

スマート情報技術研究センター 活動報告

知能数理研究室

2024年10月3日

成果項目

- ① マテリアルズ・インフォマティクス
 - NIMS SuperConデータを超電導材料に関する知識グラフとして構築・評価（言語処理学会年次大会2024発表）
- ② バイオメディカル・インフォマティクス
 - 医療用語のデータベースへのリンク（BMI論文誌掲載）
 - 同じ用語の表現を統一した、電子カルテからの用語抽出試行（言語処理学会年次大会2024発表）
- ③ スマート情報技術
 - 大規模言語モデルを利用して「関係の間の関連性」を扱う手法の開発（言語処理学会年次大会2024発表→若手奨励賞受賞）

知識グラフによる転移温度予測

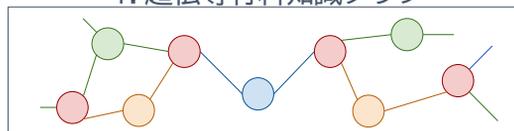
1. 超伝導材料知識グラフ：材料が他の頂点を介して繋がるように作成
2. 頂点のベクトル化：頂点を種類に応じた方法でベクトル化
3. 転移温度予測：グラフニューラルネットワークにより転移温度を予測

超伝導材料の表データ

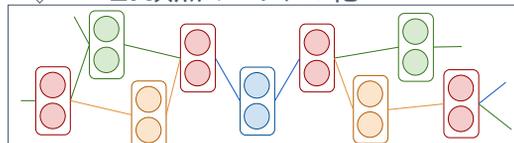
処理後組成式	処理前組成式	構造名	論文題名
$\text{La}_{1.7}\text{Sr}_{0.3}\text{Cu}_1\text{O}_4$	$(\text{La},\text{Sr})_2\text{Cu}_1\text{O}_4$	T型	xxx
$\text{La}_{1.9}\text{Sr}_{0.1}\text{Cu}_1\text{O}_4$	$(\text{La},\text{Sr})_2\text{Cu}_1\text{O}_4$	T*型	yyy



1. 超伝導材料知識グラフ



2. 頂点のベクトル化



3. 転移温度予測



GNN



3

実験設定

転移温度の予測誤差（5回分の平均）を比較する

- 評価指標

- 平均絶対値誤差 (MAE) $\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|$
- 二乗平均平方根誤差 (RMSE) $\text{RMSE} = \frac{1}{N} \sqrt{\sum_{i=1}^N (\hat{y}_i - y_i)^2}$

- 超伝導材料データとして **SuperCon** [NIMS2022] を使用

- 22,629件の事例（組成式，転移温度，構造名，論文題名）を使用
- 事例を報告年で分割し，
過去の事例から未来の事例を予測

- 比較対象

- CrabNet（既存の物性値予測手法）

	事例数	報告年
訓練データ	16,269件	1913年～2003年
開発データ	4,395件	2004年～2013年
テストデータ	1,965件	2014年～2021年

4

実験結果

- 開発データでは提案手法の性能が既存手法を上回った
- テストデータでは既存手法と同等の性能を得ることができた

手法	開発データ [K]		テストデータ [K]	
	MAE (↓)	RMSE (↓)	MAE (↓)	RMSE (↓)
CrabNet	9.26±0.09	14.62±0.15	7.20±0.18	17.05±0.20
提案手法	9.00±0.12	14.12±0.31	7.40±0.18	16.98±0.16

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|$$

$$\text{RMSE} = \frac{1}{N} \sqrt{\sum_{i=1}^N (\hat{y}_i - y_i)^2}$$