Apriori Algorithm and Game-of-Life for Predictive Analysis in Materials Science

Aparna S. Varde, Makiko Takahashi, Elke A. Rundensteiner, Matthew O. Ward, Mohammed Maniruzzaman and Richard D. Sisson Jr.

Worcester Polytechnic Institute (WPI), Worcester, MA 01609, USA

Corresponding Author: Aparna S. Varde

E-mail: aparna@wpi.edu Phone: (508)-831-5857 Fax: (508)-831-5776

Abstract

Experimental data in many domains serves as a basis for predicting useful trends. If the data and analysis are available over the Web this promotes E-Business by connecting clientele worldwide. This paper describes such a predictive tool "QuenchMiner™" in the domain "Materials Science". Data mining, more specifically the "Apriori Algorithm", is used to derive association rules that represent relationships between input conditions and results of domain experiments. This enables the tool to answer questions such as "Given cooling medium and agitation during material heat treatment, predict cooling rate". This allows users to perform case studies on the Web and use their results to optimize the involved processes, thus increasing customer satisfaction. Another interesting aspect is predicting material microstructure during heat treatment. Microstructure controls material properties such as hardness. Hence its prediction helps in making decisions about materials selection. Microstructure prediction has similarities to an artificial intelligence process called "Game-of-Life". Some challenges in our work are incorporating domain expert judgement while mining association rules, simulating microstructure evolution under different conditions, and dealing with uncertainty. These challenges and associated research issues are outlined here. To the best of our knowledge, this is the first tool performing Web-based predictive analysis in Materials Science.

1. Introduction

Data in a domain is either fully deterministic and hence is predictable, or consists of random variables and is thus unpredictable, or is somewhere in between [8]. The data that we consider in this paper falls into the in between range. It comes from the domain of Materials Science, in particular the heat treatment of materials. Heat treatment refers to the controlled heating and cooling of materials to acquire desired mechanical and thermal properties [19]. These properties determine the use of the materials in specific applications. The data is primarily experimental and consists of input conditions in heat treatment and the corresponding observations in terms of parameters such as cooling rates and heat transfer coefficients. The heat transfer coefficients (hc) [26] represent the heat extraction capacity of the process and hence characterize the experiment. Domain users are interested in determining these parameters to make decisions about corresponding processes in industry. Selecting the most suitable materials and parameters for specific processes optimizes the results, thereby enhancing business.

Computational tools assist in making decisions by analyzing the data, and discovering useful patterns for predicting future trends. If the tool is Web-based, this provides worldwide access to domain users. In the Materials Science domain it is imperative to connect materials suppliers, automobile companies, heat treatment industries, universities, researchers, aerospace agencies, manufacturing companies and other users [5, 7]. Exchange of knowledge among these users enables them to make faster and more effective decisions. For example, prior knowledge of the fact that distortion is likely to occur in a part when it is heat treated under certain conditions is useful in selecting parameters so as to minimize distortion in an industrial heat treatment process. This in turn helps to optimize processes and make better products hence improving business by satisfying customers. Thus on the whole, E-Business is promoted by facilitating worldwide exchange of knowledge useful in the domain for supporting various aspects of decision support. This paper focuses on the techniques involved in building such a tool called QuenchMiner™ [32, 33] with the main goal being predictive analysis. It has been rightly said that, "the building of predictive tools is one of the basic subjects in science" [20]. There are two important aspects to prediction in Materials Science. One is estimating parameters of interest such as cooling rates and heat transfer coefficients [26] given the input conditions in a process. This supports parameter selection to optimize processes. The other is simulating the microstructure evolution [29] of a material during heat treatment. Since microstructure controls the mechanical properties of a material, this helps in materials selection to optimize products.

In order to assist decision making in Materials Science, it is useful to discover knowledge from raw data, i.e., to perform data mining [10] and build a Materials Knowledge Base. It is imperative to assimilate the knowledge of a domain expert in the mining process. The Apriori Algorithm [2] is used in data mining to perform Association Analysis, namely the discovery of rules of the type "A=>B" where A and B are items or conditions in the given data set [1, 2]. This is useful in developing a Materials Knowledge Base with association rules representing relationships between input conditions and experimental results. However, the rules discovered by

Apriori may not all be useful with respect to the domain. Hence it is essential to prune the rules guided by basic domain knowledge. Also, some interesting rules may not be found from experimental data, in our case, heat treatment experiments. Thus it is advisable to extend the Association Analysis to other sources such as the related literature in the domain, to enhance the Materials Knowledge Base. The paper addresses the potential research issues emerging from this.

Experimental data in heat treatment is used to plot graphs such as cooling curves [26] that serve as good visual tools to represent results. A material has different microstructures [29] at different regions on a cooling curve. In predicting these microstructures, an important aspect is the visualization of experimental data. Techniques provided by the packages such as the Xmdv tool developed at WPI [34] are useful in data visualization. Domain-specific aspects such as the superimposing of cooling curves over Jominy end quench results [23] are important. There are also rules pertaining to microstructure evolution, i.e., the various phases that a material could be in at a particular stage of heat treatment and the microstructure of that phase. An artificial intelligence process called Game-of-Life [9] simulates the birth and death of cells in a society and is useful for microstructure simulations. The main challenges in this task are predicting the actual evolution of microstructure at several regions of interest on a graph.

A significant issue in estimation is uncertainty. Resolving uncertainty [24, 35] is an important aspect of predictive analysis. Artificial intelligence techniques such as conflict resolution strategies [17, 28] are useful here. These are included in our work.

This paper describes a research effort in building the QuenchMiner™ tool with the following objectives.

- Domain-type-dependent data mining using Apriori over relational and text sources, for estimating experimental parameters.
- Data visualization guided by domain knowledge and Game-of-Life, for simulating microstructure evolution.
- Treatment of uncertainty in prediction by using artificial intelligence approaches such as conflict resolution.

The rest of this paper is organized as follows. Section 2 describes building a Materials Knowledge Base using Apriori. Section 3 outlines research issues in knowledge discovery. Section 4 introduces microstructure prediction. Section 5 describes simulating microstructure evolution with Game-of-Life. Section 6 explains dealing with uncertainty. Section 7 summarizes evaluation. Section 8 describes the application of the tool to E-Business. Section 9 gives conclusions.

2. Apriori Algorithm for the Development of the Materials Knowledge Base

In order to perform predictive analysis, it is useful to discover interesting patterns in the given data set that serve as the basis for estimating future trends. Association Analysis or Association Rule Mining [1] is helpful here. This refers to the discovery of attribute-value associations that occur frequently together within a given data set [11]. An association rule is defined as follows [1, 2].

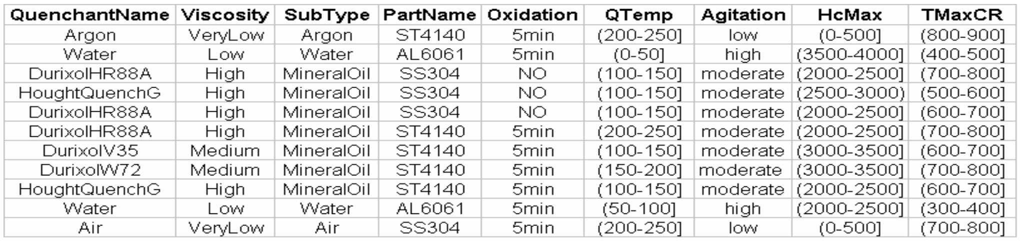
Definition of Association Rule: Let I = {i1, i2, …. im} be set of items, D be task relevant data of transactions, T be each transaction, a set of items, such that T ς I where ς denotes proper subset and TID be the Transaction Identifier. An Association Rule is defines as an implication of type A => B, where A ς I, B ς I and A ∩ B = Φ. The Rule holds in D with confidence C and support S, where C: Confidence (A=>B) = P (A U B), S: Support (A=>B) = P (B | A where P is probability.

2.1 The Apriori Algorithm

The Apriori Algorithm [2] proposed by Agrawal et. al. in 1994, finds frequent items in a given data set using the anti-monotone constraint [10, 25]. This algorithm embodies the following.

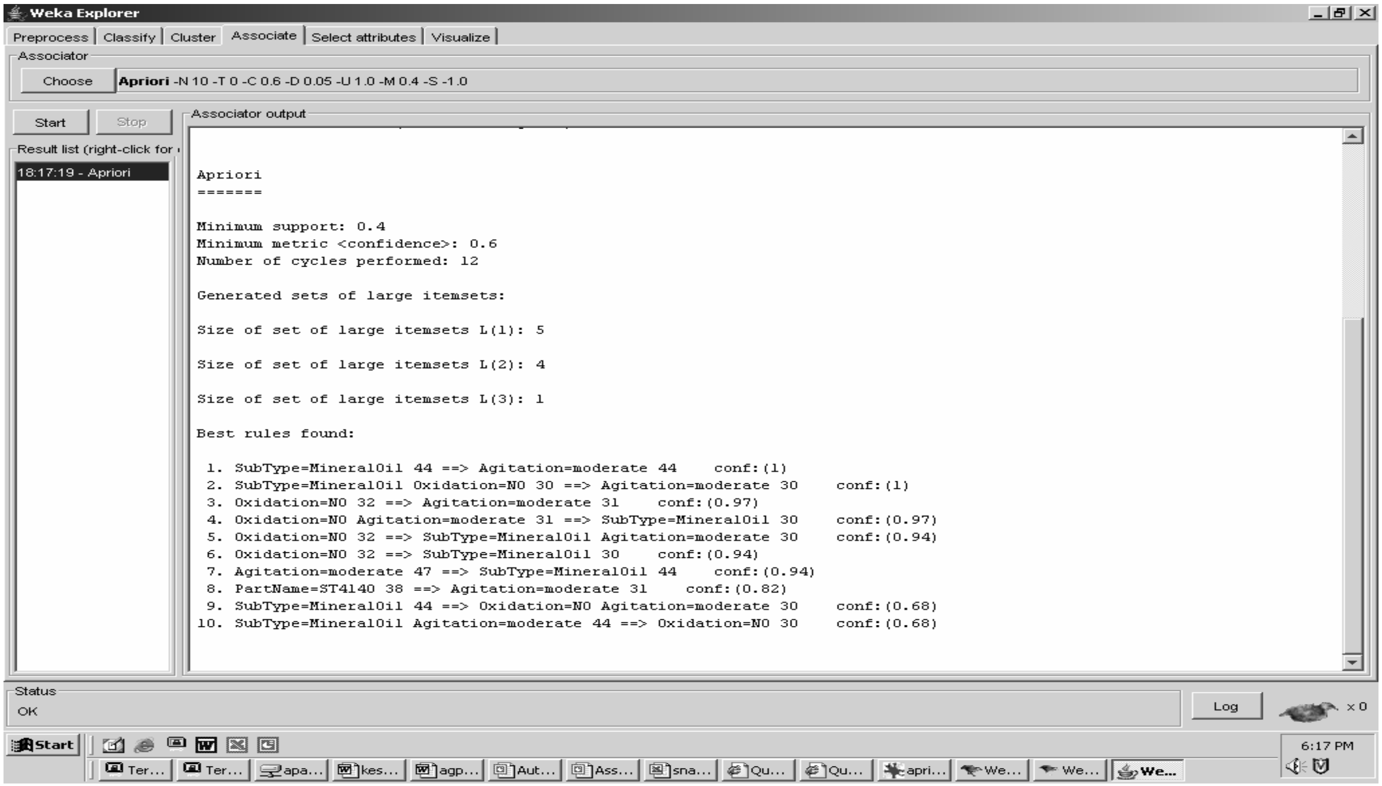
- Given a data set, the problem of association rule mining is to generate all rules that have support and confidence greater than a user-specified minimum support and minimum confidence respectively.
- Candidate sets having k items can be generated by joining large sets having k-1 items, and deleting those that contain a subset that is not large (where large refers to support above minimum support).
- Frequent sets of items with minimum support form the basis for deriving association rules with minimum confidence. For A=>B to hold with confidence C, C% of the transactions having A must also have B.

2.2 Apriori over Relational Experimental Data



QuenchantName	Viscosity	SubType	PartName	Oxidation	QTemp	Agitation	HcMax	TMaxCR
Argon	VeryLow	Argon	ST4140	5min	(200-250]	low	(0-500]	(800-900]
Water	Low	Water	AL6061	5min	(0-50]	high	(3500-4000]	(400-500]
DurixolHR88A	High	MineralOil	SS304	NO	(100-150]	moderate	(2000-2500]	(700-800]
HoughtQuenchG	High	MineralOil	SS304	NO	(100-150]	moderate	(2500-3000)	(500-600]
DurixolHR88A	High	MineralOil	SS304	NO	(100-150]	moderate	(2000-2500]	(600-700]
DurixolHR88A	High	MineralOil	ST4140	5min	(200-250]	moderate	(2000-2500]	(700-800]
DurixolV35	Medium	MineralOil	ST4140	5min	(100-150]	moderate	(3000-3500]	(600-700]
DurixolW72	Medium	MineralOil	ST4140	5min	(150-200]	moderate	(3000-3500]	(700-800]
HoughtQuenchG	High	MineralOil	ST4140	5min	(100-150]	moderate	(2000-2500]	(600-700]
Water	Low	Water	AL6061	5min	(50-100]	high	(2000-2500]	(300-400]
Air	VeryLow	Air	SS304	5min	(200-250]	low	(0-500]	(700-800]

Figure 1: Partial Snapshot of Experimental Data



Apriori -N 10 -T 0 -C 0.6 -D 0.05 -U 1.0 -M 0.4 -S -1.0

18:17:19 - Apriori

```
Apriori
=======

Minimum support: 0.4
Minimum metric <confidence>: 0.6
Number of cycles performed: 12

Generated sets of large itemsets:

Size of set of large itemsets L(1): 5

Size of set of large itemsets L(2): 4

Size of set of large itemsets L(3): 1

Best rules found:

 1. SubType=MineralOil 44 ==> Agitation=moderate 44    conf:(1)
 2. SubType=MineralOil Oxidation=NO 30 ==> Agitation=moderate 30    conf:(1)
 3. Oxidation=NO 32 ==> Agitation=moderate 31    conf:(0.97)
 4. Oxidation=NO Agitation=moderate 31 ==> SubType=MineralOil 30    conf:(0.97)
 5. Oxidation=NO 32 ==> SubType=MineralOil Agitation=moderate 30    conf:(0.94)
 6. Oxidation=NO 32 ==> SubType=MineralOil 30    conf:(0.94)
 7. Agitation=moderate 47 ==> SubType=MineralOil 44    conf:(0.94)
 8. PartName=ST4140 38 ==> Agitation=moderate 31    conf:(0.82)
 9. SubType=MineralOil 44 ==> Oxidation=NO Agitation=moderate 30    conf:(0.68)
10. SubType=MineralOil Agitation=moderate 44 ==> Oxidation=NO 30    conf:(0.68)
```

OK

Figure 2: Sample Association Rules from Experimental Data

Experimental data in the domain is integrated into a database to serve as the basis for analysis. In our context, the database is QuenchPAD™ [32], the Quenchant Performance Analysis Database, developed at the Center for Heat Treating Excellence, WPI. The Apriori algorithm is used for discovering rules from this experimental data in Materials Science to represent the knowledge of an expert. A partial snapshot of a data sample presented for Association Analysis and some rules derived are shown in the Figures 1 and 2 respectively. In Figure 2, the numbers on the left and right hand sides of the rules indicate the support for those items respectively.

2.3 Role of Basic Domain Knowledge

Using the statistical measures of interestingness, i.e., confidence and support, some of the rules derived by Apriori are obvious, e.g. "Oxidation=No => Agitation=Moderate". Since the part used to perform a heat treating experiment often has no oxide formation and the default level of agitation during the rapid cooling step in heat treatment is "moderate" [26, 32], this rule has high support and confidence. However as per the opinion of the domain experts, this rule does not represent knowledge useful for decision support. It only represents obvious information. A potential solution to the problem of obvious rules may be attribute selection during mining [11], i.e., in this case removing the attributes "Agitation" and "Oxidation". However this is not feasible since some rules involving these attributes along with others may be interesting in the domain, e.g., in this case, "Oxidation = No AND Agitation = Moderate => Subtype = Mineral Oil". It is important for the domain users to know that when a part does not have oxide formation and when the agitation is moderate, the cooling medium used is most likely to be a mineral oil. This knowledge helps in cooling medium selection in heat treatment.

Thus there is some intuition coming from the fundamental knowledge of the domain that is not captured by purely statistical measures of interestingness. Altering the levels of confidence and support, and using other measures such as lift and conviction [11] has not helped solve this problem. Hence it is proposed that in our tool, the rules derived by Apriori are pruned by using basic domain knowledge. This is summarized below. Section 6 gives another type of pruning.

Pruning using Basic Domain Knowledge

1. Consider rules derived using the Apriori Algorithm.
2. Use domain expert opinion to determine obvious and uninteresting rules.
3. If a derived rule matches an obvious rule, then prune the derived rule.
4. Store obvious rules in a rule base for future use. These represent uninteresting information.
5. Repeat this process until all rules discovered are considered interesting in the domain.

Another important aspect is the sufficiency of the rules with respect to the problem they aim to solve. Since the goal of the tool is predictive analysis, it is important to determine how many of the likely questions posed by users can be answered by the discovered rules. On analyzing the rules derived from the experimental data stored in relational databases, it was found that the rules were insufficient to answer questions such as, “Given experimental conditions, predict the tendency for distortion in the part due to rapid cooling”. Information about part distortion is not stored as an experimental observation.

However, information on distortion cases, namely the potential causes and solutions, is found in the related literature. For example, the study of several research papers indicates that “Excessive agitation during heat treatment leads to greater distortion in the part.” This can be converted into a rule of the form “Agitation=Excessive=>Distortion=High”. The confidence and support of this rule depends on the number of instances in the papers that satisfy this statement. In order to discover rules such as this, Association Analysis using Apriori is extended to text sources in the domain.

2.4 Apriori over Text Sources of Related Literature

In performing Association Rule Mining over text sources, the first step in our tool is the extraction of plain text into structured text. This is done by defining a set of domain-specific tags representing the entities in the domain and storing the properties or tendencies of the entity as the contents of the tags. For example, consider the entities “Distortion” and “Agitation” and the tendencies “High” and “Excessive”. The sentence, “Excessive agitation during heat treatment leads to greater distortion in the part.” is extracted as one instance of the process “Quenching”, quenching being the rapid cooling step during heat treatment. This is done in the following manner.

```
<Quenching>
        <Distortion>high</Distortion>
         <Agitation>excessive</Agitation>
</Quenching>
```

Thus the tags defined are <Quenching>, <Distortion> and <Agitation>. Likewise facts are extracted from several papers using the necessary tags. The resulting repository of structured text is analogous to the integrated database of experimental data. Details on the text extraction process and the issues emerging from it are discussed in Section 3.

This structured text is then converted into the format required by the Apriori algorithm software, namely the Attribute Relation File Format (ARFF) [16]. Note that this has also been done for the relational data in order to preprocess it for data mining. Next, the formatted files are used for Association Analysis, to discover rules with user-specified minimum confidence and support measures. Rule pruning is done using the same steps as outlined earlier for experimental data, i.e., using basic domain knowledge. The resulting interesting association rules are helpful in predictive analysis. On mining over several instances of facts obtained from many research papers, some of the rules discovered are presented below.

Rules from Association Analysis

1. “Agitation = Excessive => Distortion = High” Confidence = 0.92, Support = 82.
2. “Cold Plastic Deformation = Yes => Residual Stress = Likely” Confidence = 0.86, Support = 75
3. “Device = Impeller Stirrers => Cooling Nature = Uniform”, Confidence = 0.94, Support = 63

2.5 Methodology for Association Rule Mining

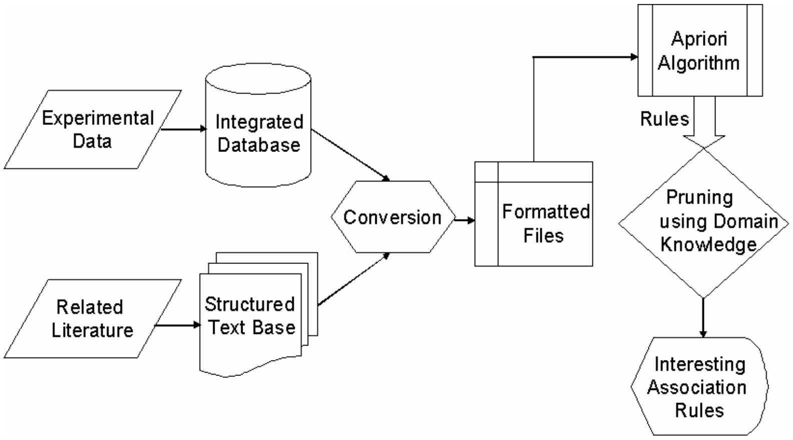


Figure 3: Association Rule Mining in QuenchMiner™

The methodology for Association Rule Mining in QuenchMiner™ is outlined in Figure 3. Each component is explained below.

- Experimental Data: This is the raw data in the domain representing the input conditions and observed results of experiments.
- Integrated Database: This is the common database into which all the relevant data is extracted for mining.
- Related Literature: This refers to research papers and other relevant documents forming text sources of domain knowledge.
- Structured Text Base: This refers to the integrated repository of structured text extracted from the related literature using the domain-specific tags for relevant entities.
- Conversion: This refers to the preprocessing involved in converting the information into the format required for data mining.
- Formatted Files: These are the files in the format required for Apriori Analysis [16].
- Apriori Algorithm: This represents the actual step of applying the Apriori algorithm in the Association Rule Mining process.
- Rules: These are the output of the Apriori Analysis over text and relational sources.
- Pruning using Domain Knowledge: This refers to the pruning done using basic domain knowledge as described earlier. It also refers to pruning using functional dependencies [11] as described in Section 6.
- Interesting Association Rules: These are the rules obtained that are useful for predictive analysis in the given domain.

These interesting association rules are used to populate a knowledge base. These represent the knowledge that a domain expert discovers on learning from experimental data and literature surveys. This can be used for decision support.

2.6 Predictive Analysis using Knowledge Discovered

The interesting association rules discovered using the methodology above representing advanced domain knowledge form the basis for analysis in QuenchMiner™. This refers primarily to the estimation of parameters given input conditions. The algorithm used for predictive analysis for parameter estimation in this tool is shown below.

Algorithm1: Predictive Analysis of Parameters

```
FOR y = 1 to m          /* number of input variables */
    Iy.value = user-entry   /* list of input variables */
FOR x = 1 to n          /* number of output parameters */
    Ox.name = user-select        /* list of output parameters */
FOR x = 1 to n          /* iterate through each output parameter */
    v1 = 0, v2 = 0    /* initialize variables for tendencies */
    FOR y = 1 to m   /* identify tendencies */
        IF Ox := Iy THEN   /* if output parameter depends on input variable */
                IF Iy.value => v1   /* v1 represents one extreme tendency */
                THEN v1 = v1 + wgt1   /* wgt1 is extent of impact */
                ELSE IF Iy.value =>  v2   /* v2 represents other extreme tendency */
                    THEN v2 = v2 + wgt2   /* update variable by weight. */
    IF v1 > v2   /* check which tendency is greater */
    THEN final-tendency ~ v1   /* overall tendency corresponds to the v1 extreme */
    ELSE IF v1 < v2
        THEN final-tendency ~ v2   /* overall tendency corresponds to the v2 extreme */
        ELSE final-tendency ~ avg (v1, v2)   /* overall tendency corresponds to average of extremes */
    Ox.value = final-tendency   /* predict overall tendency */
FOR x = 1 to n    /* for each output parameter */
    OUTPUT Ox.value   /* convey predicted decision to user */
```

The rule confidence and support derived by Apriori help to determine the impact of the input conditions such as part surface on the output parameters such as cooling rate. The extent of the impact is represented by weights in the algorithm.

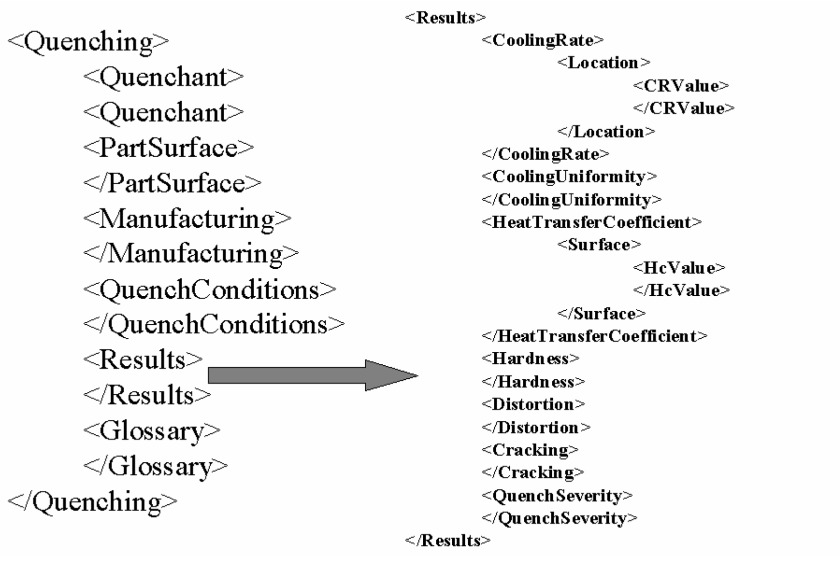
3. Issues in Knowledge Discovery

One of the steps in the methodology for Association Analysis is the extraction of plain text into structured text capturing the domain-specific aspects significant in data mining. This is proposed using Natural Language Processing [12] as follows.

Automating Text Extraction for Data Mining

1. Define a domain-specific markup language with tags and nesting representing domain entities and relationships.
2. Use Natural Language Processing to parse each plain text document.
3. Fill tags of domain-specific markup language by mapping natural language to domain semantics.
4. Address issues such as synonyms and homographs through an ontology.
5. Use basic domain knowledge to refine contents of tags, analogous to “Pruning using Basic Domain Knowledge”.

There are several challenges involved in this, especially in dealing with ambiguity. This effort is a subject of ongoing research at the Center for Heat Treating Excellence, WPI. A domain-specific markup language [31] has been defined for the heat treating of materials and is proposed to be included as a semantic extension to “MatML, the XML for Materials Property Data [7]” developed by the National Institute of Standards and technology (NIST) [22]. This markup language, in addition to facilitating data storage and exchange worldwide, would also set the stage for extraction of plain text into structured text in the domain. Figure 4 shows an overview of the proposed tags for this markup language. Using this markup language, it is proposed that plain text be converted to structured text.



```
<Quenching>
        <Quenchant>
        <Quenchant>
        <PartSurface>
        </PartSurface>
        <Manufacturing>
        </Manufacturing>
        <QuenchConditions>
        </QuenchConditions>
        <Results>
        </Results>
        <Glossary>
        </Glossary>
</Quenching>
```

```
<Results>
        <CoolingRate>
                <Location>
                        <CRValue>
                        </CRValue>
                </Location>
        </CoolingRate>
        <CoolingUniformity>
        </CoolingUniformity>
        <HeatTransferCoefficient>
                <Surface>
                        <HcValue>
                        </HcValue>
                </Surface>
        </HeatTransferCoefficient>
        <Hardness>
        </Hardness>
        <Distortion>
        </Distortion>
        <Cracking>
        </Cracking>
        <QuenchSeverity>
        </QuenchSeverity>
</Results>
```

Figure 4: Overview of Proposed Markup Language

The issue of synonyms and homographs [12] is also being addressed. Consider the following segments of structured text.

```
<Shape> geared </Shape>
<Shape> corrugated </Shape>
```

It is important to know that the terms “geared” and “corrugated” mean the same in the domain [5]. Otherwise, facts related to “geared” and “corrugated” would be extracted as referring to different shapes. This would later pose a problem in Apriori analysis, since the count of items pertaining to “geared” would not get updated if the term in another item is “corrugated”, although these two imply the same shape. Issues such as this are solved through ontological developments. In addition to a schema for the markup language, an ontology has been defined that takes into account the synonyms and homographs [12] in a domain-specific manner. Details on this are in [31]. Using this markup language, facts pertaining to “geared” and “corrugated” would be extracted as instances of the same shape, thus solving the above problem.

4. Microstructure Prediction using Visualization Techniques and Domain Knowledge

The second aspect of predictive analysis in QuenchMiner™ is the prediction of material microstructure during various stages of heat treatment of a given material. A microstructure is what one sees when an alloy specimen is cut, its surface is polished and etched to expose phases, and it is put under a microscope [5, 29]. Predicting microstructures of the alloy interests materials scientists and engineers because microstructures dictate mechanical properties such as hardness, toughness and ductility, and hence enable materials selection for specific processes based on these properties. Data visualization is useful in microstructure prediction.

4.1 Data Visualization

Data visualization is a technique to present a set of data in the form of graphical depictions [15, 27]. Visualization is applied extensively in the various disciplines of science [27, 34]. In the process of scientific visualization the collected raw data should to be filtered and smoothed as necessary since measurements taken from experiments often contain noise. Then the data is mapped to geometric primitives that effectively represent the meaning of the data. The images generated from the visualization tool should reduce the users' cognitive loads.

Time-temperature curves, cooling rate curves, and heat transfer coefficient curves [26] are conventional methods for data representation in the Materials Science domain. While they are effective to represent the quality of a single quenching process, they may not be best suited to represent multiple processes or the whole database content at the same time. A multivariate data visualization tool, called the Xmdv Tool [34] provides an alternative to determine trends, similarities, and dissimilarities among a group of quenching processes and view a large number of data sets at a glance. There are various techniques such as [34] scatter-plots, parallel coordinates plots, glyphs, line graphs and pie charts, found in this tool.

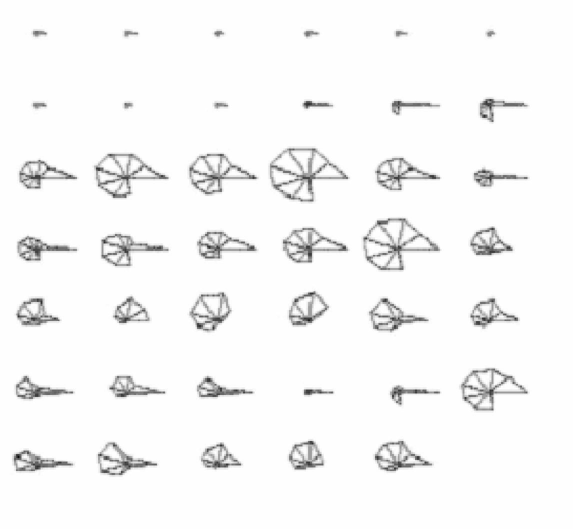


Figure 5: Star Glyphs Plot

An example of star glyphs is shown in Figure 5. The example shows the plot of heat transfer coefficients obtained from heat treatment experiment using various quenchants (cooling media) and experimental conditions. Each vertex represents a parameter and the distance from the center of the star represents its value. The number of the parameters and their combinations can be customized according to the user preferences. Clusters and similarities can be identified by comparing their shapes and sizes. In the given example plot, all the small stars are results from gas quenchants. The results from water quenchants show relatively large stars.

4.2 Role of Domain Knowledge

In addition to visualization techniques, the role of domain knowledge is critical in microstructure prediction. The basic information is provided by a time-temperature curve [3] which is a plot of temperature versus time during the rapid cooling of a material being heat treated. The volume fractions of the phases present in the resulting microstructure can be determined by tracing the regions of the time-temperature curve and their duration [23, 3]. As the cooling progresses, new phases start to form when the cooling curve reaches different regions and grains grow at the same time. The changes in volume fractions during the cooling process are commonly represented using a line graph of volume fraction versus time. The following equation [23] is the theoretical explanation behind the mechanism of the visualization model.

$$f(\theta) = \int 1 - e^{-B(T)\theta^{N(T)}}$$

Here f (θ) is the fraction of pearlite or bainite after some time θ, and B (T) and N (T) are constants that depend on the properties of each material [23]. Martensite has slightly different transformation kinetics and the volume fraction of martensite, f_m at temperature T can be obtained by the following equation [5, 23, 27].

$$f_m = (1 - e^{-\alpha(T_{ms} - T)})(1 - \sum_i F_i)$$

α in the equation is a coefficient taken from the literature [23] F_i is the volume fraction of pearlite or bainite. Thus, domain knowledge and visualization techniques both are helpful in microstructure prediction. A few domain-specific visual tools are explained.

4.3 Domain Specific Visual Tools

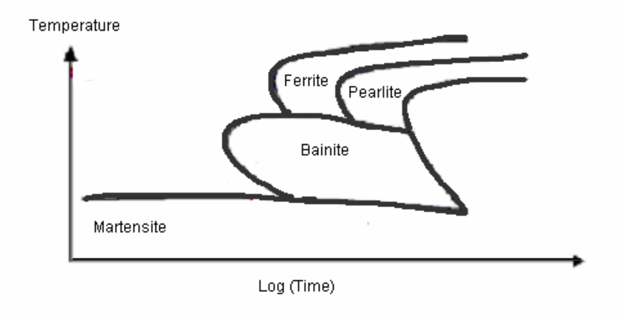


Figure 6: CCT diagram (Source: Ref. No. [3])

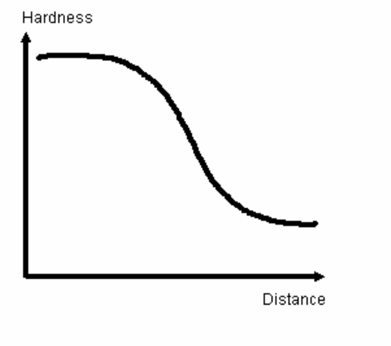


Figure 7: Jominy end quench graph (Source: [3])

Tools such as continous cooling transformation (CCT) diagrams and Jominy end quench graphs [3] embody domain-specific aspects in Materials Science. A continuous cooling transformation diagram shows which phase starts developing at what time and what temperature. It is a plot of temperature versus the logarithm of time. It depends on the chemical composition of the materials, thus different materials have different CCT diagrams. These carry phase transformation information. Figure 6 shows a CCT diagram [3]. A Jominy end quench graph is the plot of hardness versus distance, showing cooling phenomena at different locations of a material. It is based on the Jominy end quench test, which is performed by the rapid cooling (quenching) of a part from one end. The test results indicate how fast cooling occurred at different locations of the part. Therefore, this test supplies interesting information about the cooling phenomenon during the quenching of a large part. Figure 7 shows a Jominy end quench graph [3].

4.4 Methodology for Microstructure Prediction

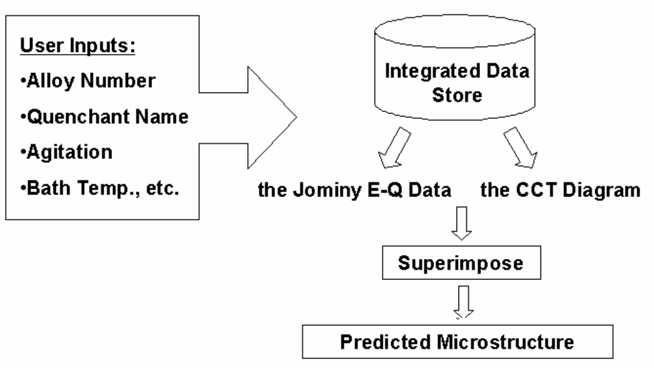


Figure 8: Microstructure Prediction in QuenchMiner™

Based on this domain knowledge and visualization techniques, the methodology for microstructure prediction is as shown in Figure 8. The superimposing of the Jominy End Quench Test results over the CCT diagram of the material of interest enables the prediction of the microstructure development through the given quenching process. This enables the visualization of the final microstructure at each point, namely at different locations for different specimens. Even more challenging is simulating the actual evolutions of microstructure during the process of rapid cooling of a material in heat treatment. This is discussed in the next section.

5. Game-of-Life for Simulating Microstructure Evolution at Different Locations

The Game-of-Life, originally created by Jon Conway in 1970, simulates the birth and death of cells in a society. A cell is born or dies according to a set of four rules [9].

1. A cell is born or dies if exactly 3 of its neighbors are alive.
2. An existing cell stays alive if there were either 2 or 3 neighbors alive.
3. A cell will die from isolation if there are fewer than 2 neighbors alive at any given time.
4. A cell will die from overcrowding if there are more than 3 neighbors at any given time.

This is a classical computer science problem and can be solved with relatively little effort using two-dimensional arrays. The Game-of-Life is often used in the studies of cellular automata and artificial intelligence [9]. The simulation of microstructure evolution is based on this Game-of-Life process, which operates in the domain-specific sets of rules. The rules relevant to the Materials Science domain are listed below [23, 29] and their application in the tool is explained in Example1.

- Each pixel in the image field stores the likelihood for being transformed into different phases, i.e, whether it is still available (can be transformed), and which phase it belongs to, if it is already a part of a phase.
- In each iteration the pixels with the highest likelihood to become a particular phase are picked and get transformed to be parts of the phase.

Example1: Microstructure evolution in Steel: At time 0, the only phase present is Austenite [3, 19]. Therefore, all he pixels are marked as Austenite crystal or Austenite crystal boundary. Since crystallization of any phase always starts from Austenite crystal boundaries, a pixel adjacent to a pixel that is marked as boundary has higher likelihood to be transformed into other phases. At the implementation level, this pixel gets 1 point each for the likelihood to be Ferrite, Pearlite, Bainite, or Martensite [3, 23]. If the pixel is adjacent to multiple pixels that were marked as boundary, the points accumulate. The pixel that is adjacent to 3 boundary pixels has 3 points. This pixel will be picked before the pixels with fewer points. If this pixel gets picked and gets transformed into Ferrite, it is marked as Ferrite and "not available" (only Austenite crystals can be transformed into some other crystals during quenching). Pixels that are adjacent to the Ferrite pixel now get 1 point each for the likelihood to become another Ferrite pixel. Since different types of crystals have different rules for their growth, different sets of rules apply for different phases in keeping with the points [3, 19, 23]. For the case of ST4140, the process starts from 100 % Austenite [3, 23, 29]. As the cooling progress, Austenite phase transforms into other phases, such as Ferrite, Pearlite, Bainite and Martensite. The phases present in the quenched specimen vary depending on the cooling rate. The changes in volume fractions during microstructure evolution can be represented in two ways, a line graph or a pie-chart [29, 34]. These represent how the fractions evolve.

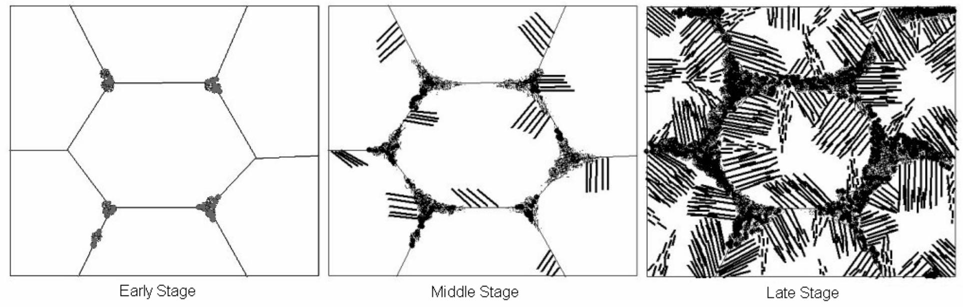


Figure 9: Visualizing Microstructure Evolution

Figure 9 shows the microstructure evolution for ST4140 at the location of 2 inches from one end of the specimen. It represents three snapshots at early, middle and late stages of evolution respectively. The Xmdv tool [34] provides some of the techniques needed to generate these pictures. These snapshots are screen-dumps taken during a demo. The evolution is seen more clearly in a live demo, which is available on the Web for the authorized users of this tool.

6. Dealing with Uncertainty

Uncertainty in prediction occurs because the system may not have access to the whole truth about the environment, or because there may be incompleteness and incorrectness in understanding the properties of the environment [28, 30]. Treatment of uncertainty has long been studied in artificial intelligence [30] following different perspectives. It has been argued by Zadeh that probability theory is not adequate for the treatment of uncertainty [35]. There are various aspects of uncertainty as follows.

6.1 Occurrence of Conflicts

This refers to two or more input conditions leading to opposing results. For example, in our context, one rule may indicate that the nature of the cooling is uniform, while another may indicate that it is non-uniform. This problem is solved through the use of good conflict resolution strategies [17, 28]. The various means of conflict resolution as incorporated in QuenchMiner™ are:

- No duplication: Do not execute the same rule on the same arguments twice.
- Recency: Prefer rules that refer to recently created working memory elements.
- Specificity: Prefer rules that are more specific. For example, in case of conflicts, prefer "X and Y => Z" over "X=>Z".
- Weights: Attach weights to each rule based on the extent of impact and estimate the overall tendency for each parameter, after considering all the weights.

The forward chaining [6] principle that finds every conclusion possible based on a given set of premises is used here. This ensures that once a rule fires, it is removed from the rule list, and rule application stops only if no more rules can be fired. This also helps to address uncertainty due to conflicts. No matter where a rule occurs in the list its effect is considered and no variable gets updated more than once due to the same rule.

6.2 Semantic Deficiency

Consider the following scenario [24] that could arise in any domain even after probability theory is used.

- Rule i: IF Ai THEN *B* for i = 1 to k.
- As more Ai'*s* are confirmed, B becomes more credible.
- What if all the Ai*'s* are correlated.

In our domain such a situation could arise. For example, "Cooling Rate = Fast => Heat Transfer Coefficient = High". However, fast cooling itself is correlated with excessive agitation, i.e., excessive agitation is one of the causes of fast cooling. Since the Ai's of Rule i are correlated, the variable for "High Heat Transfer Coefficient" would get updated twice, thus leading to a higher prediction of "Heat Transfer Coefficient" than expected. This is a semantic deficiency. This issue is related to correlations and hence functional dependencies between variables. The solution proposed to such problems is pruning using functional dependencies. It is important to note the difference between a functional dependency and an association rule [11]. A functional dependency is a statement of certainty, while an association rule represents a probability. For example, the fact that heat transfer coefficient depends on cooling rate is a definite statement. This can be represented as "CR → hc" or "hc := CR", where "CR" is "Cooling Rate" and "hc" is "Heat Transfer

Coefficient." This is true in all cases, i.e., there is no issue of confidence and support. A dependency is thus a more solid relationship than an association rule. Pruning rules using functional dependencies shown below overcomes the problem of semantics deficiencies.

Pruning Rules using Functional Dependencies

1. Identify functional dependencies [5, 11] between variables using cause-effect analysis [21].

2. If C depends on B and B depends on A, then C depends on A. Hence prune the rule(s) with lower confidence. (If A=>C has lower confidence than the two individual rules, prune it.)

3. Test the remaining rules on a validation set [28].

4. If semantic deficiencies arise, then continue pruning, else stop.

5. Output the set of rules as rules without semantic deficiencies.

6.3 Degrees of Uncertainty

For each parameter that is being estimated by the tool, the default levels are high and low. For example, consider the parameter "distortion". This represents the tendency of a part to undergo deformation in shape and / or size due to mechanical processes [5]. The distortion either occurs or does not. An extreme case of distortion is "cracking", where a part actually breaks during a process [5]. The presence or absence of cracking is an important aspect of estimation. However, in many situations, it is not possible to provide a categorical estimate, i.e., a "yes-no" answer. Hence, we define multiple levels of abstractions to represent degrees of uncertainty in order to account for the gray areas. These are:

- Level 1: High, Low (Most Certain)
- Level 2: On higher side, On lower side (Less Certain)
- Level 3: Medium (Uncertain)

Note that level 3 could be inferred, if all the input conditions are such that the high and low tendencies of a parameter are equally likely. For instance, QuenchMiner™ could conclude in a distortion case, that the tendency for distortion is "medium or cannot be determined using given conditions". The user could then continue with the analysis of the case by altering input conditions if desired.

7. Experimental Evaluation

QuenchMiner™ has been subject to rigorous evaluation to determine its effectiveness. Several experiments have been carried out using this tool and the corresponding results have been compared with real experiments in the domain. The evaluation criteria are:

- Accuracy: This is measured in terms of the deviation of predicted results from real domain results. If the error is less than 5%, the accuracy is acceptable.
- Efficiency*:* This is measured as the response time taken by the predictive tool to estimate the results. If it is less than 5 minutes, the efficiency is acceptable.

The process of conducting the experiments is demonstrated in Example2. This refers to a case submitted by domain users.

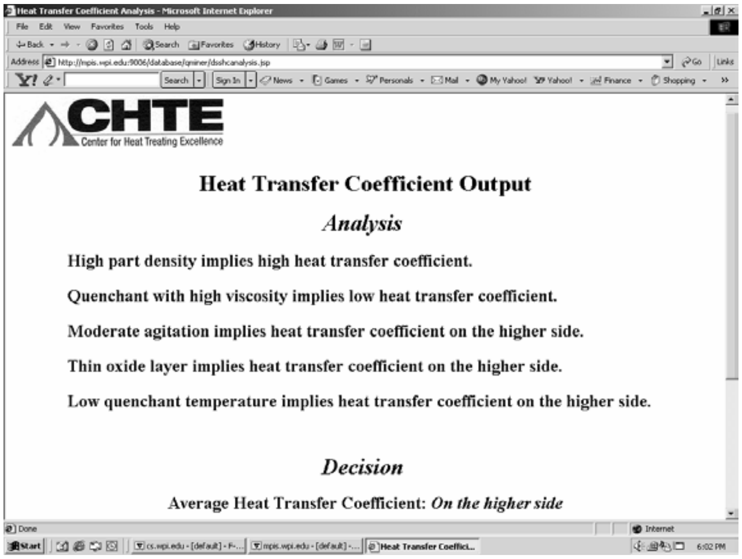
Example2: Estimate the average heat transfer coefficient (heat extraction capacity) in this quenching (rapid cooling) process, given the following inputs.

1. Temperature of Quenchant (cooling medium) : Low
2. Agitation Velocity: Moderate
3. Quenchant Viscosity: High
4. Part Density: High
5. Oxide Layer: Thin

Processing in Example2 using Algorithm1 and Levels of Abstraction

```
m = 5    /* 5 input variables */
FOR y = 1 to 5   / * read values of input variables */
    I1.value = "Low", I2.value = "Moderate", I3.value = "High", I4.value = "High", I5.value = "Thin"
n = 1     /* 1 output parameter */
For x = 1
    Ox.name = "Heat_Transfer_Coefficient"       /* read the output parameter(s) selected by user */
For x = 1    /* the only output parameter is heat transfer coefficient */
    v1 = 0, v2 = 0             /* initialize variables for tendencies, v1 = high, v2 = low */
    For y = 1 to 5             /* identify tendencies */
            Ox := I1   /* heat transfer coefficient depends on quenchant temperature */
                    I1.value = "Low" => v1   /* low quenchant temperature implies heat transfer coefficient on higher side */
                    v1 = 0 + 1   /* weight is 1 for impact of low quenchant temperature on high heat transfer coefficient */
            Ox := I2      /* heat transfer coefficient depends on agitation velocity */
                    I2.value = "Moderate" => v1   /* moderate agitation implies heat transfer coefficient on higher side */
                    v1 = 1 + 1   /* weight is 1 for impact of moderate agitation on high heat transfer coefficient */
            Ox := I3   /* heat transfer coefficient depends on quenchant viscosity */
                    I3.value = "High" => v2              /* quenchant with high viscosity implies low heat transfer coefficient */
                    v2 = 0 + 4    /* weight is 4 for impact of high viscosity on low heat transfer coefficient */
            Ox := I4      /* heat transfer coefficient depends on part density */
                    I4.value = "High" => v1              /* high part density implies high heat transfer coefficient */
                    v1 = 2 + 3   /* weight is 3 for impact of high part density on high heat transfer coefficient */
            Ox := I5     /* heat transfer coefficient depends on oxide layer */
                    I5.value = "Thin" => v1   /* thin oxide layer implies heat transfer coefficient on the higher side */
                    v1 = 5 + 2   /* weight is 2 for impact of thin oxide layer on high heat transfer coefficient */
    v1= 7;   v2 = 4;   v1 > v2   /* variable for high tendency has greater weight than variable for low tendency */
    final-tendency ~ v1  /*  overall tendency corresponds to high heat transfer coefficient */
For x = 1   /* only output parameter is heat transfer coefficient
    Ox.value = "high" /* heat transfer coefficient is on higher side considering abstraction levels and the difference (v1- v2) */
    Output Ox.value   /* Convey decision to user "Average Heat Transfer Coefficient: On the higher side" */
```

Following the above steps QuenchMiner™ returns the predicted output for Example 2, as shown in Figure 10. On using the same inputs to conduct a real domain experiment, and using its results to plot heat transfer coefficients versus temperature, a graph is obtained as shown in Figure 11. On analyzing this graph, a domain expert would infer that the average heat transfer coefficient is on the higher side, which is identical to the estimation of the tool. Thus, in this case the error is zero. The time taken to compute this response is approximately 1 second. Thus response time and error are both within acceptable limits, as verified by domain experts.


Heat Transfer Coefficient Analysis - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites History
Address http://mpis.wpi.edu:9006/database/qminer/dsshcanalysis.jsp
Go Links
Search Sign In News Games Personals Mail My Yahoo! Yahoo! Finance Shopping
CHTE
Center for Heat Treating Excellence
Done
Internet
Start
6:02 PM

Heat Transfer Coefficient Output

Analysis

High part density implies high heat transfer coefficient.

Quenchant with high viscosity implies low heat transfer coefficient.

Moderate agitation implies heat transfer coefficient on the higher side.

Thin oxide layer implies heat transfer coefficient on the higher side.

Low quenchant temperature implies heat transfer coefficient on the higher side.

Decision

Average Heat Transfer Coefficient: *On the higher side*

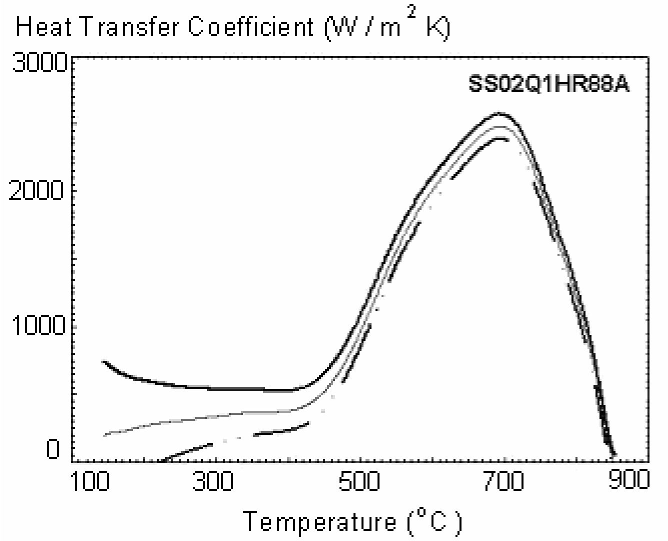


Figure 10: Predicted Output for Example 2

Figure 11: Actual Heat Transfer Coefficient Graph

Likewise, several experiments were conducted using QuenchMiner™ for other cases of parameter estimation and for microstructure prediction. The results of these experiments are summarized below.

- Parameter Estimation Experiments:
  - Number of Tool Experiments: 200
  - Number of Inputs (conditions): 12
  - Number of Outputs (parameters): 8
  - Range of Error in Tool: Between 1 and 3%
  - Average Response Time of Tool: 1 second
  - Domain Experiment Time: 4 hours
- Microstructure Simulation Experiments:
  - Number of Tool Experiments: 100
  - Number of Inputs (location, material): 15
  - Number of Outputs (evolution stages): 10
  - Range of Error in Tool: Between 2 and 4%
  - Average Response Time of Tool: 1 minute
  - Domain Experiment Time: 6 hours

The general observations from all the experiments are as follows.

1. As the number of inputs and outputs increase, the time for parameter estimation increases only by a fraction of a second.

2. Increasing the number of input conditions supplied by the user increases the accuracy of the parameter estimation.

3. Increasing the number of output parameters to be estimated does not reduce the accuracy of the estimation.

4. Taking snapshots of microstructure evolutions at later stages of evolution needs more time.

5. Accuracy in microstructure prediction is almost identical for all locations in all materials.

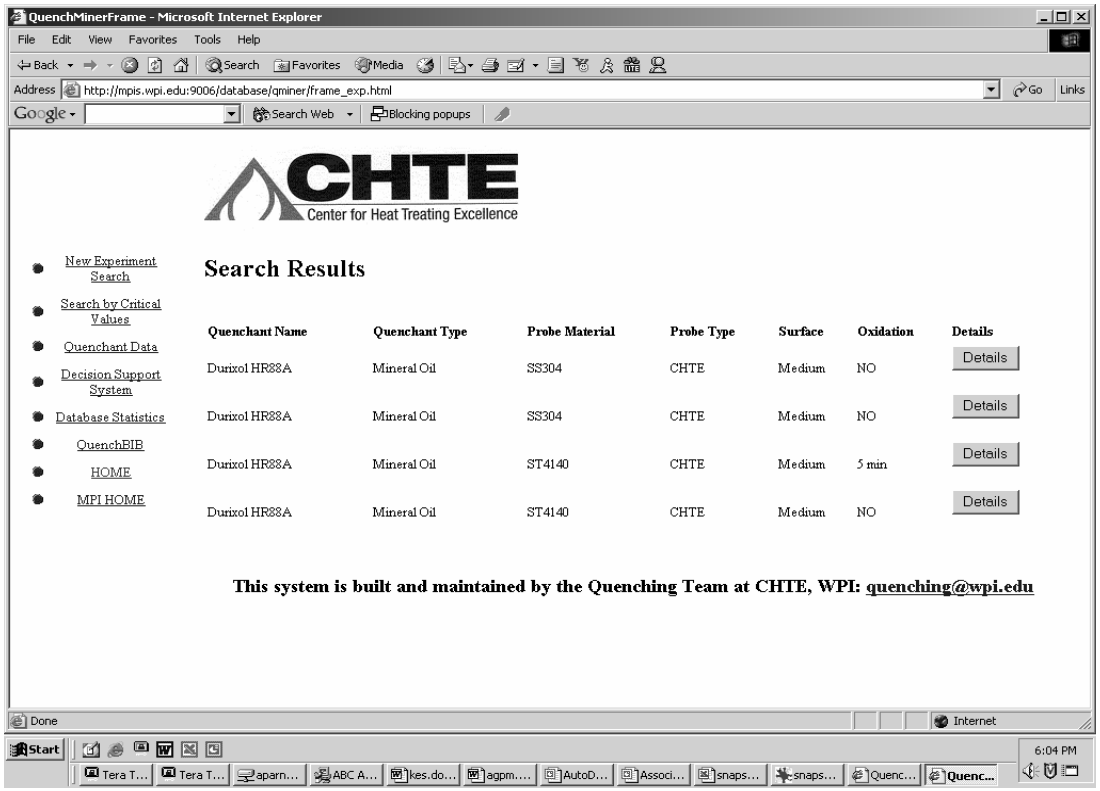
On considering various experimental criteria and comparing he predicted results with the real results, the QuenchMiner™ tool is found to be satisfactory for predictive analysis in Materials Science as verified by domain experts.

8. Application to E-Business

E- Business (Electronic Business) is, in its simplest form, the conducting of business on the Internet. It is a more generic term than E-Commerce because it refers to not only buying and selling but also servicing customers and collaborating with business partners [14]. Today, many corporations are reorganizing their businesses in terms of the Internet and its capabilities. Many organizations are considering computer automation as a means of better serving their customers, improving efficiency, reducing costs, and providing a positive impact for their company [18]. Some interesting elements of E-Business are summary tables, aggregated information, query mechanisms, graphs and pictures [10, 14, 25, 34]. These are provided in our tool for retrieval and analysis of data with the goal of decision support for a particular community of knowledge workers, in our case Materials Scientists.

The main aim in the development of the QuenchMiner™ tool is to provide at-a-glance information to Materials Science users worldwide. These include materials suppliers, automobile companies, heat treatment industries, universities, researchers, aerospace agencies, manufacturing companies and others [5, 7]. The information helps to connect these different categories of users, assisting them in several aspects of knowledge exchange and business decision support.

Figure 12 is an example of at-a-glance information for decision support. This refers to a basic search engine functionality of QuenchMiner™ [32]. It pulls out the experimental input conditions and results from an underlying integrated database, QuenchPAD™ [32], in response to a user query. In Figure 12, the tool returns a response to the query “retrieve all experiments conducted using the CHTE probe and with the mineral oil DHR88A as a cooling medium”. The user interaction occurs through the Web. This provides Web-based information retrieval. This was one of the earliest accomplishments of QuenchMiner™ that motivated the development of more advanced features such as prediction of experimental parameters and material microstructure to achieve decision support for optimization of the involved processes [33]. Predicting parameters such as cooling rates and heat transfer coefficients given experimental input conditions is helpful because these parameters characterize the experiments and hence help in the optimization of the corresponding real processes in the industry. Predicting microstructures of the alloy interests materials scientists and engineers because microstructures control the mechanical properties of materials such as their hardness, toughness and ductility. Hence this enables materials selection for specific processes based on these properties, in turn helping to optimize products. Optimization of products and processes increases customer satisfaction, thus enhancing business.


QuenchMinerFrame - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Media
Address
http://mpis.wpi.edu:9006/database/qminer/frame_exp.html
Go
Links
Google
Search Web
Blocking popups
CHTE
Center for Heat Treating Excellence
New Experiment Search
Search by Critical Values
Quenchant Data
Decision Support System
Database Statistics
QuenchBIB
HOME
MPI HOME
Done
Internet
Start
6:04 PM

Search Results

Quenchant Name	Quenchant Type	Probe Material	Probe Type	Surface	Oxidation	Details
Durixol HR88A	Mineral Oil	SS304	CHTE	Medium	NO	Details
Durixol HR88A	Mineral Oil	SS304	CHTE	Medium	NO	Details
Durixol HR88A	Mineral Oil	ST4140	CHTE	Medium	5 min	Details
Durixol HR88A	Mineral Oil	ST4140	CHTE	Medium	NO	Details

This system is built and maintained by the Quenching Team at CHTE, WPI: quenching@wpi.edu

Figure 12: At-a-glance information for decision support

A computational tool that provides predictive analysis helps in making faster, easier and more sophisticated decisions. For example, a materials supplier may want to know what quenchants (cooling media) a particular heat treating company requires in a fiscal quarter. The company users could run case studies with this predictive tool, to estimate the types of quenchants needed to achieve a desired output, as per their targeted goal for that quarter. The materials supplier could be informed accordingly. Since the response time of the tool is fast, this enables a quick and accurate estimate, resulting in more effective buying and selling decisions.

9. Conclusions

The Apriori algorithm and the Game-of-Life process have been used as the basis for predictive analysis to build a tool called QuenchMiner™ for decision support in Materials Science. This provides parameter estimation and microstructure simulation for heat treating processes. To the best of our knowledge, this tool is the first of its kind integrating domain-type-dependent data mining and data visualization for decision support of mechanical engineering processes. Since it is a Web-based tool, it connects clientele worldwide and allows them to exchange useful knowledge for business decision support. This promotes E-Business in Materials Science. Several aspects of predictive analysis are topics of our ongoing research at WPI.

Acknowledgements

This work is supported by the Center for Heat Treating Excellence (CHTE) and its member companies, and by the Department of Energy – Office of Industrial Technology (DOE-OIT) Award Number DE-FC-07-011D14197. The authors thank the DKBRG / AIRG groups in the Computer Science Department and the Quenching team in the Materials Science Department at WPI.

References

[1] Agrawal R., Imilinski T. and Swami A., Mining association rules between sets of items in large databases, ACM SIGMOD (1993), 207-216.

[2] Agrawal R. and Srikant R., Fast algorithms for mining association rules, VLDB (1994), 487-499.

[3] American Society for Metals, Atlas of isothermal transformation and cooling, ASM Publishers, OH, USA, 1977.

[4] Armstrong W. and Deobel C., Decompositions and functional dependencies in relations, In: ACM Transactions on Database Systems, 1980, (5):4, 404-430.

[5] Askeland D., The science and engineering of materials, Wadsworth Inc., CA, USA, 1984.

[6] Bacchus F. and Teh Y.W., Making forward chaining relevant, AI Planning Systems (1998), 54 - 61.

[7] Begley E.F., MatML version 3.0 schema, NIST 6939, National Institute of Standards and Technology Report, USA, Jan 2003.

[8] Cristea A. and Okamoto T., Energy function construction and implementation for stock exchange prediction NNs, In: KES, 1998, (3), pp.403-410.

[9] Dekker A.H., The game of life: A CLEAN programming tutorial and case study, In: SIGPLAN Notices, 1994, (29): 9, pp.91-104.

[10] Fayyad U., Piatetsky-Shapiro G. and Smyth P., From data mining to knowledge discovery in databases, In: AAAI Magazine, 1996, (Fall), pp.37-53.

[11] Han J. and Kamber M., Data mining: concepts and techniques, Morgan Kaufmann Publishers, CA, USA, 2001

[12] Hickok G. and Poeppel D., Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language, In: Cognition, 2004, (92):1, pp.67-99.

[13] Huang D. Arimoto K., Lee K., Lambert D. and Narazaki M., Prediction of quench distortion on steel shaft with keyway by computer simulation", ASM HTS (2000), 708–712.

[14] IBM, Connecting people to your e-business efficiently, IBM WebSphere White Paper, Dec 2002.

[15] Keim D., Information visualization techniques for exploring large databases, Visual Databases (2000).

[16] Kirkby R., Frank E., Holmes G., Hall M., Racing committees for large datasets, Discovery Science (2002).

[17] Klein M., Supporting conflict resolution in cooperative design systems, In: IEEE Transactions on Systems, Man, and Cybernetics, 1991, (21): 6, pp.1379-1390.

[18] Microsoft, Microsoft solution for Internet business, Microsoft Corporation White Paper, Oct 2002.

[19] Mills A., Heat and mass transfer, Irwin Publishers, CA, USA, 1995.

[20] Miyano T., Girosi F., Forecasting global temperature variations by neural networks, MIT AI Lab, 1994.

[21] Nakajo T.and Kume H., A case history analysis of software error cause-effect relationships, In: IEEE Transactions on Software Engineering, 2002, pp. 830-838.

[22] National Institute of Standards and Technology, http://www.nist.gov/ , MD, USA.

[23] Pakalapati R., Jin L., Farris T.N., and Chandrasekar S., Simulation of quenching of steels: effect of different multiphase constitute models, ASM HTS (1999), 416-423.

[24] Pearl J., Probabilistic reasoning in intelligent systems: networks of plausible inference, Morgan Kaufmann, CA, USA, 1988.

[25] Petrucelli J., Nandram B. and Chen M., Applied statistics for scientists and engineers, NJ, USA, 1999.

[26] Totten G., Bayes C., Clinton N., Handbook of quench technology and quenchants, ASM International, OH, USA, 1993.

[27] Rohrer M., Seeing is believing: the importance of visualization manufacturing simulation, Simulation (2000), 1211 - 1216.

[28] Russell S. and Norvig P., Artificial intelligence: a modern approach, Prentice Hall, USA, 1995.

[29] Santella M.L., Babu S.S., Biemer B.W. and Feng Z., Influence of Microstructure on the Properties of Resistance Spot Weld, Trends in Welding (1998), 605-609.

[30] Saotti A., An AI View of the Treatment of Uncertainty, In: The Knowledge Engineering Review, 1987, (2): pp. 75-97.

[31] Varde A., Rundensteiner E., Mani M., Maniruzzaman M. and Sisson R. Jr. Augmenting MatML with Heat Treating Semantics, Submitted to ASM Materials Solutions Conference, Columbus, OH, Oct 2004.

[32] Varde A., Takahashi M., Maniruzzaman M. and Sisson R. Jr., Web-based data mining for uenching analysis, IFHTSE (2002).

[33] Varde A., Takahashi M., Rundensteiner E., Ward M., Maniruzzaman M. and Sisson R. Jr, QuenchMiner™: decision support for optimization of heat treating processes, IICAI (2003), 993-1003.

[34] Ward M., Xmdv Tool: integrating multiple methods for visualizing multivariate data, Visualization (1994) pp. 326-333.

[35] Zadeh L.A., Possibility theory and soft data analysis, In Mathematical Frontiers of the Social and Policy Sciences, WestView Press, CO, USA, 1981, pp 69-129..