 K
- ☐ 800 K
- ☐ 900 K
- ☐ 1000 K

Plot size

8.00

-

+

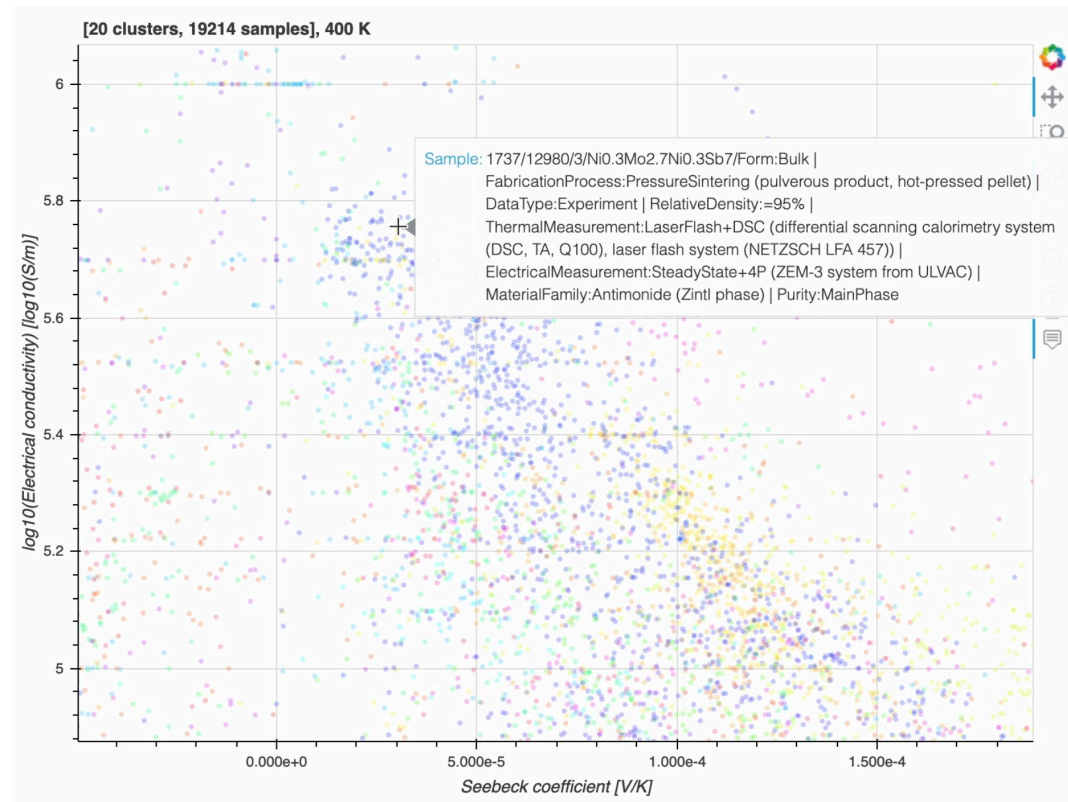
Opacity

0.20

-

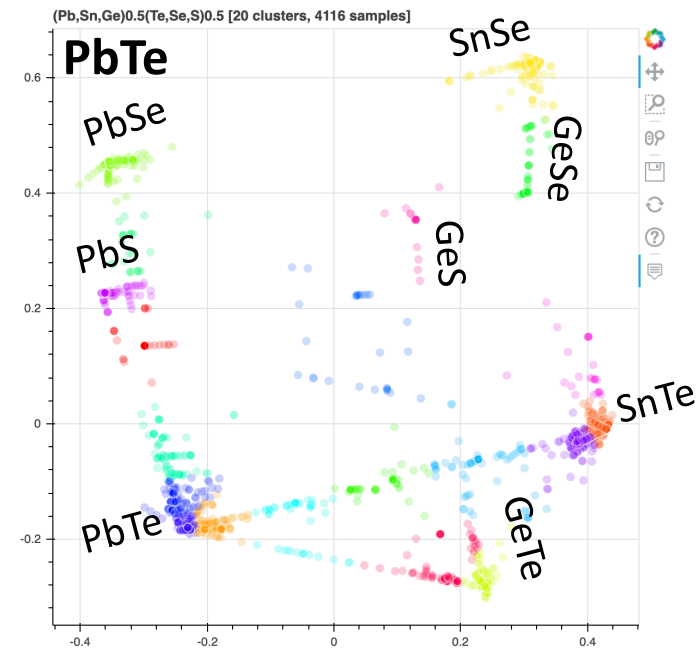
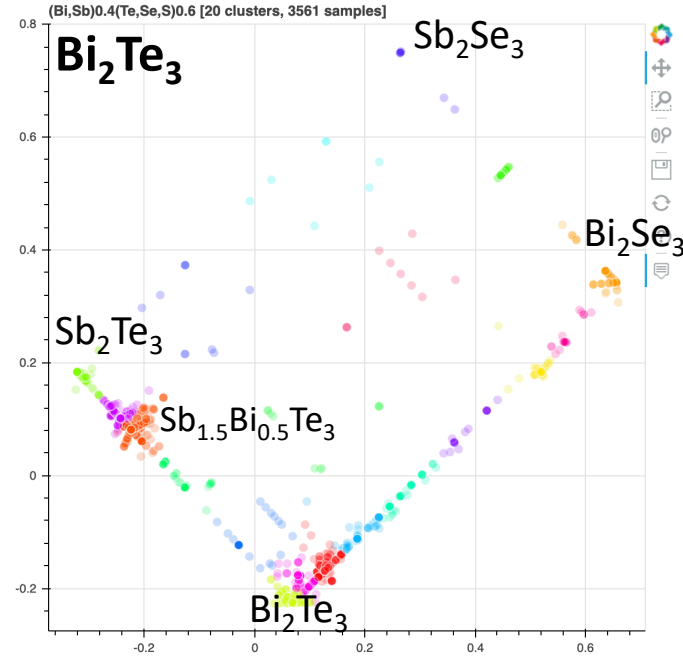
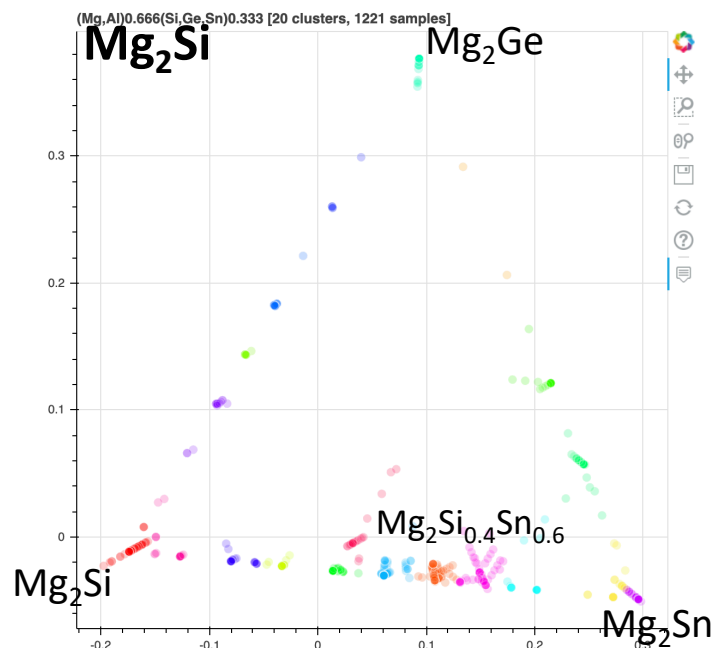
+

Generate scatter plot



想定される用途

- 過去に合成された試料の化学組成の分布とそれらの物性値の分布



想定される用途

● 機械学習による目標物性の予測

Starrydataの大規模実験データ

特性を予測したい候補物質リスト

sampleid	composition	T	S	sigma	kappa	ZT
2	Pb1Te0.99450.0055	400	4.17E-05	513404	4.44767	0.0801
3	In0.01Ag0.01Sn0.98Te1	400	7.38E-05	342811	4.85333	0.1538
5	Pb1Te0.99880.0012	400	0.000114	207089	2.27498	0.4738
7	Fe0.96Ti0.04Si2	400	0.0001991	0.12631	7.95434	0.0009
9	Pb1Te1	400	-8.08E-05	238693	2.12848	0.303
10	In0.01Sn0.99Te0.990.01	400	7.92E-05	152236	3.31108	0.1154
11	Pb0.98Na0.02Te1	400	0.0001053	174319	2.66781	0.2839
12	In0.01Ag0.005Sn0.985Te1	400	8.19E-05	284926	4.30773	0.1775
13	Pb1Te0.75Se0.25	400	-7.07E-05	91201.5	1.28134	0.1356
14	Pb1.0039Te0.75S0.15Se0.100.0078	400	-4.46E-05	204539	2.3742	0.0685
15	Pb1Te1Cl0.01	400	-5.22E-05	287416	4.444	0.0704
16	Pb1Te1In0.18	400	0.000194	10700.3	1.2213	0.1319
17	Na0.02Pb0.98Te0.85Se0.15	400	0.0001217	88889.1	1.52449	0.3654
18	Sb0.002Pb0.998Te0.88S0.12	400	0.000202	38374	0.86805	0.5951
20	Pb1.005Te0.5Se0.50.01	400	-5.10E-05	332384	3.01317	0.1149
22	Bi0.0044Pb0.9956Te0.88S0.12	400	-7.07E-05	140811	1.40811	0.2013
23	Pb0.94Mn0.06Te1Na0.04	400	1.55578	0.6039	1.55578	0.6039
24	Sn0.98Bi0.02Te1	400	4.99641	0.0593	4.99641	0.0593
25	Pb1Te1Na0.04	400	2.5171	0.3008	2.5171	0.3008
26	Pb0.9975La0.0025Te1	400	2.82719	0.2854	2.82719	0.2854
27	In0.01Sn0.99Te1	400	2.59906	0.1083	2.59906	0.1083
30	Pb1Te0.99720.0028	400	3.31312	0.2575	3.31312	0.2575
31	Pb150.9987Cl0.0013	400	-0.000136	99397.4	2.29027	0.3153
32	Na0.008Pb0.962Cd0.03Te1	400	0.0001411	91043.1	2.36821	0.2943
33	Pb1Te0.5Se0.50.0002	400	-0.000228	39435	1.63534	0.5002
34	Pb1Te0.9960.004	400	5.26E-05	459097	3.89794	0.1305
35	Fe0.98Nb0.02Si2	400	7.51E-05	812.602	8.14862	0.0002
36	Fe0.98Ti0.02Si2	400	0.0002243	362.253	8.49596	0.0008
37	Ba8Ni1.62Zn4.21Ge39.05Sn1.12	400	-9.97E-05	98295.4	0.2655	0.2655
39	Pb1.00075Zn0.01Te1.010.0015	400	-0.00011	120666	2.11952	0.275
40	Pb0.88Mn0.12Te1Na0.04	400	0.0002143	39650	1.32679	0.5488
41	In0.01Sn0.99Te0.9850.015	400	7.66E-05	216878	3.10736	0.1639
45	Pb0.99970.0002Te0.9940.006	400	-0.0002	349.35	1.61264	0.3466
46	Na0.0125Pb0.958Cd0.03Te1	400	0.0001232	116549	3.31312	0.3303
47	Fe0.99990.0001Te1.010.0005	400	-0.000177	89367.5	1.94773	0.5787
48	Bk0.00068Pb0.999312Te0.88S0.12	400	-0.000156	75298.8	0.90967	0.2962

モデル

機械学習

Output

material_id	full_formula	e_above_hull	band_gap	ZT (300 K)	ZT (400 K)	...
mp-765063	Li6V3Fe3P6H6O30	0	1.7907	?	?	?
mp-14424	Sr8Al16S32	0	3.8928	?	?	?
mp-705889	Cs8Zr6Mn2P12O48	0	2.7516	?	?	?
mp-1103722	Al4Pb2S8	0	2.5887	?	?	?
mp-1206883	K2Co1F6	0	1.1385	?	?	?
mp-3123	P8Pd8Se8	0	0.9677	?	?	?
mp-39160	Ca2Ti2Fe2P6O24	0	2.2467	?	?	?
mp-1029275	Sr2Zr2N4	0	1.2166	?	?	?
mp-1209897	Ni2Pt6O12	0	0.2204	?	?	?
mp-14388	Ho12Al20O48	0	4.7536	?	?	?
mp-769404	Y8Se8O24	0	1.8121	?	?	?
mp-15219	Rb4V2Cu2S8	0	1.2783	?	?	?
mp-14827	Li56Nb8N32	0	3.3305	?	?	?
mp-1208625	Sr2Li2V2F12	0	2.5539	?	?	?
mp-32635	Sr24P16	0	0.3855	?	?	?
mp-1211183	Li4Tm24B12O56	0	4.8715	?	?	?
mp-675651	Mg2P2S6	0	2.7986	?	?	?
mp-1106233	K2Nb2Cu4S8	0	1.7014	?	?	?
mp-24922	K4Cr4P8O28	0	2.8756	?	?	?
mp-2629	Ce7O12	0	0.0126	?	?	?
mp-3332	Na1Cr1Se2	0	0.7174	?	?	?
mp-4287	Sr8Sn4O16	0	2.6455	?	?	?
mp-1213727	Cs8P12Se24	0	1.3958	?	?	?
mp-30926	La4B20O36	0	4.4878	?	?	?
mp-1208687	Tb18P6O42	0	4.5326	?	?	?
mp-23336	Cs3Mn3Cl9	0	1.485	?	?	?
mp-1204388	Li4P4O12	0	2.7394	?	?	?
mp-7929	Rb2Pt1S2	0	1.2521	?	?	?

モデル

予測

Input

職人の勘の
ような形で
それらしい
特性を予測
↓
候補物質の
選別に利用

Output

想定される用途

● 機械学習による目標物性の予測

Starrydataチームが収集

論文からの
実験データ
(公知データ)

同一フォーマットで
実験データを整理

社内に眠る実験データ
(非公開データ)

①教師データとして開発

②性能が上がったら
教師データを追加

独自の機械学習
プログラム

予測

自社だけが予測
できる独自材料

実用化に向けた課題

- データ収集者らの人件費が足りない
- 材料開発における数値目標の達成を
実施前から約束できるわけではない
- 新しい開発スタイルの導入という
目的に投資してくれる企業が少ない

企業への期待

- MIを導入したい材料分野の開発部とNIMSとの共同研究契約の締結
- 当該分野の知識の豊富な社員様と、プログラミングのできる社員様のご協力
- データ収集者（および開発者）の人件費

企業への貢献、PRポイント

- グラフを丸ごと収集して、成長を続けている世界的に競合の現れていないデータベース
- 多様なMIに使える大規模・高品質・高情報量の実験データを管理できるWebシステム
- 熱意を持って専属データ収集者が研究動向調査と収集作業に取り組めるチームの支援体制
- 実験データMIのアイディアや進め方の提案

本技術に関する知的財産権

Starrydata webシステム

<https://www.starrydata2.org>

特許は未出願だが、論文データ収集に最適化されたユーザーインターフェイスを実現し、そのソースコードを秘匿する形で知的財産権を独占している。

産学連携の経歴

- 2018年-2020年 企業寄付金（自動車メーカーA）
- 2019年-2020年 **企業連合研究費**
（自動車用内燃機関技術研究組合(AICE) プロジェクト研究）
- 2019年-2020年 企業共同研究（材料メーカーB）
- 2021年-2023年 企業共同研究（材料メーカーC）
- 2022年-2024年 企業共同研究（材料メーカーD）
- 2022年-2024年 企業共同研究（材料メーカーD）
- 2022年 企業技術指導（材料メーカーE）
- 2023年 企業共同研究（電子部品メーカーF）
- 2024年 **企業系財団寄付金**（大倉和親記念財団）

お問い合わせ先

**国立研究開発法人物質・材料研究機構
外部連携部門 企業連携室**

企業様向け総合窓口HP(スマホ対応)

<https://technology-transfer.nims.go.jp/>

