

調査研究助成課題の成果概要(その1)

熱電変換材料に関する過去の実験データのオープンデータベース化

東京大学大学院新領域創成科学研究科物質系専攻・助教
桂 ゆかり

1. はじめに

熱電材料は、温度差から電気を作り出す熱電発電に利用できる材料です。身の回りの熱エネルギーを電気として回収できれば、電池交換のいらぬ電源として利用できます。また、熱電材料に電流を流すと、片面が温まり、反対側の面が冷たくなります。このため、ポータブル冷蔵庫や精密温度制御素子としても利用されています。

熱電材料を広く普及させるためには、変換効率を高くする必要があります。変換効率は、性能指数 ZT が高くなるほど高くなります。 ZT はゼーベック係数¹⁾ S 、電気伝導率 σ 、熱伝導率 κ 、温度 T の関数として、 $S^2\sigma T/\kappa$ と表せます。よって、 ZT を高くするには S 、 σ が大きく、 κ が低い材料を探せばいいということになります。熱電材料の候補物質は非常に多いのですが、実際に高い ZT を示す材料を探すことは簡単なことではありません。 S 、 σ 、 κ のどれかを良くしようとすると、別のどれかが悪くなってしまうからです。

そんな複雑な熱電材料研究をもっと効率的にしたい—そんな思いで、熱電材料に関する過去の論文から実験データを集め、ひとつのデータベースにまとめる調査研究を始めました。

2. 調査研究の概要

熱電材料の実験の論文には、 T を横軸として、縦軸に S 、 σ 、 κ 、 ZT を示したグラフが標準的に載っています。そこで、これらのグラフから、元の数値データを復元しました。論文中のグラフは画像データですが、専用のソフトでデータ点を読み取れば、元の数値データを復元できます。

このため、私たちはStarrydata2 (<https://www.starrydata2.org/>)というWebシステムを開発してきました²⁾。これは論文を自動的に読んでくれるロボットではないのですが、人間が論文中のグラフから数値データを回収し説明を記入して保存する作業を大幅に効率化してくれるシステムです(図1)。大学や研究所など、論文を無料で読むことのできるインターネット環境で使えば、効率的にたくさんの論文データを集めることができます。Starrydata2では、ユーザーが論文からデータを取得すると、その数値データは別のユーザーも使うことができます。こうしているような人が自分の興味のあるデータを集めるためにStarrydata2を使ってくると、いろいろな種類のデータが集まっていき、最終的には大きなデータベースに成長すると期待できます。

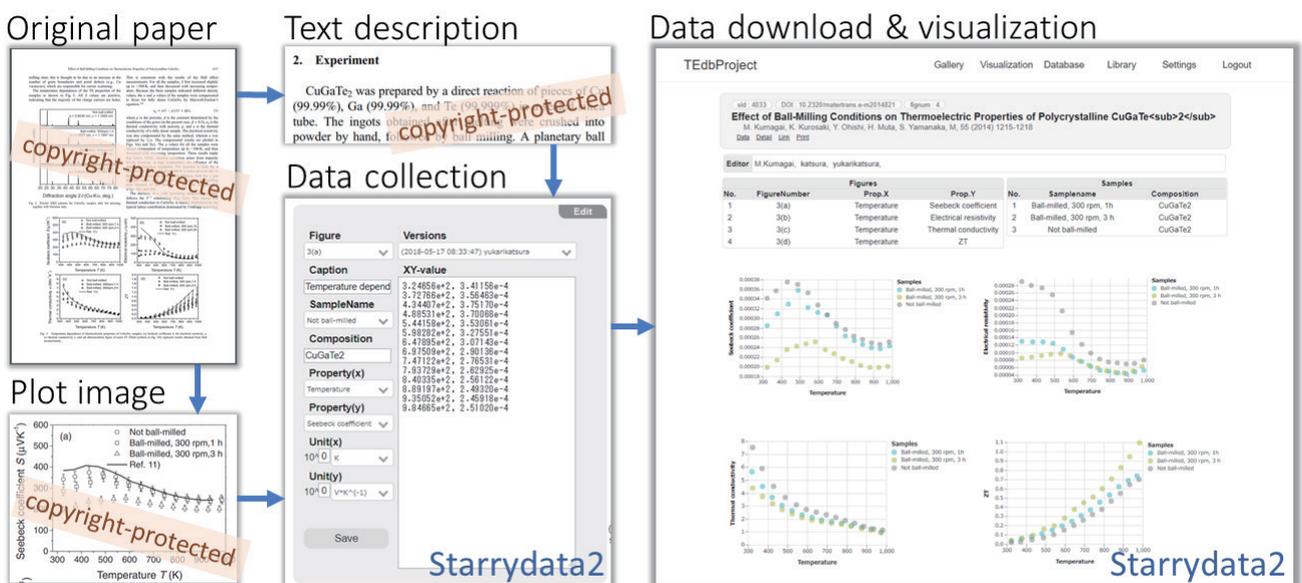


図1 Starrydata2のデータ収集の仕組み²⁾

(論文中のグラフから数値データを集め、それらの客観的な情報を記入することで、著作権を侵害せずにデータベースを作成できます。)

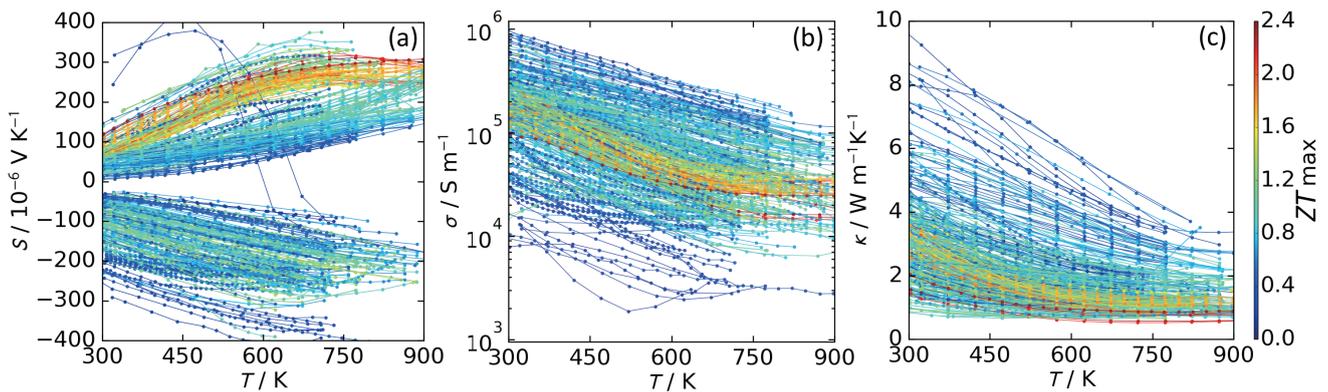


図2 論文から収集した熱電特性 (S , σ , κ) の温度依存性のグラフの例²⁾
 (434個の試料のデータを各試料の性能指数 ZT の最大値に応じて色分けし、一枚のグラフ上に示すことに成功しました。)

Starrydata2には2019年現在、4,000本以上の論文から集めた20,000試料以上の熱電材料の試料のデータが登録されています。カーブ数で数えると70,000本以上です。その一部を図2に示します。これは熱電材料のデータベースとして世界最大の規模のもので、試料の選別をできるだけ行わず、特性の悪い試料も含めて幅広い試料のデータを集めました。また、測定方法の正確さに基づくデータの選別も行いませんでした。これにより、大量のデータを集めることに成功しました。

これらの熱電特性のデータをx軸を $\log\sigma$ 、y軸を S としてプロットすると、物質ごとに異なる直線にまとまることが確認でき、電子構造の違いが観察できました。また、第一原理計算³⁾によるシミュレーションデータと実験データの整合性も直接確認できるようになりました。そして第一原理計算データと実験データを組み合わせることで、電子緩和時間⁴⁾の推定に始めて成功し、これが長くなるほど ZT が高くなるという傾向を発見しました。このように大量のデータを集めることで熱電材料の研究が深くなり、さまざまな新しい知見を得ることができるようになりました。今後は機械学習によって、熱電材料開発に役立つツールを作りたいと考えております。

3. おわりに

この高度にデジタル化された時代にあっても、論文に掲載されるほど重要な実験データが未だにデジタル検索・入手できないという現状は、残念なことだと思っています。多額の税金を投じてようやく得た大事

なデータなのに、読み取るのが面倒くさいからとか、自分のデータ以外信頼できないからなどという理由で葬り去ってしまうのは、非常にもったいないことだとも思っています。

Starrydata2に集めたデータはすべて無償で公開しています。この理由は、大学や研究機関でダウンロードした論文が商用利用できないためでもあります。多くのユーザーにとって使いやすいデータにするためでもあります。有料のデータベースでは、データを盗まれないように管理するシステムを作るために余計な手間がかかります。その上、何万件ものデータを手元にダウンロードして機械学習をかけたり、その結果を元データと合わせて人に渡したり、論文に載せたり、新しくデータベースを作り出したりといった利用が一切できなくなってしまうのです。

データは研究のインフラであり、無償であるべきものです。無償データベースという研究のインフラがきちんと整備されることによって、革新的なデータ科学研究と、革新的なビジネスが成長する環境が整うのです。今後は熱電材料以外にも、いろいろな材料科学分野のデータベースを作っていくたいです。そして材料科学に限らず、いろいろな科学分野のデータも集めていきたいと考えています。

機械学習応用を見据えた地道なデータ収集作業は、まだ重要性が広く認識されていないものの、今後さまざまな分野で重要な役割を担うと思います。これらを支えられるのは調査研究であると考えています。本調査研究を支援いただきました新技術振興渡辺記念会の皆様に、深く感謝申し上げます。

1) ゼーベック係数：熱起電力とも言い、温度差1Kあたり何Vの電圧を発生できるかを表します。
 2) Starrydata2については、次の論文で詳しく説明しています。
 Yukari Katsura, et al., "Data-driven analysis of electron relaxation times in PbTe-type thermoelectric materials", Sci. Tech. Adv. Mater., 20, 1 (2019) 511-520. DOI: 10.1080/14686996.2019.1603885
 3) 第一原理計算：物質の結晶構造に対して、量子力学の原理に基づいて物質中の電子の状態をシミュレーションする計算。
 4) 電子緩和時間：物質中の電気伝導を担う電子が、何にもぶつからずにまっすぐに進むことのできる時間。