ISSN No. 0976-5697

Volume 7, No. 3, May-June 2016

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Material Selection using Association Rule Mining

Vandit Hedau
School of Data Sciences and Forecasting
DAVV
Indore, India

Pushpesh Pant
School of Data Sciences and Forecasting
DAVV
Indore, India

Kuldeep Sharma
School of Data Sciences and Forecasting
DAVV
Indore, India

Abstract: Data mining is multi-disciplinary field of study in its broader area of research. From a research point of view, it is not often to implement data mining concept in material science. This paper attempts to investigate the data mining techniques for material selection which are beneficial for design, manufacturing and other application of industrial engineering. Predictive data mining technique was used to model a knowledge discovery system for the selection of materials that must satisfy the design specifications. Well known predictive method Association Rule Mining with the help of WEKA software were used for material selection for specific application. The knowledge discovery from the engineering materials, datasets is proposed for effective decision making in advance engineering materials design application.The main aim of this article is thus to introduce a data mining techniques for material selection and also highlighting its value to the modern material scientist and researcher.

Keywords:Association Rule Mining, Engineering Materials, Support, Confidence, WEKA software

INTRODUCTION

Data Mining is an interdisciplinary field of study that develops methods to extract or identify meaningful information or pattern from the data. It has already expanded its uses to the scientific areas such as statistics, graph theory, linear algebra, databases, mathematics, and computer science. With the development of information era, the need of these techniques has increased drastically in information related applications including commerce, finance, etc. Data mining has also become an essential tool in the area of genomics whose primary techniques involves an orderly sifting through millions of genes to discover similarities or pattern among them [6].

Recently material scientists have started to inquire data mining ideas for the selection of steel alloy for different types of applications. The following section gives a brief overview of a technique used in data mining. Whenever possible, an attempt is made to give examples from materials where the techniques can be applicable.

Among the many problems that are solved by data mining, two are of most importance. The first is 'unsupervised clustering' which is the task of finding subsets of the data such that the items from the same subset are most similar and items from the distinct subsets are more dissimilar. The second is (predictive modeling, supervised learning) where predictive mining tasks perform induction on the current data in order to make predictions [7].

In order to perform these tasks, it is necessary to have dataset containing useful information which can be helpful while predicting materials and their applications. In material informatics, data mining can be used in the following tasks [1].

(1) ***Association Analysis****:* Association rule mining is good at finding patterns, and can be used to develop heuristic rules for materials behavior based on large data sets.

(2) ***Predict Modeling****:* Some machine learning algorithms can be used for material class prediction, and materials classification models such as Neural Network (NN) can be built up predict models.

(3) ***Cluster Analysis****:* As an exploratory data analysis tool, it has the task of finding subsets of the data such that the items from the same subset are most similar and items from the distinct subsets are more dissimilar.

(4) ***Outlier Analysis****:* In properties analysis or combinatorial, outlier analysis is used to identify peculiarities, especially to compute the uncertainty and accuracy of results, and

Distinguish between true discoveries and false positive result.

(5) ***Material Visualization****:* Reconstruction of material structure information based on materials data would help researchers to analyze the relationships between material structure and material properties.

SCOPE OF DATA MINING IN MATERIAL SCIENCE

Data Mining is now a valuable tool in the broad area of material selection and applications [8], [9]; there are good reasons why this area is particularly rich in engineering materials[10], [11], [12]. There is a huge range of materials and it is a difficult task to physically model the relationships between constituent, processing and properties. Therefore materials are primarily selected for a particular application by quantitative and trial-and-error method, where

researchers are guided by experience and heuristic rules for materials selection. These rules are applicable to somewhat limited materials, data sets. This is basically human data mining, where one's brain, instead of a computer, is being used to find correlations, make predictions and selections. Transferring data mining tasks to computer offers the way to enhance accuracy.

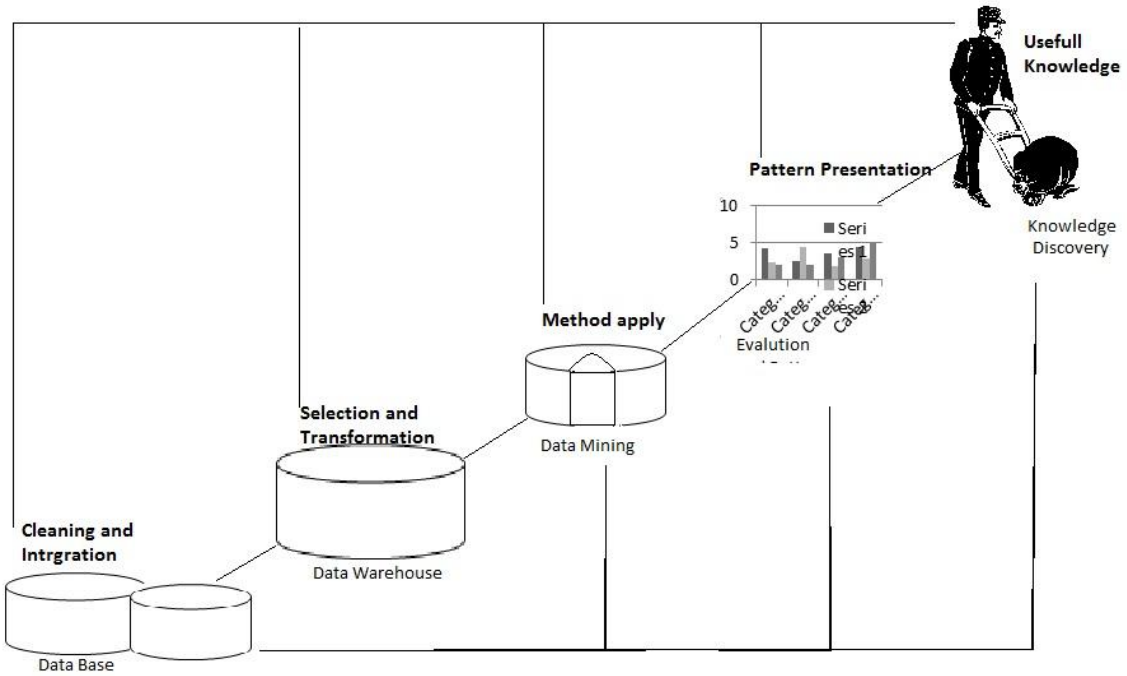

FIGURE.1: Knowledge Discovery Process [2]

The research areas of material science involving data mining are mainly focused on [3]:

(1) ***Data Standards:*** There are thousands of materials databases in different format and they are difficult to interpret with each other. So it is very important to integrate all the data into a single, or coherent database, data pre-processing is the first important task [2,3] to enable knowledge discovery.

(2) ***Management of Material Data:*** We have a huge amount of material data, but only the part of it is useful for materials researcher. To satisfy the need of research and production, and to construct the materials, data sets into single coherent database, efficient Materials Database Management Systems are very necessary [14].

(3) ***Data Mining on Materials Data:*** Data mining has the abilities for selecting, exploring, and modeling large amounts of data to unveil previously unknown patterns from large materials databases. Data mining involves some computational algorithms [14], [15], [16] such as neural network, genetic algorithm etc.

MATERIALS DATABASE

In the recent era, science and technology is advancing rapidly. Accordingly, material science and engineering change with each passing day, and new ideas, methods, techniques and materials appear one after another. This leads to the huge amount of materials database which cannot be tackled manually, so data mining is becoming one of the general areas of material selection and development [4] [5].

TableI. Steel Alloys with their Mechanical Properties

	Steel Alloy	Mechanical Properties
1.	Ferrium M54	High Strength, Toughness, Corrosion Resistant
2.	Crucible Steel	Hardness, Toughness, Abrasion Resistant
3.	Hadfield Steel	Toughness, Abrasion Resistant, High Impact Strength
4.	High Speed Steel	Toughness, Abrasion Resistant, Hot Hardness, High Bending Strength, High Thermal Conductivity
5.	Maraging Steel	High Strength, Toughness, Corrosion Resistant
6.	Reynolds 531	Toughness, Light Weight, High Strength
7.	Reynolds 525	Toughness, Light Weight, High Strength
8.	Reynolds 520	Toughness, Light Weight, High Strength
9.	Tool Steel	Hot Hardness, Toughness, Abrasion Resistant, High Thermal Conductivity
10.	Weathering Steel	High Strength, Corrosion Resistant

11.	Wootz Steel	Toughness, High Impact Hardness, Plastic Properties
12.	Al-6XN	Corrosion Resistant, Weldable, High Strength, Formability
13.	Celestrium	High Resistant, Corrosion Resistant
14.	Marine Grade Steel	Corrosion Resistant, Heat Resistant, Weldable, High Fatigue Strength
15.	Alloy 28	Weldable, Formability, Corrosion Resistant, Abrasion Resistant
16.	Surgical Stainless Steel	High Strength, Formability, Corrosion Resistant
17.	Zeron 100	High Strength, Corrosion Resistant, Pitting Resistant
18.	Ferrium S53	High Strength, Corrosion Resistant
19.	Ferrium C61	High Strength, Toughness, Hardness

In the above table, we just take the sample of 19 steel alloys to show how a material database with lack of alloys can be handled to choose a specific material for a particular application.

DATA MINING TECHNIQUE

Prediction is one of the core tasks in data mining. In the above materials database, we use *Association Rule Mining* to predict the material for specific applications.

A. ASSOCIATION RULE MINING:

It is an "If-Then" relationship. If it happens, what is most likely to happen next? It is a popular and well researched method for discovering interesting relation between variables in large databases. Association rule show attributes value conditions that occur frequently together in a given dataset [2].

Association Rule:

- $I = \{ i_1, i_2, i_3, - - - - i_0 \}$: A set of all the items
- Transaction, T: A set of items such that $T \subseteq I$
- Transaction database, D : A set of transactions
- A transaction $T \subseteq I$ contains a set $X \subseteq I$ of some items, if $X \subseteq T$
- An association rule is an implication of the form $X \rightarrow Y$, where $X,Y \subseteq I$

A set of items is referred as an item-set that contains k items i.e. k- item set

- Support is a rule to find the value of X with respect to T which is defined as the proportion of transaction in the database that contains the item-set X

Support = {(X U Y). count}/ (n)	(1)

- Confidence gives a value of the rule, $X \rightarrow Y$, with respect to a set of transaction that contains X which also contains Y.

Confidence = {(X U Y). count}/ (X. count)	(2)

Table II. Support and confidence for each pair of transaction

X → Y (Transaction)	**Support**	**Confidence**
Toughness → High Strength	42%	73%
High Strength → Toughness	42%	57%
Corrosion Resistant → High Strength	47%	90%
Light Weight → Toughness	16%	100%
Light Weight → High Strength	16%	100%
Hardness → Toughness	26%	100%

```
Apriori
=======

Minimum support: 0.89 (17 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 11

Generated sets of large itemsets:

Size of set of large itemsets L(1): 4

Size of set of large itemsets L(2): 1

Best rules found:

1. High Impact Strength=no 18 ==> Heat Resistant=no 17    conf:(0.94)
2. High Strength Bending=no 18 ==> Heat Resistant=no 17    conf:(0.94)
3. Plastic Properties=no 18 ==> Heat Resistant=no 17    conf:(0.94)
4. Pitting Resistant=no 18 ==> Heat Resistant=no 17    conf:(0.94)
5. High Impact Strength=no Plastic Properties=no 18 ==> Heat Resistant=no 17    conf:(0.94)
```

Figure.2**:** Material Selection Rule in WEKA

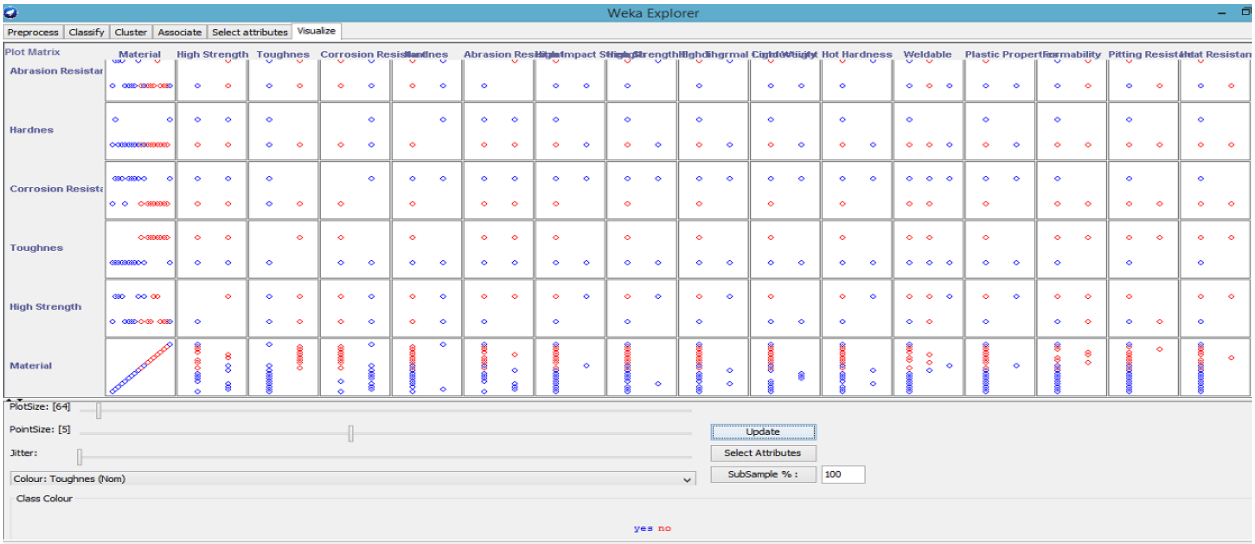

Figure.3**:** Materials Associated Properties visualization

The above figure shows the graphical representation between the mechanical properties where blue color indicates positive association and red color indicates negative association.

RESULTS AND DISCUSSION

In this research paper, association rule mining is applied to the material data set to reveal the need of particular material for specific applications. The best known constraints are minimum thresholds on support and confidence. The support

conf (X→Y) = supp (XUY)/supp(X)	(3)

(X) of an item set X is defined as the proportion of transactions in the data set which contain the item set.
The confidence of a rule is defined as:

This is the manual method of conducting association rule mining just to show how it actually works. In the above table, we calculated support for different transaction (transaction between mechanical properties) to find the value of X with respect to T which is defined as the proportion of transaction in the database that contains the item-set X and confidence gives a value of the rule, X→Y, with respect to a set of transaction that contains X which also contain Y.

Apriori is a method of association rule mining. In conducting apriori in WEKA, we get the following result which is shown in the figure below. This result is based on our material database (containing 19 steel alloy).

CONCLUSION

The primary aim of this paper was to show how a few simple data mining techniques can be applied to answer a few specific questions on materials. Generally it is very difficult to choose a particular material for specific application manually because of very large and complex database available to us. This paper focused on how data mining technique such as association rule would be helpful in predicting suitable material for different applications. By applying data mining to material database we have got the following results obtained,

- Collaboration of association rule mining with a material database can be helpful in selecting the right type of material for a particular application.
- It will minimize material selection time.
- It provides a way to switch from conventionally used material for a particular application with the new one.

REFERENCES

[1] Doreswamy& Hemant K S, "Hybrid data Mining Technique for Knowledge Discovery from Engineering Materials Data Sets,"

[2] Jiawei Han, Micheline Kamber& Jian Pei.: "Data Mining: Concepts and Technique," (2012)

[3] Krishna and Rajan.: "Materials informatics," *Materials Today*. (2005)

[4] William D. Callister.:"Materials science and engineering: an introduction," (2007)

[5] Donald R. Askeland, Pradeep P. Fulay&Wendelin J. Wright.: "The Science and Engineering of Materials," (2012)

[6] Florin Gorunescu.: "Data Mining- Concepts, Models and Techniques" (2011)

[7] Ian H. witten, Eibe Frank & Mark A. Hall.: "Data Mining-Practical Machine Learning Tools and Techniques" (2011)

[8] Doreswamy, Sharma, S, C .: "An Expert Decision Support System for Engineering Materials Selections And Their Performance Classifications on Design Parameters", International Journal of Computing and Applications (ICJA) (2006), Vol.1, No.1,pp.17-34.

[9] Langseth, H. & Nielsen, T.: "Classification using Hierarchical Naïve Bayes models", Machine Learning, Vol. 63, No. 2, pp. 135-159. (2006)

[10] Ferris,F., Peurrung,L.M., Marder,J .: "Materials informatics: Fast track to new materials," ADVANCED MATERIALS & PROCESSES, Vol. 165, pp. 50-51. (2007)

[11] Gang Yu.,Jingzhong Chen., Li Zhu.: Data mining techniques for materials informatics: Datasets Preparing and Applications", Proc. 2009 Second International Symposium on Knowledge Acquisition and Modeling, pp.181-189. (2009) "

[12] Rodgers, J.R., Cebon, D.: "Materials informatics," MRS BULLETIN, Vol. 31, pp. 975- 980. (2006)

[13] Wei,Q.Y., Peng,X. D., Liu, X.G., Xie,W.D .: , "Materials informatics and study on its further development," CHINESE SCIENCE BULLETIN, Vol. 51, pp 498-504. (2006)

[14] Rajan .: "Informatics and Integrated Computational Materials Engineering: Part II," JOM, Vol. 61, pp. 47-47. (2009)

[15] Rodgers,J.R .: "Materials informatics: Knowledge acquisition for materials design.," ABSTRACTS OF PAPERS OF THE AMERICAN CHEMICAL SOCIETY, Vol. 226, pp. 302-303. (2003)

[16] Rajan,K.: "Combinatorial materials sciences: Experimental strategies for accelerated knowledge discovery," ANNUAL REVIEW OF MATERIALS RESEARCH, Vol. 38, pp. 299-322.(2008)