What can data science tell us about finding new superconductors?

Turab Lookman* and Alejandro Lopez-Bezanilla**

*AI Materials Research LLC, Santa Fe, NM 87501, USA  
**Theoretical Division T-4, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

What can data science tell us about finding new superconductors? We examine this question in light of a study in this issue of Patterns by Liu et al., who find that the superconducting transition temperature and certain computed energy intervals of the valence band are correlated.

Can data science guide researchers toward understanding superconductivity or discover new superconductors? We examine this question in light of a study in this issue of Patterns by Liu et al., who find that the superconducting transition temperature and certain computed energy intervals of the valence band are correlated.

Very few topics in condensed matter physics have generated as much excitement, disappointment, and speculation about their fundamentals as superconductivity. This exotic state of matter has fascinated scientists ever since its discovery in mercury by Kammerlingh Onnes in 1911. 1 By offering little resistance to the flow of electric current below a transition temperature, Tc, superconductors find applications in multiple technological areas. Finding new superconductors at normal pressures with ever-higher Tc’s so that they approach ambient temperatures remains an outstanding problem with much emphasis on deciphering the mechanisms underpinning superconductivity. The ground-breaking theory of Bardeen, Cooper, and Schrieffer (BCS) in 1957 2 provides a robust framework in which the emergence of superconductivity is associated with the occurrence of a small but effective attractive potential between charges in a material and the subsequent creation of electron or hole pairs. Intensive (temperature) and extensive (pressure) thermodynamic variables play a fundamental role in this pairing mechanism, allowing the charge carriers to stick together and flow with no dissipation below Tc. The attractive interaction between charges is mediated by phonons, namely the quantized vibrational modes of the atomic lattice. Above Tc, the pairing mechanism vanishes. This mutual attraction between charges is the key ingredient of the microscopic interpretation of superconductivity and assumes that one charge polarizes the lattice while the other charge interacts with this polarization to glue the carriers. Putting aside the physical phenomena and experimental results accumulated over a hundred years, we can ask, is it possible to rely on data and infer hidden patterns and correlations between physical properties to glean aspects of the mechanism, and perhaps even predict superconductivity in new or existing materials? There have been a number of studies that have recently applied data science tools to address this problem 3,4 that have identified the number of d valence electrons and atomic mass as relevant features. What Liu et al. 5 have proposed in this issue of Patterns is a close correspondence between the upper limit of Tc (Tc,max) in classes or groups of superconductors and the energy levels of valence electrons.

Liu et al. 5 conclude that the energy interval between the last occupied and first unoccupied valence bands of the material appears to be linearly correlated on a log-log scale with Tc,max in groups of related superconductors, such as cuprates and Fe-based and heavy fermions. The features incorporated in the data model relate to electron-electron interactions between orbitals as a component necessary to interpret high-Tc superconductivity. Does data analysis lead to this result? Not directly. The authors use data of roughly 1,000 superconductors from publicly available databases, and invoke about 440 features related to orbital properties. After following a recipe of down selecting features that includes eliminating those that are highly correlated among themselves, they construct a surrogate data model with four remaining features and Tc,max. This is merely a fitting exercise to the data and is done in a “black-box” mode. The model is checked against the Tc,max prediction of various materials but, astonishingly, also of insulators rather than superconductors. Therein lies the difficulty of using features that are by no means complete, however excellent the fitting. Unless a fair set of relevant features, which are still unknown, in addition to an accurate modeling framework describing the essential physics are taken into account, one needs to be skeptical of much of the model building and predictions of new superconductors from data alone. Such models certainly risk of failing in answering the inverse design question: what superconductor will exhibit Tc,max? However, Liu et al. 5 speculate from the statistics of their features that a parameter related to the band energy differences, ΔE, could be relevant. For each subgroup of superconductors, they then perform first-principles electronic calculations (DFT) to evaluate ΔE using the code CASTEP to observe the correlation with Tc,max. This is where care is needed in the use of approximate information from computationally modeled superconductors. Better approximations by choosing appropriate energy functional and/or energy corrections often need to be considered.

Can the correlation of Tc,max with ΔE be explained? It is difficult to say given that relevant aspects such as spin fluctuations, phonon coupling, and polaronic effects are absent in the statistics. It is undeniable, though, that an interesting trend has been observed—even if it may be fortuitous—as the authors’ rationale is based on observing that certain features connected to the orbitals have similar values for the cuprates. Given that these seem to be correlated, one would have
anticipated that one of these would have been filtered in the feature engineering step. Therefore, care needs to be exercised with the authors’ claim that “additional inter-orbital electron-electron interaction should be considered in the interpretation of high-Tc superconductivity.” Also, Tc and Tc,max are not the same, and the relationship between the two needs to be clarified.

Discovery from data has had some occasional successes, largely for materials where the relevant physics is reasonably well understood. These include shape memory alloys, ferroelectrics, and thermoelectrics to name a few systems where one can gain some confidence about the identification of some of the essential features.6 However, that is not the case for superconductivity. It has not been demonstrated that orbital energies suffice for estimating Tc,max and whether phonon-mediated mechanisms contained somehow in the data escaped the analysis. In any case, this should motivate the construction of a theory relating Tc,max and energy intervals. Identifying descriptors, as this study has tried to do, or exploring promising regions of the chemical space via high-throughput screening are worthwhile tasks to be carried out. A more thorough analysis based on the use of other physical quantities, even if only computational, such as dielectric constants,7 elastic moduli,8 electron-phonon coupling magnitude,9 that could directly or indirectly relate to parameters defining superconductivity would benefit from understanding the contribution of individual features. Data science methods can certainly help to reduce computational costs associated with DFT calculations by building surrogate models from high-fidelity data. However, the jury is very much out on whether data science can provide insights into the mechanisms underlying superconductivity or predict ambient temperature superconductors. Perhaps the need is not just to use even more data but incorporate a lot more science.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Kammerling Onnes, H. (1911). Leiden Communications: 120b, 122b, 124c. Proceedings of the Koninklijke Akademie van Wetenschappen te Amsterdam, 1479–1481.
2. Bardeen, J., Cooper, L.N., and Schrieffer, J.R. (1957). Theory of superconductivity. Phys. Rev. 108, 1175–1204. doi:10.1103/physrev.108.1175.
3. Stanev, V., Oses, C., Kuene, A.G., Rodrigue, E., Paglione, J., Curtaro, S., and Takeuchi, I. (2018). Machine learning modeling of superconducting critical temperature. npj Computational Materials 4, 29. doi:10.1038/s41524-018-0085-8.
4. Liu, Z.-L., Kang, P., Zhu, Y., Liu, L., and Guo, H. (2020). Material informatics for layered high-Tc superconductors. Apl. Mater. 8, 061104. doi:10.1063/5.0004641.
5. Liu, Y., Huang, H., Yuan, J., Zhang, Y., Feng, H., Chen, N., Yang, L., Teng, J., Jin, K., Xue, D., and Su, Y. (2022). Upper limit of the transition temperature of superconducting materials. Patterns 3, 100609.
6. Lookman, T., Balachandran, P.V., Xue, D., and Yuan, R. (2019). Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. npj Computational Materials 5, 21. doi:10.1038/s41524-019-0153-8.
7. Morita, K., Davies, D.W., Butler, K.T., and Walsh, A. (2020). Modeling the dielectric constants of crystals using machine learning. J. Chem. Phys. 153, 024503. doi:10.1063/5.0013136.
8. de Jong, M., Chen, W., Angsten, T., Jain, A., Notestine, R., Gamst, A., Sluiter, M., Krishna Ande, C., van der Zwaag, S., Plata, J.J., et al. (2015). Charting the complete elastic properties of inorganic crystalline compounds. Sci. Data 2, 150009. doi:10.1038/sdata.2015.9.
9. Sun, Y., Zhang, F., Wang, C.Z., Ho, K.M., Mazin, I.I., and Antropov, V. (2022). Electron-phonon coupling strength from ab initio frozen-phonon approach. Phys. Rev. Materials 6, 074801. doi:10.1103/physrevmaterials.6.074801.