Research Article

Visualization of superconducting materials

Anton Matasov[1,2] · Varvara Krasavina[3]

Received: 31 March 2020 / Accepted: 24 July 2020 / Published online: 31 July 2020

Abstract
The most extensive database of superconducting materials has been pre-processed. On this database, methods for reducing dimensions, pairwise display of features, a heat map and Pearson's correlation criterion, visualization of features with color are considered. The dependences of the critical temperature of superconductors on atomic mass, radius, ionization energy, electron affinity, heat of fusion, thermal conductivity, and valence are considered.

Keywords Superconductivity · Superconductor · Machine learning · Critical temperature

1 Introduction

At the moment, there is no general theory of superconductivity. The most complete Bardeen–Cooper–Schrieffer theory [1, 2] does not allow predicting the most important parameter of superconductors—their critical temperature. The problem of the transition from conventional low-temperature superconductors to unconventional high-temperature superconductors remains unexplored.

The inability to predict the critical temperature of superconducting materials, as well as the dependence of the critical temperature on the parameters of the material, leads to the difficult practical use of these materials. Also, to create superconducting materials with desired properties, it is important to understand the mutual dependencies of other material parameters besides the critical temperature from each other. The problem of the transition from one class of superconductors to another is key in the search for new higher-temperature superconducting materials.

Therefore, in order to be able to control the properties of superconducting materials, to develop a general theory of superconductivity, it is necessary to generalize the standard theory [3, 4], create a new theory [5] or search for an alternative approach.

As an alternative approach in this work, it is proposed to use data visualization methods to solve the problems of determining the dependences of the critical temperature on material parameters, the dependence of the parameters of the material from each other, considering the separation or connection of classes of superconducting materials.

Visualization methods in the field of superconductivity were used mainly as visualization of magnetic structures [6–8] and other physical effects [9–11]. There are several works [12–16] in which statistical analysis is carried out in superconducting materials, and models for predicting the critical temperature are constructed. But the general approach in applying standard methods of machine learning visualization to a database of superconducting materials has not yet been applied. Thus, this approach was first applied in this paper.

The standard methods that are often used for data preprocessing in machine learning [17–21] can be considered the methods from Seaborn library in Python, as well as various methods of reducing the dimension. In particular, the methods from Seaborn allow to determine the

✉ Anton Matasov, matasov_av93@mail.ru; Varvara Krasavina, varvara.krasavina@gmail.com | [1]Department of Physics and Technologies of Electrical Materials and Components, National Research University "MPEI", Moscow, Russia. [2]MIREA – Russian Technological University (RTU MIREA), Moscow, Russia. [3]Epoch8, Moscow, Russia.

SN Applied Sciences (2020) 2:1463 | https://doi.org/10.1007/s42452-020-03260-6

correlation between the target variable and all features in the data, which is useful before building the model. Dimension reduction methods allow to determine hidden dependencies in the data, identify clusters. Thus, these all these methods should be applicable to solve the problems posed.

For the appropriate application of visualization methods, it is necessary to use the most comprehensive database of superconducting materials. In [16] data is based on open materials published by the National Institute of Materials Science in Japan (NIMS). At the moment, this is the widest base of superconducting materials, the data of which are taken mainly from journal articles and books. The data contains the values of 83 attributes for 21,263 superconducting materials.

The NIMS data were pre-processed for further research in [16], which makes their analysis more convenient in the future than the NIMS data. As a result of the pre-treatment, materials that had some missing features were removed. Also, preliminary processing included the formation of new features based on existing ones. Atomic mass, density, first ionization energy, atomic radius, density, electron affinity, fusion heat, thermal conductivity, valence were taken as initial features (Table 1). That is, the chemical formula of the material was considered and based on the meanings of each features, the mean, weighted mean, geometric mean, weighted geometric mean, entropy, entropy weighted, range, weighted range, standard deviation and weighted standard deviation were calculated (Table 2).

As a result, the initial 10 available features, name of the material, number of chemical elements and critical temperature add up to 83 features.

We suppose that this approach to the formation of features is quite general and suitable for the study of superconducting materials due to the general uncertainty of the dependence of the critical temperature. Therefore, the data of work [16] are the initial data for our study. The application of visualization methods to this most extensive database of superconducting materials is also a specialty of this article.

Thus, to solve the problems posed, the structure of the manuscript is as follows. Pre-processing of initial data, which presents an algorithm for pre-processing and obtaining data for research. The results of applying the methods of reducing the dimension, methods of pairwise mapping of features, heat map, method of visualization of features with colors, its algorithm, which was first developed in this work. Discussion of the results of applying visualization methods and conclusion.

Table 1 Initial features of atoms and their units

Feature	Units
Atomic mass	Atomic mass unit (amu)
Ionization energy	Kilojoule per mole (kJ/mol)
Atomic radius	Picometer (pm)
Density	Kilogram per cubic meter (kg/m^3)
Electron affinity energy	Kilojoule per mole (kJ/mol)
Heat of fusion	Kilojoule per mole (kJ/mol)
Thermal conductivity	Watt per Meter–Kelvin (W/(m * K))
Valence	–

Table 2 An example of calculation formulas for the formation of characteristic values for the number of chemical elements equal to 2

(t_1, t_2—feature value for the atoms in the compound; p_1, p_2—ratios of the atoms in the compound.)

$w_1 = \frac{t_1}{t_1+t_2}$ $\quad A = \frac{p_1 w_1}{p_1 w_1 + p_2 w_2}$

$w_2 = \frac{t_1}{t_1+t_2}$ $\quad B = \frac{p_2 w_2}{p_1 w_1 + p_2 w_2}$

Feature	Expression
Mean	$\mu = \frac{t_1+t_2}{2}$
Weighted mean	$\nu = p_1 t_1 + p_2 t_2$
Geometric mean	$\sqrt{t_1 t_2}$
Weighted geometric mean	$t_1^{p_1} * t_2^{p_2}$
Entropy	$-w_1 \ln(w_1) - w_2 \ln(w_2)$
Entropy weighted	$-A\ln(A) - B\ln(B)$
Range	$t_1 - t_2 (t_1 > t_2)$
Weighted range	$p_1 t_1 - p_2 t_2$
Standard deviation	$\left(0.5(t_1-\mu)^2 + (t_2-\mu)^2\right)^{0.5}$
Weighted standard deviation	$\left(p_1(t_1-\nu)^2 + p_2(t_2-\nu)^2\right)^{0.5}$

2 Visualisation of superconducting materials

2.1 Data preparation

After a preliminary analysis of the data, it was noticed that the initial data of [16] contain data for the same materials, with all the attributes besides the critical temperature having the same values. Apparently, this is due to the fact that during the preliminary data processing, there were materials with different oxygen contents in their composition (for example, compounds of the type $YBa_2Cu_3O_{7-x}$), but the exact oxygen content was not indicated, therefore, as a result, the same material names appeared in the data. The difference for such materials only at the critical temperature can lead to the fact that the model will additionally make mistakes on the same materials, and the likelihood of retraining on such data increases.

Therefore, to solve this problem, the following data processing algorithm is proposed:

1. Create a database s_0, which we will fill, different from the main one. As the first line, you can take the line for the first material from the main base.
2. Further in the cycle, until the condition that the source data base s_1 has no elements is satisfied:
   - 2.1 We filter the source database by the name of the first material. If the same materials are available, we get a list.
   - 2.2 At this stage, you should process the list so that at the end you get one line of values for this material. As mentioned above, the lines differ from each other only in the value of the critical temperature. Therefore, three processing options are proposed, using as the new critical temperature the average, minimum or maximum value of the critical temperature of this list.
   - 2.3 Add a line to the data s_0.
   - 2.4 Delete from the data s_1 all materials that have the name of the first element of point 2.1
   - 2.5 Transition to the new first data material s_1.

Thus, we obtain three new data sets: with a minimum (sc_min), an average (sc_mean) and a maximum value (sc_max) of the critical temperature for each material. Each of which now contains 15,542 original materials. It turns out that the old data contained 5721 materials, which are repeated or 27% of the total data, which can significantly affect the model in terms of its adequate component of predictive ability.

In addition, outliers may still be present in the data. In order to get rid of them, we will remove all materials for which the values of the features are not included in the range of values of 3 standard deviations from the mean value of any feature.

After removing the outliers from the data, we get 12,105, 12,113, 12,119 materials for sc_min, sc_mean, sc_max, respectively.

2.2 Dimension reduction methods

Often, in machine learning problems, before training a model, it is convenient to use data visualization methods or methods to reduce the dimension of a data space to find hidden dependencies in data.

For this purpose, we will use four methods of reducing the dimension: the random projection method, the principal component method (PCA), the multidimensional scaling method (MDS), and the t-SNE method. A detailed description of these methods can be found in the works [17–20].

The random projection method is a linear method in which the formation of new features occurs on the basis of old ones with some weight, which is determined by the normal distribution. The principal component method is also linear and, unlike the random projection method, weights are expressed through singular vectors of the "feature-signs" matrix. These methods cannot determine nonlinear dependencies in the data; therefore, the last two methods are of the greatest interest.

The multidimensional scaling method requires that pairwise distances between objects vary as little as possible. But this problem is quite complicated mathematically and the final separation between classes in two-dimensional space is obtained so that the classes are located close to each other. The t-SNE method assumes that distances between similar classes are reduced, and between dissimilar classes are increased, which makes this method the most informative.

Due to the similarity of the data sc_min, sc_mean, sc_max, we will consider the visualization of one received data frame. The results of data visualization are presented in Fig. 1. The values of the feature "number of chemical elements in material" n are indicated by the corresponding color.

This feature was chosen to visualize a decrease in dimension, since it can be considered categorical and in view of the general effect of this feature on the critical temperature. For low temperature superconductors n mainly changes from 1 to 3. For high temperature superconductors $n \approx 4$ to 9. Thus, it is convenient to track classifications of materials.

On the axes of Fig. 1, the relative distance between superconducting materials in the space of reduced dimension is plotted. The presence of classes can be determined

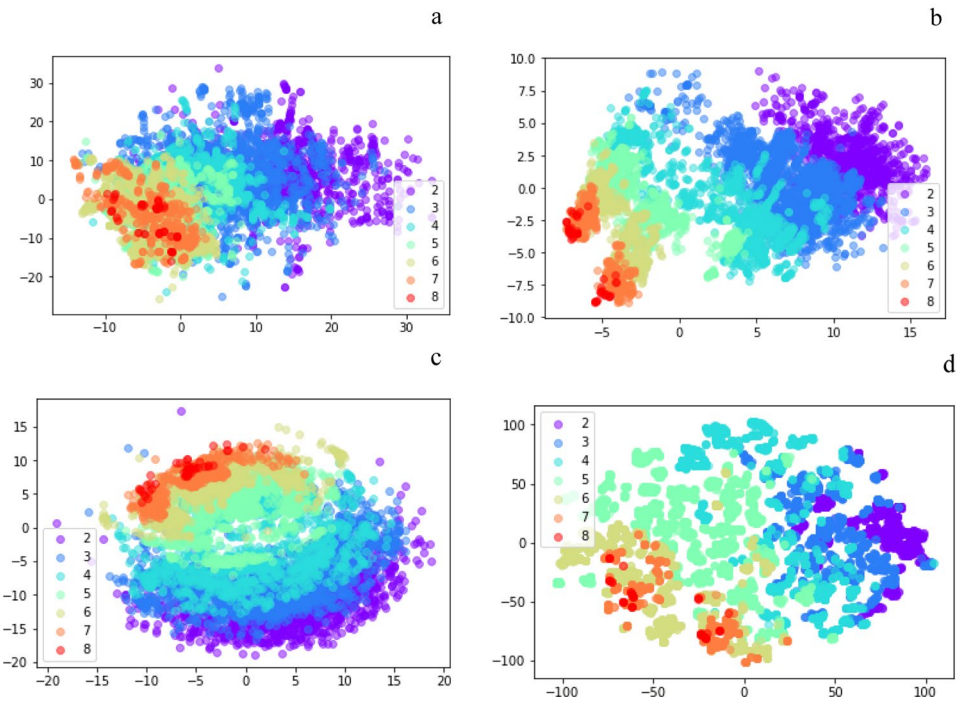

Fig. 1 Visualization of data on superconducting materials by random projection (**a**), PCA (**b**), MDS (**c**), t-SNE (**d**)

by visual assessment of proximity and the accumulation of points on the plane of space.

Note that the application of all four methods of reducing the dimension is due to a more complex analysis of the results of visualization.

2.3 Pairwise mapping of features

For pairwise mapping of features, the method "pairplot()" from Seaborn in Python was used [21]. This method allows to build paired relationships in a data set. By default, this function creates a grid of axes so that each variable in the data is shared along the Y axis in one row and the X axis in one column. The diagonal axes are processed differently, drawing a graph showing the one-dimensional data distribution for the variable in this column.

Due to the fact that in these 83 signs it is extremely difficult to present pairwise dependencies clearly. Therefore, we present the dependencies between the critical temperature and all "mean", "range" features. (Fig. 2). These features were selected for visualization due to the simplicity of evaluating these parameters for researchers who are engaged in the synthesis of new superconducting materials. The critical temperatures and the corresponding features in units of measurement from Table 1 are plotted along the dependency axes. The construction of these dependences is important for a visual assessment of the dependence of the critical temperature on various parameters of the material, as well as determining the ranges of values of the attributes in which the critical temperature takes the highest values.

Note that the method "pairplot()" allows one to construct the distribution of features (Fig. 3), which allows to determine what value each of the features takes most often in superconducting materials. Consider the distribution data and values for the "mean" and "range" features. Due to the fact that the peaks are wider for "mean" features, we give approximately the region of the most frequent values, and for "range" features values of peaks (Table 3).

The most frequent values of the parameters of superconducting materials are required to assess the probability of synthesis of new possible superconducting materials. Using the obtained graphic results and the most common values, it will be extremely convenient for researchers to make a preliminary calculation of the parameters for their new possible superconducting material. Characteristic distribution data represents the dependence of the frequency of parameter values relative to the maximum value on parameter values in units of measurement from Table 1.

2.4 Heat map and Pearson criterion

The "heat map" method presents data in the form of a color matrix or heat map [21].

We suppose that, first of all, the physical meaning of the features in the data, excluding the critical temperature, and their relationship to each other, reflect

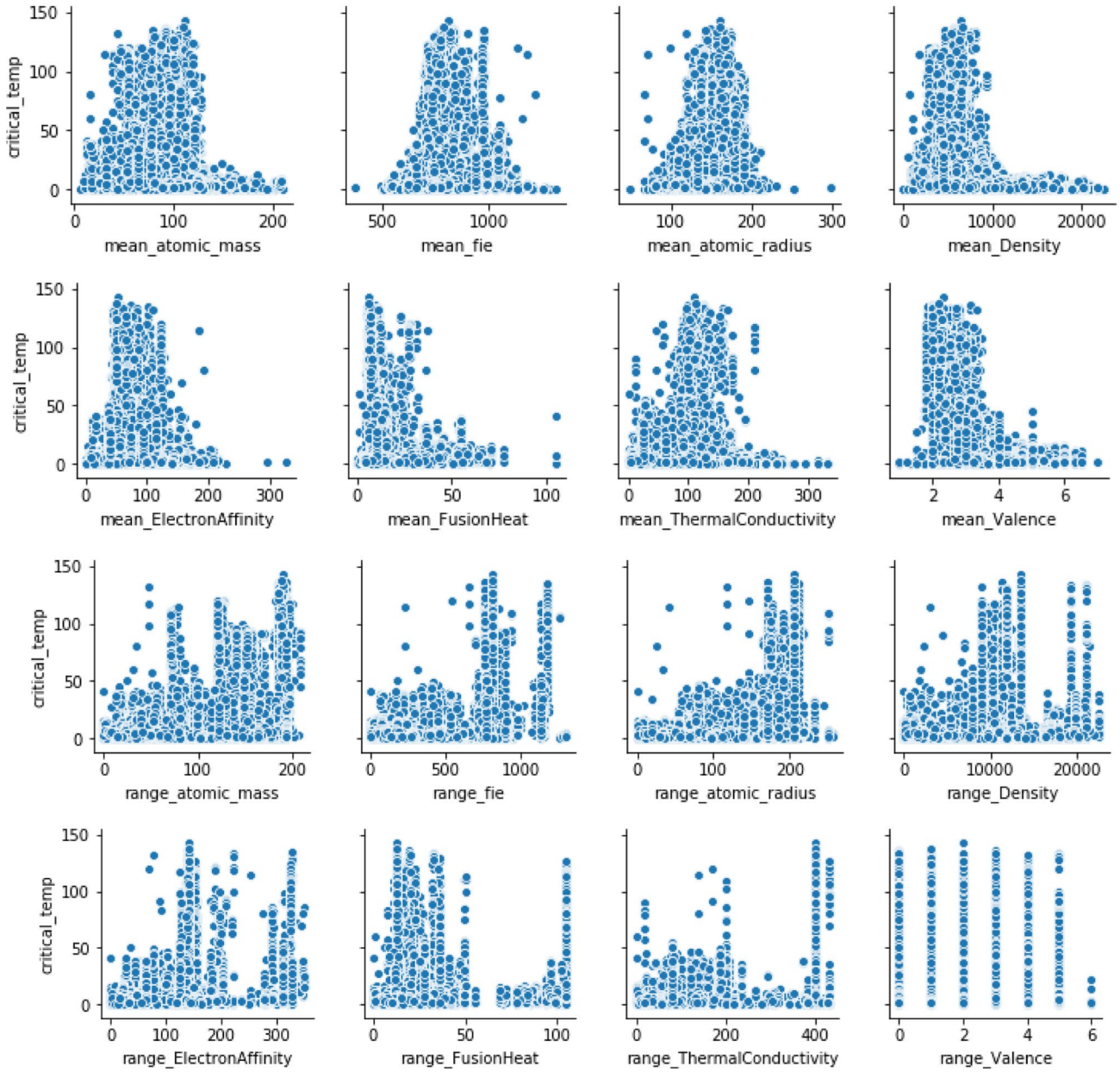

Fig. 2 The dependencies between the critical temperature and all "mean", "range" features

"weighted mean" features. Thus, we apply "heat map" method to display the correlation coefficient between "weighted mean" features and critical temperature (Fig. 4). As the correlation coefficient, we take the Pearson criterion, which is a method of parametric statistics that allows to determine the presence or absence of a linear relationship between two quantitative indicators. This coefficient r_{xy} is closer to 1 or − 1, the stronger the linear relationship between the two features [22]:

$$r_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

The construction of such heat maps can lead not only to the determination of the relationship between the critical temperature and features in the data, but also to the determination of the relationships between all parameters of the superconducting material, which is important for

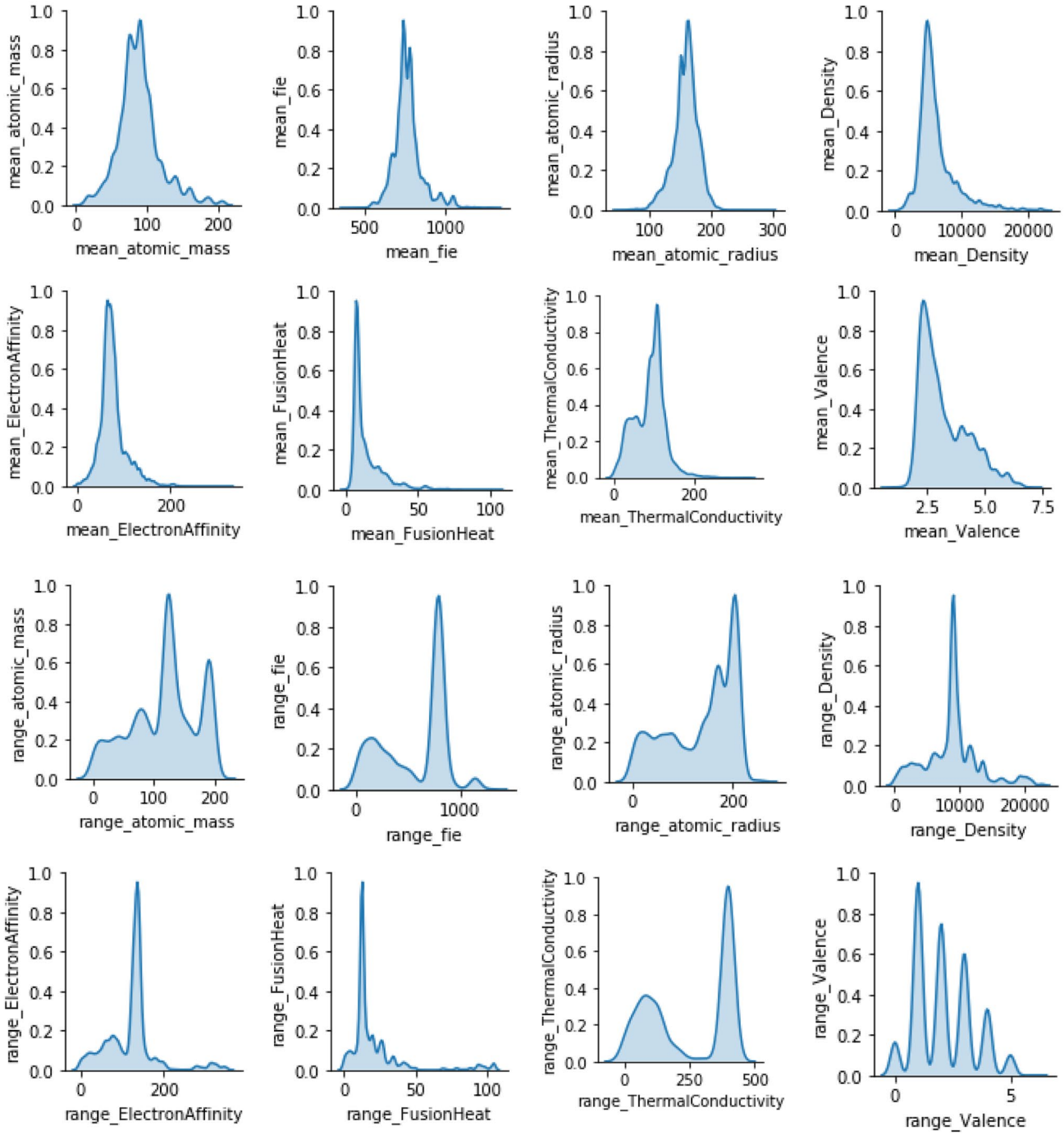

Fig. 3 Distribution of "mean", "range" features

constructing a complete theory of superconductivity. The display of these heat maps solves the problems of simultaneous pairwise graphic mapping of features and is a simpler way to assess the presence of dependence.

Due to the impossibility of displaying all the correlations between each other, we present a table for the calculated correlation coefficient for the critical temperature and all the features in the data. Based on the obtained values of the correlation coefficient, it is possible to establish how close the linear connection the critical temperature has with the features in the data. Table 4 presents the features and the correlation coefficient, which have a noticeable or strong tightness of connection.

Table 3 The most frequent values of "mean", "range" features

"mean" features	Region of the most frequent values	"range" features	Values of peaks
Atomic mass	60–70	Atomic mass	110, 175, 60
Ionization energy	750–800	Ionization energy	750, 250
Atomic radius	150–170	Atomic radius	200, 160
Density	15,000–18,000	Density	8700
Electron affinity energy	70–90	Electron affinity energy	120
Heat of fusion	6–8	Heat of fusion	10
Thermal conductivity	90–120	Thermal conductivity	300, 75
Valence	2.5	Valence	–

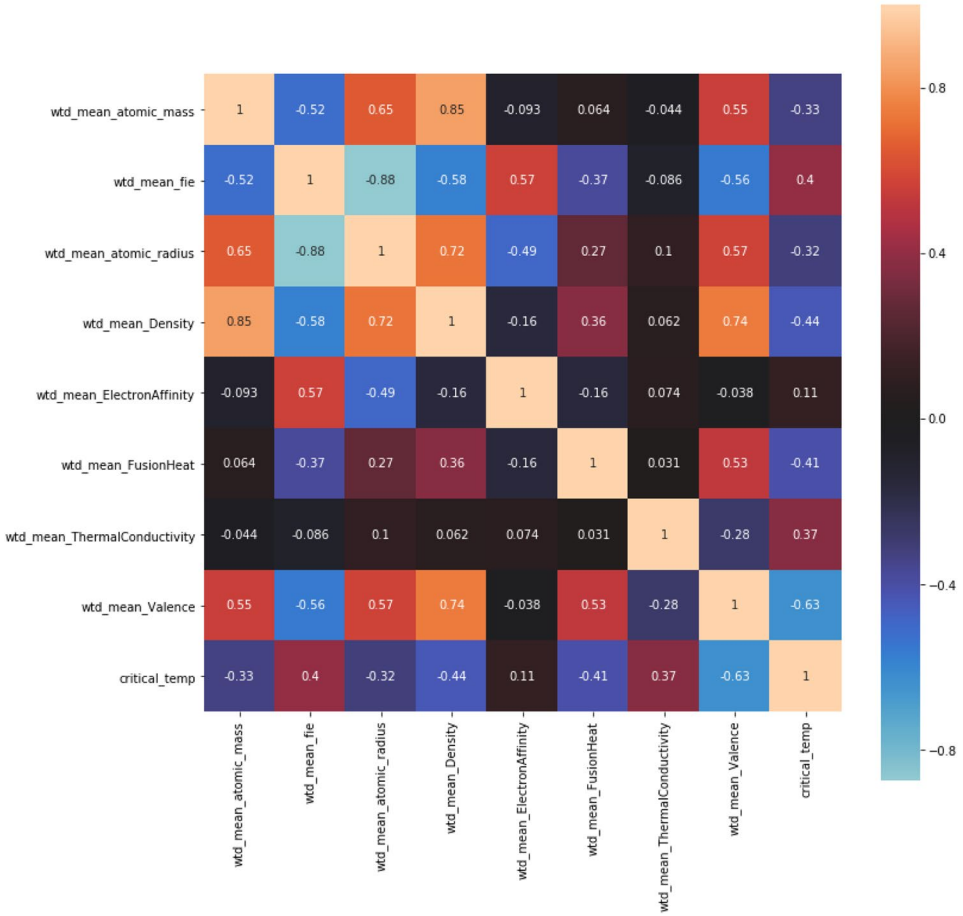

Fig. 4 Heat map of "weighted mean" features

2.5 Visualisation features with colors

The idea of the method is to represent the values of a particular attribute with color. The application of this method to the database of superconducting materials can be represented as follows:

1. The source data is sorted so that the materials are arranged as the critical temperature increases.
2. Each of the feature is normalized to a range of values [− 1,1].
3. For each feature, we set its values in accordance with the color. In this work, the maximum value of the attribute corresponds to yellow, the minimum blue, medium, respectively, green.
4. It turns out that all the data is sorted in the direction of increasing critical temperature and, displaying the values of the feature in the form of color for each mate-

Table 4 The value of the correlation coefficient for the critical temperature and all features in the data that have a noticeable or strong correlation with the critical temperature

Feature	Correlation coefficient	Feature	Correlation coefficient
wtd_std_ThermalConductivity	0.718081	wtd_mean_Valence	−0.635837
range_ThermalConductivity	0.686366	wtd_gmean_Valence	−0.619912
range_atomic_radius	0.655599	mean_Valence	−0.600995
std_ThermalConductivity	0.646382	gmean_Valence	−0.576322
wtd_entropy_atomic_mass	0.630703	gmean_Density	−0.546297
number_of_elements	0.628652	wtd_gmean_Density	−0.544344
entropy_Valence	0.625061		
range_fie	0.610423		
wtd_entropy_atomic_radius	0.609974		
entropy_fie	0.599203		
wtd_std_atomic_radius	0.598388		
wtd_entropy_Valence	0.597178		
entropy_atomic_radius	0.591918		
wtd_std_fie	0.589628		
entropy_FusionHeat	0.574341		
entropy_atomic_mass	0.569139		
wtd_entropy_FusionHeat	0.568757		
std_atomic_radius	0.559283		
std_fie	0.549440		

rial, the dependence of the feature on the critical temperature is traced.

Consider the example of the feature "number of chemical elements". The data obtained after sorting by critical temperature are presented in Fig. 5a.

Each line of Fig. 5a represents a material, more precisely since we visualize only one feature, the value of the critical temperature of the material.

The feature "number of chemical elements" after sorting the data by critical temperature is shown in Fig. 5b.

The number of chemical elements does not change explicitly with increasing critical temperature, but visually can see that the high values of the critical temperature often correspond to light green values of the number of chemical elements that take values greater than 5.

Next, we sort the features in such a way that we put together features with the same physical meaning, that is, for example, all features that have the meaning of atomic mass, etc. Get 8 of these sets of features. We divide the graph field into 10 columns and display these feature sets by analogy with the number of chemical elements (Fig. 6). The order of attributes for each set is as follows: average, weighted average, geometric mean, geometric weighted average, entropy, entropy weighted, difference, weighted difference, standard deviation, and weighted standard deviation.

The method of visualization features with colors, the algorithm of this method is first described in this paper.

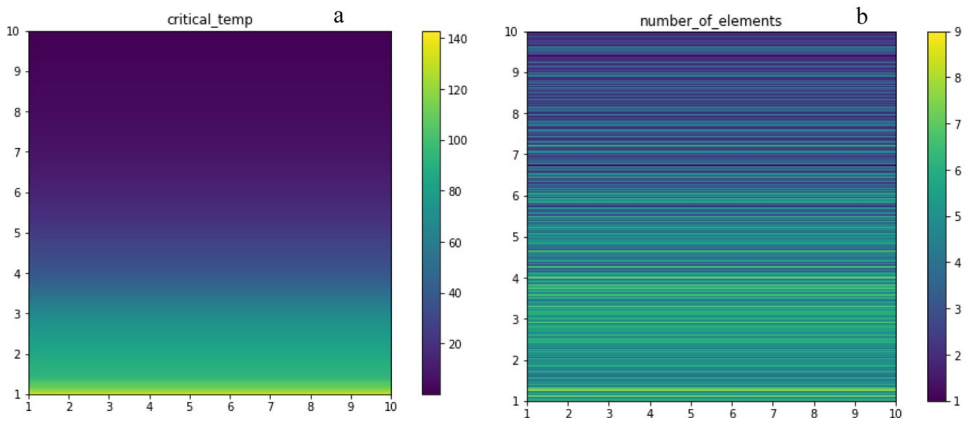

Fig. 5 Presentation of data on the basis of feature "critical temperature" (**a**) and "number of chemical elements" (**b**) after sorting

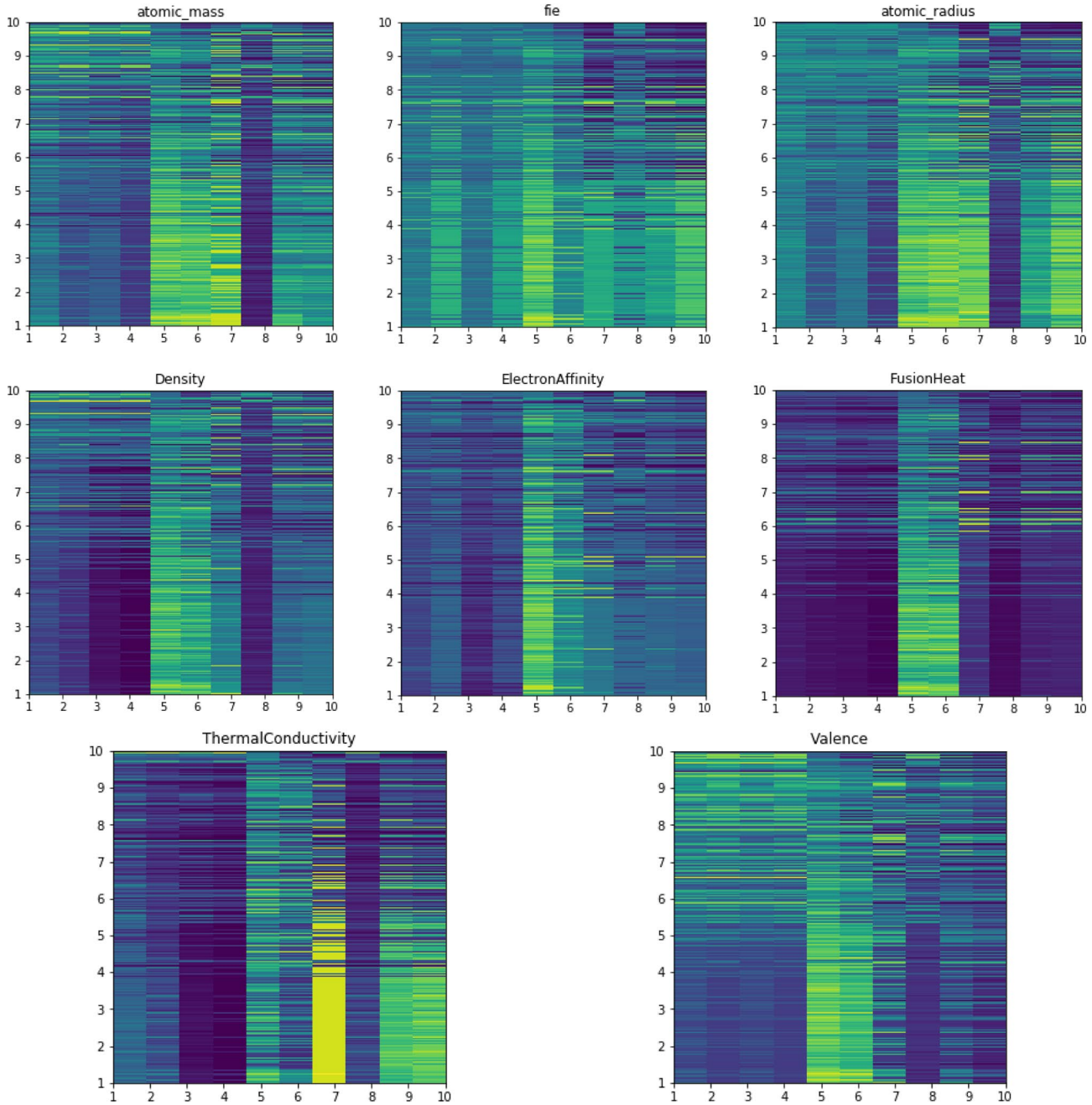

Fig. 6 Presentation of data on all features except the number of chemical elements

This method allows to simultaneously monitor the dependence of the target variable on many attributes in the data, when the direct construction of graphical dependencies is not possible due to the presence of a large number of features.

3 Discussion

Note that instead of the proposed algorithm for data preparation, we can use the built-in libraries for data

processing (for example, the method "drop_duplicates()" from Pandas in Python [23, 24]). But using this method, it is impossible to trace exactly which duplicates we remove, in contrast to the algorithm described above. As a result of data preprocessing, three new databases with unique critical temperatures were obtained. The uniqueness of critical temperature values is important for the further construction of critical temperature prediction models. Since the presence of different values for the same material can lead to the creation of models that do not have physical meaning and worse estimates of the quality of the model.

The feature "number of chemical elements" has a value from 1 to 9. Visualization of dimension reduction methods was carried out after removal of emissions and it can be seen that data with the number of chemical elements equal to 1 is very noisy, which means that it is not recommended to use them in the data to predict new materials. Apparently, the presence of emissions among materials with $n = 1$ is due to the fact that in addition to conventional superconducting metals, the data also contained various modifications of carbon, superconducting films under pressure, and other materials that do not exhibit superconductivity under normal conditions. Therefore, the emission removal algorithm deleted these materials, since the parameter values in this case can go beyond three standard deviations. In order for these materials not to be recognized as emissions, such parameters as the crystal structure, pressure, and sample size should be added to the data. But there are only 58 such materials in the data, which makes 0.37% of the total amount of materials before the removal of emissions. We can assume that these materials do not affect the results of visualization and will not affect further research.

From the point of view of physical theories, it remains unclear whether there is a continuous connection between low-temperature and high-temperature superconductors, since the mechanism of formation of superconducting particles [2, 25] should be different. As can be seen from Fig. 1, the methods of random projectionsnd multidimensional scaling show that the data is not divided into clusters. PCA and t-SNE methods show some separation. In PCA visualization, there is a "jumper" between two big clusters, but still this "jumper" is continuous and there is no strict separation into classes. The t-SNE method is the most obvious method in dividing data into clusters, as it seeks to reduce the distance between similar objects and increase between different ones. If there were different classes in the data, the result of lowering the dimension by this method would look much more discharged. So, the existing small clusters can be explained by the presence in these classes of superconducting materials, such as cuprate, iron, superconductors with heavy fermions and all others. Understanding the presence of classes among all superconducting materials is important for creating a complete theory of superconductivity.

Based on the results of pairwise mapping of the dependences of the critical temperature on the "mean" and "range" features, it can be concluded that it is difficult to establish an explicit dependence. But the obtained graphical dependences can be used to estimate the interval of critical temperature values of new possible superconducting materials. The obtained most frequent values of the "mean" and "range" features for superconducting materials can also be used for further synthesis of new materials. Certainly, if the values of the feature are outside the obtained range, this does not mean the absence of the possibility of superconducting properties of the material, but statistically the probability will be less.

Based on the constructed heat map, it can be concluded that the critical temperature is most strongly associated with the weighted mean valency of superconducting materials. Note that a noticeable correlation is shown by pairs of "weighting mean" features: density—atomic mass, density—atomic radius, atomic radius—ionization energy, valency—density, which is a physically trivial moment. This may indicate more general mechanisms of the formation of a superconducting state than is commonly accepted [1]. Based on the obtained values of the correlation coefficient between the critical temperature and all the features in the data, it is possible to determine the most important material parameters that affect the critical temperature. As a result, thermal conductivity has the largest positive value of the correlation coefficient, valency has the largest negative value, and these parameters can be considered the most important, on which the critical temperature depends.

Then, as a first approximation, we can assume that the more particles capable of transferring heat in a solid, the higher the critical temperature. And if we assume that these same particles carry a superconducting current, then the critical temperature T_C should be directly proportional to the concentration of superconducting particles n_S. The concentration of superconducting particles is associated with one of the characteristic lengths that describe the superconducting state—the London penetration depth λ. Then it turns out that: $T_C \sim n_S \sim 1/\lambda^2$. Articles [26, 27] discuss the mechanism of formation and destruction of a superconducting state, which relates the critical temperature to the London penetration depth and the coherence length. The expression for the critical temperature in these works was tested on many superconducting materials. It is shown that the dependence of the critical temperature on the coherence length is much weaker than on the London penetration depth and it is possible to show: $T_C \sim 1/\lambda^{1.5}$. Considering that the correlation coefficient is

not equal to unity for thermal conductivity, we can assume that the result of this work coincides with the theoretical result of works [26, 27] and is additional evidence in favor of this mechanism. In the same way, the results of this paper can be used to confirm and create a theory of superconductivity.

A negative correlation coefficient for valency indicates that with increasing valency, the critical temperature decreases. This fact can be compared with experimental data on the change in the critical temperature as a function of the oxygen content in cuprate superconductors [28]. In these materials, due to oxygen vacancies, the average valence of copper changes. Note that the data of the critical temperature dependence, in addition to the negative correlation, also have a positive correlation with the peak of the critical temperature. Thus, the result of this work represents a more general physical meaning than the dependence of the critical temperature on the oxygen content.

In this paper, for the first time, a method for visualizing features with colors and its algorithm are presented. This method is a general data presentation method with a large number of features. Based on the visualisation features with colors, we can conclude that the critical temperature increases with the number of chemical elements, geometric, entropy, weighted entropy value of atomic mass, entropy value, difference in ionization energy, standard and weighted standard value of standard deviation of ionization energy. Features related to atomic radius and density, to a first approximation, repeat the distribution of atomic mass. The critical temperature increases with increasing entropy value of electron affinity, entropy, weighted entropy value of fusion heat, range in thermal conductivity, standard and weighted standard values of standard deviation of heat conductivity, entropy, weighted entropy value of valency. With an increase in the remaining features, the critical temperature either decreases or is clearly independent.

4 Conclusion

For the first time, the methods of visualization of superconducting materials were applied in the framework of considering the problem of the transition of low-temperature superconductors to high-temperature ones, a general analysis of the dependences and trends of the critical temperature on material parameters. What is extremely important for controlling the properties of superconducting materials and the synthesis of new superconducting materials.

Based on the results of dimensional reduction methods, it can be concluded that superconducting materials do not have a division into classes. Thus, there must be a general theory of superconductivity for all known classes of superconductors, and the results of dimensionality reduction methods obtained can be applied to create this theory.

Based on the results of pairwise mapping of features, the dependences of the critical temperature on "mean" and "range" features are visualized, distributions of "mean" and "range" features, and the most frequent values of these features are obtained. These results can be used to estimate the manifestation of superconducting properties of the synthesized material, the critical temperature of new materials.

A heat map is constructed with the calculated Pearson correlation criterion for the weighted mean features, which is important for creating a general theory of superconductivity and establishing the dependences of the critical temperature on material parameters, dependencies of material parameters with each other. The Pearson correlation criterion is calculated for the critical temperature and all the features in the data. Features that show a strong or noticeable correlation with the critical temperature are selected, which determines the most important features that have the greatest influence on the critical temperature. A limitation of the results of this method is that the Pearson correlation criterion is able to establish only linear relationships.

A new method for visualization of features depending on the target variable has been developed. The method of visualization of features with colors allows to consider the dependence of the critical temperature on all features in the data. The trends in the dependences of the critical temperature on the material parameters are established, which is important in the synthesis of higher-temperature superconductors. The developed visualization method can be used to determine the trends of any data, where a direct representation of the dependences of the target variable on features in the data is impossible due to the large number of these features.

The results of this work are limited by the set of data used. To achieve more accurate results, more data on the experimental values of the critical temperatures are required. Also, consideration of applying visualization techniques to other material parameters may provide a different end result.

A further research direction, in addition to the possibility of creating a general theory of superconductivity, synthesis of new superconducting materials, and controlling their properties, may be creating a model for predicting the critical temperature based on machine learning methods.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Code availability The code required to reproduce these findings are available to download from [https://github.com/matasovav/DATA_SC].

Availability of data and material The initial data required to reproduce these findings are available to download from [https://archive.ics.uci.edu/ml/datasets/Superconductivty%20Data], [5]. The processed data required to reproduce these findings are available to download from [https://github.com/matasovav/DATA_SC].

References

1. Cooper LN (1956) Bound electron pairs in a degenerate Fermi gas. Phys Rev 104:1189
2. Bardeen J, Cooper LN, Schrieffer JR (1957) Theory of superconductivity. Phys Rev 108:1175
3. Che Y, Zhang L, Wang J, Chen Q (2017) Impurity effects on BCS-BEC crossover in ultracold atomic Fermi gases. Phys Rev B 95:014504
4. Palestini F, Strinati GC (2013) Systematic investigation of the effects of disorder at the lowest order throughout the BCS-BEC crossover. Phys Rev B 88:174504
5. Lakhno V (2019) Superconducting properties of a nonideal bipolaron gas. Phys C 561:1–8
6. Guo H, Peterka T, Glatz A (2017) In situ magnetic flux vortex visualization in time-dependent Ginzburg-Landau superconductor simulations. In: 2017 IEEE Pacific visualization symposium (PacificVis), Seoul, pp 71–80
7. Wang T, Jiang CY, Bilheux HZ et al (2019) Improving polarized neutron imaging for visualization of the Meissner effect in superconductors. Rev Sci Instrum 90(3):033705
8. Indenbom MV, Uspenskaya LS, Beek VCJ, Konczykowski M (2002) Visualization of magnetic structures in superconductors and magnetic materials using magneto-optical garnet films. Trans Magn Soc Japan 2(4):244–247
9. Ruan W, Li X, Hu C et al (2018) Visualization of the periodic modulation of Cooper pairing in a cuprate superconductor. Nat Phys 14:1178–1182
10. Kim J, Chua V, Fiete G et al (2012) Visualization of geometric influences on proximity effects in heterogeneous superconductor thin films. Nat Phys 8:464–469
11. Choubey P, Berlijn T, Kreisel A, Cao C, Hirschfeld PJ (2014) Visualization of atomic-scale phenomena in superconductors: application to FeSe. Phys Rev B 90(13):134520
12. Matthias BT (1955) Empirical relation between superconductivity and the number of electrons per atom. Phys Rev 97:74–76
13. Owolabi TO, Akande KO, Olatunji SO (2015) Estimation of superconducting transition temperature T C for superconductors of the doped MgB2 system from the crystal lattice parameters using support vector regression. J Supercond Nov Magn 28:75–81
14. Owolabi TO, Akande KO, Olatunji SO (2014) Prediction of superconducting transition temperatures for Fe-based superconductors using support vector machine. Adv Phys Theor Appl 35:12–26
15. Stanev V, Oses C, Kusne AG et al (2018) Machine learning modeling of superconducting critical temperature. NPJ Comput Mater 4:29
16. Hamidieh Kam (2018) A data-driven statistical model for predicting the critical temperature of a superconductor. Comput Mater Sci 154:346–354
17. Bingham E, Mannila H (2001) Random projection in dimensionality reduction: applications to image and text data. In: Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining (KDD'01). ACM, New York, NY, USA, pp 245–250
18. Tipping ME, Bishop CM (1999) Probabilistic principal component analysis. J Roy Stat Soc Ser B (Stat Methodol) 61(3):611–622
19. Borg I, Groenen P (1997) Modern multidimensional scaling—theory and applications. Springer Series in Statistics, Berlin
20. van der Maaten LJP (2014) Accelerating t-SNE using tree-based algorithms. J Mach Learn Res 15:3221–3245
21. Michael W et al (2016) Seaborn: v0.5.0. Zenodo
22. Liu XS (2019) A probabilistic explanation of Pearson's correlation. Teach Stat 41(3):115–117
23. McKinney (2010) Proceedings of the 9th Python in science conference, vol 445
24. McKinney (2011) Python High Perform Sci Comput, vol 14
25. Romanov EP, Blinova YV, Sudareva SV et al (2006) Mechanism of formation, fine structure, and superconducting properties of high-temperature superconductors and superconducting composites. Phys Met Metallogr 101:27–44
26. Matasov AV (2019) Characteristic lengths and Plasmon superconductivity mechanism of some high-temperature superconductors. In: 2019 international youth conference on radio electronics, electrical and power engineering (REEPE), Moscow, Russia, pp 1–4
27. Matasov AV (2019) AIP conference proceedings 2163, pp 020005-1–020005-5
28. Tallon JL (2005) Oxygen in high-Tc cuprate superconductors. In: Narlikar AV (ed) Frontiers in superconducting materials. Springer, Berlin