

磁気転移温度に潜む「数字の偏り」を発見

— 5万6千件の文献抽出データを精査し、キュリー温度がベンフォード則から外れることを確認 —

発表者：

和達 大樹（兵庫県立大学 大学院理学研究科 教授）

発表のポイント：

- 科学論文から AI で自動抽出された 56,037 件の磁気転移温度データを、単位、温度範囲、物理的文脈、重複の観点から段階的に精査しました。
- 同じ材料の重複を除いた後も、強磁性体のキュリー温度はベンフォード則から明確に外れる一方、反強磁性体のネール温度は同則に大きく近づくことを明らかにしました。
- この違いは、磁気転移温度がもつ物理的な尺度制約と、研究対象の選ばれ方に由来する文献選択効果が、大規模データの数字の分布に現れる可能性を示します。

発表概要：

多くの自然・社会データでは、数値の最初の桁は均等には現れず、「1」から始まる数が多く、「9」から始まる数が少なくなるベンフォード則に従うことがあります。本研究では、AIを用いて科学論文から自動抽出された大規模な磁気転移温度データを対象に、強磁性体のキュリー温度と反強磁性体のネール温度がこの法則に従うかを調べました。

56,037件の元データに対し、ケルビン単位のみを残す、0～2000 Kの範囲に限定する、磁気転移を明示した記録を選ぶ、非磁性の文脈を除く、同一材料の重複をまとめる、という処理を行いました。最終的にキュリー温度2,754件、ネール温度1,323件からなる4,077件の材料データセットを構築しました。

解析の結果、キュリー温度はデータを精査してもベンフォード則からのずれが残り、ネール温度は重複除去後に同則へ大きく近づきました。これは単なるAI抽出の誤りや同一材料の重複だけでは説明できず、磁性材料の物理的な温度スケールや、室温付近など応用上重要な材料が多く研究・報告される傾向が関係していると考えられます。本成果は、AIにより構築された科学データベースの特性評価に加え、物性値の背後にある物理的制約や研究活動の偏りを統計的に読み解く新しい視点を提供します。

発表内容：

① 研究の背景

ベンフォード則（注1）は、さまざまな桁数にまたがる数値データにおいて、最初の有効数字が特定の割合で現れるという経験則です。先頭数字が1である確率は約30.1%である一方、9である確率は約4.6%となります。会計、地球科学、生物学など多様な分野で知られ、近年は大規模科学データの性質や品質を調べる指標としても注目されています。

一方、自然言語処理と材料インフォマティクスの発展により、膨大な論文から材料名や物性値を自動抽出したデータベースが構築されるようになりました。ただし、自動抽出されたデータには、単位の不一致、文脈の取り違い、同一材料の重複などが含まれます。また、データを十分に清浄化した後に残る数字の偏りが、データベースの誤差によるものか、材料の物理的性質によるものかは明らかではありませんでした。

② データの精査

本研究では、文献から自動抽出された56,037件の磁気転移温度記録を出発点としました。ケルビン単位の記録のみを残し、 $0 < T < 2000$ Kに限定したうえで、キュリー温度またはネール温度であることが文脈上明確な記録を選別しました。さらに、強誘電転移や超伝導転移など明らかに非磁性の文脈を除き、同一材料の複数記録は中央値を代表値として統合しました（図1）。

ベンフォード則と解析データの構築過程

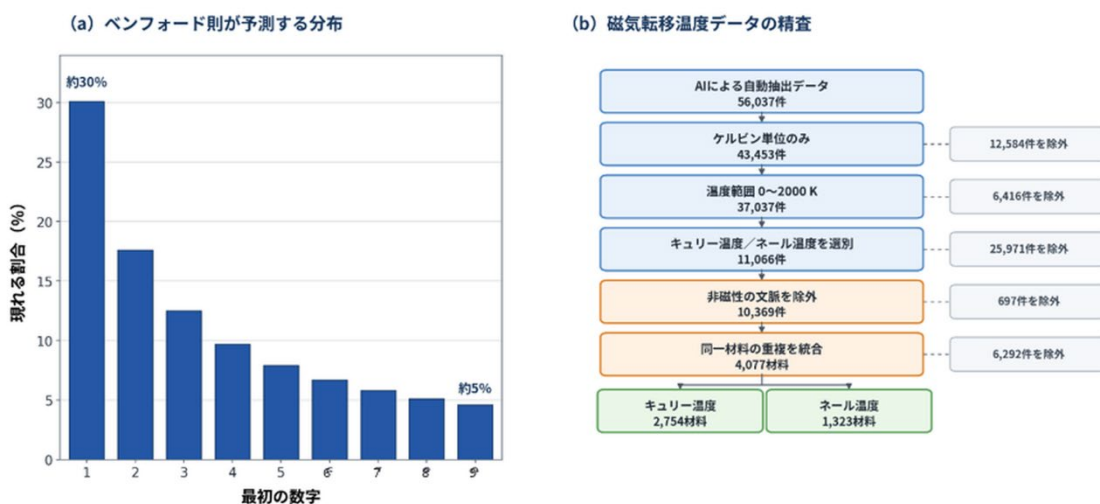


図1 ベンフォード則と解析データの構築過程。(a) ベンフォード則が予測する先頭数字の出現割合。「1」から始まる数は約30%であるのに対し、「9」から始まる数は約5%となります。(b) AIにより科学論文から自動抽出された56,037件の記録を段階的に精査し、キュリー温度2,754材料、ネール温度1,323材料からなる4,077材料のデータセットを構築しました。

③ 研究内容と成果

各転移温度の最初の有効数字を取り出し、ベンフォード則が予測する分布と比較しました。また、観測された数字の分布がベンフォード則からどの程度離れているかを、統計的な指標を用いて数値化しました。この指標は、0に近いほどベンフォード則によく一致することを表します。

キュリー温度では、先頭数字「1」がベンフォード則の予測より少なく、「3」や「5」が相対的に多いという系統的な偏りが、データ精査と材料重複除去の後にも残りました。一方、ネール温度の分布は、重複除去後にベンフォード則へ大きく近づきました（図2）。

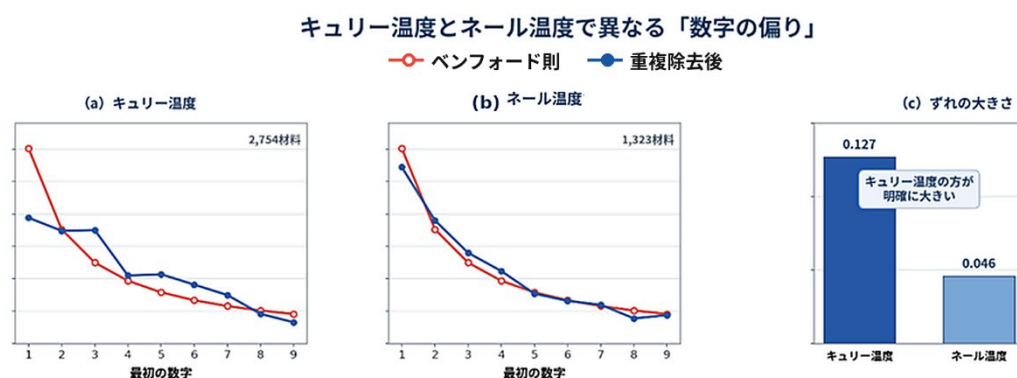


図2 キュリー温度とネール温度の先頭数字分布。(a) キュリー温度、(b) ネール温度。赤線はベンフォード則、青線は重複除去後のデータを示します。(c) ベンフォード則からのずれの大きさ。値が0に近いほどベンフォード則によく一致します。キュリー温度ではデータ精査後も明確なずれが残る一方、ネール温度はベンフォード則に大きく近づきました。

この対照は、キュリー温度の偏りが、AI抽出時の誤りや同一材料の重複だけでは説明できないことを示します。キュリー温度の分布が有限の温度範囲に集中するという物理的な尺度制約に加え、室温付近など技術的に重要な材料が選択的に研究・報告される傾向があります。こうした物理的尺度制約と文献選択効果の組み合わせが、ベンフォード則の前提となる広い尺度不変性を崩している可能性があります。

④ 社会的意義・今後の予定

AIを用いた文献データの自動収集は、材料探索を大きく加速する一方、得られたデータの重複、曖昧さ、隠れた偏りを把握することが不可欠です。本研究は、先頭数字という単純な統計量が、データ品質の点検だけでなく、材料物性の分布や研究対象の選ばれ方を調べる手掛かりにもなることを示しました。

今後は、磁気転移温度以外の材料物性データにも同様の解析を広げ、物理的制約、データベース構造、文献選択効果がどのように統計分布へ反映されるかを検証します。予測を目的とする機械学習と、データセット全体を診断する統計解析を組み合わせることで、材

料インフォマティクスの信頼性と解釈可能性の向上が期待されます。

発表雑誌：

雑誌名：Physica A: Statistical Mechanics and its Applications

論文タイトル：Scale-constrained Curie temperature distributions and the breakdown of Benford's law

著者：Hiroki Wadati

DOI番号：<https://doi.org/10.1016/j.physa.2026.132027>

掲載日：2026年9月15日（火）

問い合わせ先：

【研究内容に関すること】

兵庫県公立大学法人兵庫県立大学

大学院理学研究科

教授 和達 大樹 TEL：0791-58-0157 E-mail：wadati@sci.u-hyogo.ac.jp

【報道に関すること】

兵庫県公立大学法人兵庫県立大学

播磨理学キャンパス経営部 総務課

TEL：0791-58-0101（内線212） E-mail：soumu_harima@ofc.u-hyogo.ac.jp

用語解説：

注1：ベンフォード則

多様な桁数にまたがる数値データにおいて、最初の有効数字dが現れる確率を $P(d)=\log_{10}(1+1/d)$ で表す法則。1から始まる数は約30.1%、9から始まる数は約4.6%と予測されます。すべてのデータが従う法則ではなく、広い範囲に分布し、尺度を変えても性質が大きく変わらないデータで現れやすいとされています。

注2：キュリー温度／ネール温度

キュリー温度は、強磁性体が自発磁化を失う転移温度です。ネール温度は、反強磁性体における規則的なスピン配列が失われる転移温度です。

注3：AI抽出データベース

ここでのAI抽出は、生成AIが数値を作ったという意味ではありません。自然言語処理を用いて、既存の科学論文に記載された材料名と磁気転移温度を自動的に読み取り、整理したデータベースを指します。

研究費

本研究は、JSPS科研費 JP25H01251の支援を受けて実施されました。