

蛍光体中の Eu 価数の文献情報と機械学習

Text mining and machine learning of Eu valence states in phosphor materials

物材機構, °小山 幸典, 広崎 尚登, 高力 由香子, 武田 隆史

NIMS, °Yukinori Koyama, Naoto Hirosaki, Yukako Kohriki, Takashi Takeda

E-mail: KOYAMA.Yukinori@nims.go.jp

Eu は蛍光体の発光中心として広く研究されている元素のひとつである。発光中心としてホスト結晶中に導入された Eu は 2 価と 3 価の状態をとりうるが、2 価と 3 価では発光メカニズムが異なるため、所望の価数となるホスト化合物を選択する必要がある。しかし、ホスト結晶中における Eu の価数は自明ではない。そこで、蛍光体中の Eu の価数に関する情報を文献から抽出し、ホスト化合物に添加された Eu の価数に対する機械学習を行った。

蛍光体材料は「 $\text{SrLiAl}_3\text{N}_4\text{:Eu}^{2+}$ 」のように、ホスト化合物($\text{SrLiAl}_3\text{N}_4$)と発光中心(Eu^{2+})をコロンの(:)で挟んで表記することが多い。そこで、蛍光体に関する約 12 万件の学術論文から上記のような表記を機械的に抽出した。発光中心は Eu^{2+} と Eu^{3+} に限定した。ホスト化合物の化学式は機械的に抽出したが、不適切な化学式をルールベースで除外した。この結果、 Eu^{2+} に対して 21,607 件の論文から 831 種のホスト化合物を、 Eu^{3+} に対して 24,430 件の論文から 1,261 種のホスト化合物を抽出した。387 種のホスト化合物は Eu^{2+} と Eu^{3+} の両方の論文があった。

このデータを使って、ホスト化合物の組成から Eu の価数を分類する機械学習モデルを構築した。本データの特徴として、両方の価数で報告されているホスト化合物が多数存在することと、ホスト化合物の論文数が大きく異なることが挙げられる。そこで、訓練データのラベルを論文数が多い方の価数で二値化する場合と、両方の価数とも論文数の比で正解として与える場合を検討した。また、サンプル重みとして論文数に比例した重みを与えることも検討した。検証では、未知化合物に対する実験の実施を判断する場面を想定し、予測結果は二値化し、正答ラベルは論文数比で与え、サンプル重みは均一とした。特徴量は化学組成に由来する汎用特徴量を使用し、特徴量選択をした後に、2 価を正例としてロジスティック回帰で学習した。10 分割交差検証の検証データに対する評価指標を Table 1 に示す。正答ラベルの取り扱いによる指標の差は小さかった。また、論文数のサンプル重みを与えると recall が向上し、precision は悪化した。詳細は当日に議論する。本研究は JST CREST (JPMJCR19J2)の支援を受けたものである。

Table 1. Summary of predictive performance for validation data in the 10-fold cross validation.

Ground truth	Sample weight	Accuracy	Precision	Recall	F-score
Binary	Even	0.847	0.798	0.811	0.803
Binary	Num. of papers	0.839	0.761	0.851	0.803
Fractional	Even	0.846	0.804	0.798	0.799
Fractional	Num. of papers	0.836	0.759	0.847	0.799