

# Estimating Heat Transfer Coefficients as a Function of Temperature by Data Mining

Aparna S. Varde, Elke A. Rundensteiner, Mohammed Maniruzzaman and Richard D. Sisson Jr.

Worcester Polytechnic Institute (WPI),  
Worcester, MA 01609, USA.  
( aparna | rundenst | maniruzz | sisson ) @ wpi.edu

## Abstract

This paper describes our proposed technique AutoDomainMine that performs data mining guided by fundamental knowledge of the domain. The data being mined consists of input conditions from quenching experiments and the resulting heat transfer curves, i.e., plots of heat transfer coefficients versus part temperature. Since heat transfer coefficients characterize quenching, the estimation assists decision-making. This avoids running laboratory experiments which consume considerable time and resources. AutoDomainMine integrates two data mining techniques, clustering and classification, into a learning strategy. It clusters curves resulting from existing experiments and uses decision tree classifiers to learn the clustering criteria, i.e., input conditions characterizing the clusters. The learned criteria are used to design a representative pair of input conditions and heat transfer curve per cluster. The decision trees and representatives serve as the basis for estimation. When input conditions of an unperformed experiment are submitted, the decision tree path is traced to estimate its cluster and hence the corresponding heat transfer curve. Also when a desired heat transfer curve is submitted, it is compared with the representative curves. The input conditions of the closest matching curve are the estimated conditions to achieve the desired curve. AutoDomainMine on evaluation gives accuracy higher than state-of-the-art estimation techniques.

## Keywords

Computational Estimation, Knowledge Base, Decision Support, Data Analysis, Numerical Prediction.

## 1. Introduction

The results of experiments in quenching heat treatment are often plotted as heat transfer curves to aid analysis and comparison of the corresponding processes. Performing laboratory experiments consumes significant time and resources. This motivates the need for computational estimation of the resulting curves given the input conditions of an experiment. That is, there is a need to computationally estimate heat transfer curves as a function of temperature.

This computational estimation can be used in various applications such as simulations and decision support. It is also desirable to estimate the conditions that would obtain a particular heat transfer curve, if the laboratory experiment were performed. This estimation can be useful in selecting process parameters for industrial heat treatment processes.

The goals of the required estimation technique are stated below [19] with specific reference to quenching heat treatment.

1. Given the input conditions of a quenching experiment, estimate the resulting heat transfer curve that would be obtained.
2. Given the desired heat transfer curve in a quenching experiment, estimate a set of input conditions that would obtain it.

Techniques such as naïve and weighted similarity search [7, 22], case-based reasoning approaches [1, 8, 13, 15] and mathematical modeling in heat treating [14, 4] are not accurate enough. This elaborated in the related work section of this paper. There is a need to develop a technique that performs the desired estimation in less time than a laboratory experiment, with minimal domain expert intervention and with accuracy acceptable for decision support in the domain.

We propose a computational estimation technique based on data mining [19]. Data mining is the process of discovering interesting patterns and trends in large datasets to guide decisions about future activities [7]. In AutoDomainMine, the two data mining techniques of clustering [9] and decision tree classification [10] are integrated into a learning strategy for estimation. This approach automates a typical learning method of scientists. In this paper, the AutoDomainMine approach is described with particular reference to heat treating.

AutoDomainMine has been subjected to rigorous evaluation [19] with quenching experimental data. It is found that estimation given by AutoDomainMine is better than what would be obtained state-of-the-art approaches such as similarity search [7, 22]. The estimation of heat transfer coefficients as a function of temperature using AutoDomainMine, is considered useful for decision support in heat treating, as corroborated by domain experts.

AutoDomainMine has several applications. The main application is enhancing the decision support functionality of our earlier system QuenchMiner™ [21]. Also, it can be used

to select process parameters for other industrial applications in quenching. More applications include serving as the input to software tools such as DANTE [6], DEFORM [16], SYSWELD [5], and CHT-bf [12, 11] that perform simulations. In addition, the AutoDomainMine estimation could be useful for intelligent tutoring systems in heat treatment [3].

The rest of this paper is organized as follows. Section 2 explains the basic AutoDomainMine technique and its use in computational estimation. Section 3 describes the applications of this estimation in the real world. Section 4 summarizes significant enhancements to AutoDomainMine. Section 5 gives the user evaluation of the software tool developed using the AutoDomainMine technique. It includes the evaluation before and after the enhancement. Section 6 outlines related work. Section 7 gives the conclusions and ongoing research.

## 2. Computational Estimation of Heat Transfer Curves with AutoDomainMine

### 2.1 Proposed Approach

The proposed AutoDomainMine approach [19] can be used to computationally estimate heat transfer curves with its two-step process of knowledge discovery and estimation, as described below. AutoDomainMine first discovers knowledge from experimental results by integrating clustering and classification. It then uses the discovered knowledge to estimate curves resulting from unperformed experiments given their input conditions.

Clustering is the process of placing a set of physical or abstract objects into groups of similar objects [9, 7]. Classification is a form of data analysis that can be used to extract models to predict categorical labels [22, 10]. These two data mining techniques are integrated in AutoDomainMine as explained here with illustration.

### 2.2 Knowledge Discovery in AutoDomainMine

The knowledge discovery step is depicted in Fig. 1. In discovering knowledge from existing experiments, clustering [9] is done over the heat transfer curves from actual experiments. Once the clusters of experiments are identified by grouping their heat transfer curves, a syntactic label is obtained for each cluster. The cluster labels form the classification target. Decision tree classification [10] is then used to learn the clustering criteria, namely, the input conditions that characterize each cluster. The decision tree helps understand the relative importance of the input conditions such as quenchant, part material, agitation etc.

The paths of each decision tree are then traced to build a representative pair of input conditions and heat transfer curve for each cluster. The decision trees and representative pairs form the discovered knowledge. Hence this step is referred to as the knowledge discovery step. This is a one-time step that is performed on the experimental data stored in the database to learn in AutoDomainMine.

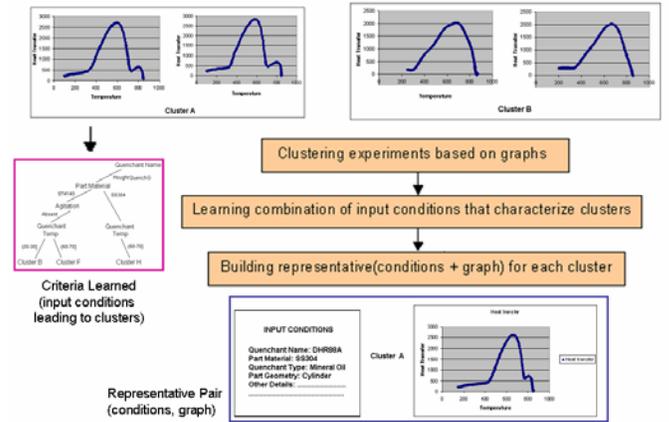


Figure 1: AutoDomainMine Knowledge Discovery Step

### 2.3 Estimation in AutoDomainMine

Estimation is performed using the discovered knowledge as shown in Fig. 2. This step is a recurrent one that has to be performed each time the user submits a new experiment to AutoDomainMine for estimation.

In order to estimate a heat transfer curve, given a new set of input conditions, the decision tree is searched to find the closest matching cluster. The representative heat transfer curve of that cluster is returned as the estimated curve for the given set of conditions. To estimate input conditions to obtain a given heat transfer curve, the representative curves are searched to find the closest match. The representative conditions corresponding to the match are conveyed as the estimated input conditions that would obtain the desired curve. Recall that the relative importance of the conditions has already been learned in the knowledge discovery step. The estimation thus takes into account such domain semantics [19].

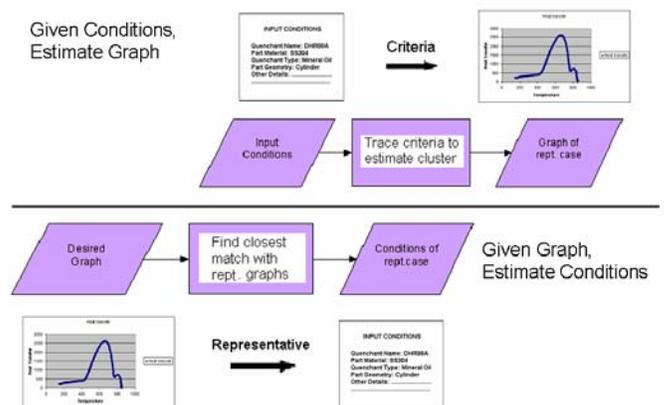


Figure 2: AutoDomainMine Estimation Step

### 2.4 Learning Analogous to Scientists

The proposed approach aims to automate a learning strategy of scientists [17]. They often group experiments based on the similarity of the resulting graphical plots, such as heat transfer curves. They then reason the causes of similarity between

groups in terms of the impact of the input conditions on the resulting plots. This is illustrated in Fig. 3. For example, the following facts were learned by Materials Scientists from the results of experiments [17].

- *Thin* oxide on a part surface causes vapor blanket around the part to break, resulting in *fast* cooling.
- *Thick* oxide on a part surface acts as an insulator, resulting in *slow* cooling.

The learning in these experiments was done by [17]:

- Performing quenching experiments with *thick* or *thin* oxide respectively, with all other conditions similar.
- Grouping based on heat transfer curves
- Reasoning based on quenching input conditions

This strategy is automated in our approach by integrating clustering and classification. Since a typical learning strategy of scientists is used in knowledge discovery, estimation obtained using this knowledge is expected to be more meaningful with respect to the domain. Thus the estimation accuracy is higher than with state-of-the-art techniques as confirmed by the evaluation.

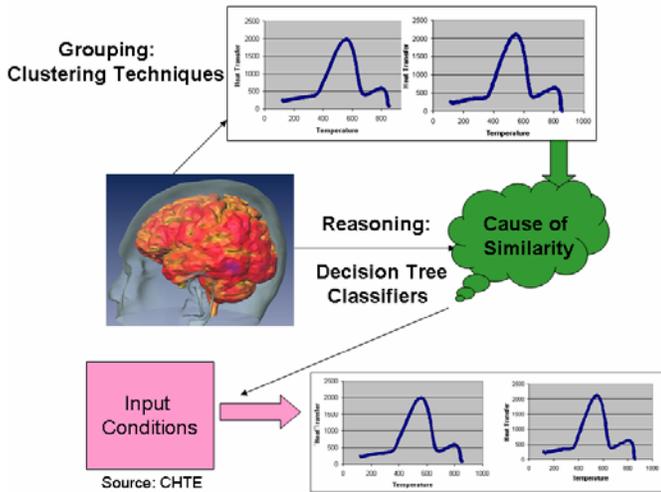


Figure 3: Automating Typical Learning Strategies of Scientists

### 3. Applications of AutoDomainMine

#### 3.1 QuenchMiner™

The main application of AutoDomainMine is to enhance the decision support functionality of our earlier system QuenchMiner™ [VTRWMS-03]. The decision support system of QuenchMiner™ discovers knowledge from existing experimental data stored in relational databases in plain text and numbers. It uses data mining techniques such as association rules [2] and applies the knowledge discovered to estimate parameters of interest in experiments given their input conditions. These parameters are typically ranges of cooling rates and heat transfer coefficients. They could even

be other parameters such as distortion tendencies. QuenchMiner™ also acts as a search engine for existing experimental data. It retrieves the heat transfer curves obtained from existing experiments stored in the database, based on the given queries. Thus it provides at-a-glance information for decision-support.

An example of the decision support functionality of QuenchMiner™ is illustrated here. In Fig. 5, the user input to a particular case is shown. The user submits information on quenching conditions, for the system to estimate the average heat transfer coefficients that would be obtained in the process. QuenchMiner™ estimates ranges of heat transfer coefficients in terms of categories such as high, medium and low. This is shown in Fig. 6.



Figure 5: Example of QuenchMiner Estimation Input



Figure 6: Example of QuenchMiner Estimation Output

However, QuenchMiner™ only *estimates ranges* of parameters such as cooling rates and heat transfer coefficients. In addition, it retrieves *existing experimental data*, on heat transfer curves. AutoDomainMine goes a step further than QuenchMiner™ by estimating the heat transfer curve that

would be obtained in an unperformed experiment, given its input conditions. This further assists the users by providing more information helpful in making decisions about the corresponding real processes. Hence the AutoDomainMine estimation enhances decision support in heat treating. This is elaborated with illustration in the user evaluation section.

### 3.2 DANTE

The DANTE software is a set of subroutines describing the thermal, mechanical and metallurgical response of steel to heating and cooling [6]. The term DANTE is an acronym for Distortion Analysis of Thermal Engineering. Using DANTE, simulations can be developed and executed to study the heat treatment of steel.

The DANTE software aids in predicting the following parameters in steel heat treatment [6]:

- Residual stress state
- Distortion of quenched parts
- Hardness profile
- Metallurgical phase fraction and distribution in carburized or through-hardened parts.

The software tool includes model building and meshing, assignment of boundary conditions, defining process variables, and generating finite element models. The material model used in the finite element analysis is a multiphase internal state variable model. In this the mechanical behavior of the composite structure is calculated from the behavior of each individual phase [6].

The DANTE software can use the boundary conditions from the heat transfer curves obtained with different quench conditions to perform simulations. Since the heat transfer curve is a plot of the heat transfer coefficients as a function of part temperature, it is useful for applications that need the values of these heat transfer coefficients at various temperatures. The AutoDomainMine estimation could thus be useful to a software tool such as DANTE. Since the heat transfer curve is estimated in AutoDomainMine given the input conditions of an experiment, the resulting estimate can be used to provide the boundary conditions. The accuracy of the estimation is likely to affect the quality of the simulations produced by DANTE. Hence it is desirable to use AutoDomainMine, since it is more accurate than state-of-the-art estimation techniques [19].

### 3.3 DEFORM-HT

The simulation software DEFORM, namely Design Environment for FORMing has Heat Treatment component called DEFORM-HT [16]. This component provides a finite element modeling system for simulating heat treatment processes. It predicts thermal, mechanical and metallurgical responses of parts during heat treatment, for example, heat treat distortion, quench cracking and residual stresses. DEFORM-HT can also provide information on phase transformation and phase volume fraction. Typical heat treatment processes in DEFORM-HT are [16]:

- Normalizing

- Quenching
- Austenizing
- Tempering
- Carburizing
- Aging
- Solution treatments
- Stress relieving

The DEFORM-HT software gives important information about process parameters needed for controlling and optimizing heat treatment. It can visualize microstructure, temperature and stress during heat treating that would be extremely difficult to do with laboratory experiments [16]. The estimation provided by AutoDomainMine could be useful for such simulations in DEFORM-HT. The estimation of heat transfer coefficients as a function of temperature could provide some of the variables needed for analysis and modeling for the simulation. The output of the AutoDomainMine system could thus be used as the input to tools such as DEFORM-HT.

### 3.4 SYSWELD

The SYSWELD tool [5] is built for simulating heat treatment, welding and welding assembly processes. It incorporates various aspects of material behavior, design and process. In the area of heat treatment in particular, SYSWELD provides the coupled modeling of complex physical phenomena such as [5]:

- Electromagnetism
- Heat transfer
- Diffusion and precipitation of chemical elements
- Phase transformation and mechanics

This software in performing simulations takes into account several factors such as processes, parameters, part geometry, thermal, metallurgical and mechanical material behavior [5]. This is where the AutoDomainMine estimation can be useful. It can provide some of the parameters needed in terms of heat transfer coefficients at various part temperatures. The AutoDomainMine estimation can also be used to infer the cooling rates at different temperatures, since the heat transfer coefficients are themselves derived from cooling rates. Since parameters such as cooling rates and heat transfer coefficients can be predicted by AutoDomainMine, this system could be useful in providing inputs to SYSWELD. Thus the application of AutoDomainMine can be extended to some extent in the welding area as well, in particular the overlapping area between welding and heat treatment.

### 3.5 CHTE Software Tools

Software tools such as CHT-bf and CHT-cf [12, 11] developed at CHTE, WPI could make use of the estimations produced by. CHT-bf stands for Center for Heat Treating – batch furnace. This tool can be used to simulate the heat treating of parts in a furnace. It contains a database for materials of the parts undergoing heat treatment. The database also includes furnace elements, furnaces, furnace atmospheres

and fuels. This data is helpful to the user to execute the software without explicitly defining all parameters.

CHT-bf has the ability to calculate important heat treatment terms. In particular, some of its features include the following [12, 11]:

- Calculating different heat losses from the furnace.
- Predicting the heat needed for the load under various conditions.
- Simulating the effect of using different fixtures.
- Plotting the heat stored in the furnace and load as a function of time.

CHT-cf is a tool developed to perform similar functions for continuous furnaces [12, 11]. These tools currently use QuenchMiner™ [VTRWMS-03] for some tasks. They use heat transfer coefficients obtained from performed experiments to perform further analysis. However, at present the only heat transfer coefficients provided by QuenchMiner™ are from experiments already performed. That is, QuenchMiner™ retrieves the heat transfer curve corresponding to specific input conditions submitted as search criteria, thus providing the required data.

AutoDomainMine provides heat transfer coefficients given input conditions of experiments not performed by estimating heat transfer curves given their input conditions. Thus it effectively increases the sample space of the experiments. Analysis can now be performed for a wider range of input conditions, even if the corresponding experiment was not conducted in the laboratory. Thus more inputs can be provided to tools such as CHT-bf and CHT-cf, hence providing more robust analysis. The effectiveness of this analysis depends on the accuracy of the AutoDomainMine estimation.

### 3.6 Quenching Processes in the Industry

The quenching conditions estimated by AutoDomainMine when a desired heat transfer curve is given, can be useful in selecting parameters for quenching processes in the industry. If the user needs a particular nature of heat transfer characterized by a given heat transfer curve, then the AutoDomainMine system can estimate the quenching conditions needed. The user can experiment with the system until a satisfactory nature of heat transfer is achieved and the corresponding input conditions are predicted. This saves the cost of performing a laboratory experiment to determine the process parameters needed for real quenching in the industry. Note that addition to augmenting the QuenchMiner™ functionality, this application of AutoDomainMine can be used for other systems requiring computational estimation for parameter selection.

### 3.7 Intelligent Tutoring Systems

AutoDomainMine is likely to be useful tutoring systems in heat treatment and related areas, designed using artificial intelligence. For example, the AutoDomainMine estimation could be used to extend the functionality of computerized tutors such as the Computer Coach project in the University of Amsterdam [3]. The original goal of this project was to serve

as a coach for thermodynamics. However, this fragmented into a number of research projects on different aspects of heat and temperature.

A system such as AutoDomainMine could be useful in further extending the scope of such a project in understanding heat treatment of materials. The heat transfer curves estimated using various input conditions, and also the conditions estimated for obtaining heat transfer curves, can help students understand heat treatment processes. The relative importance of process parameters, the nature of heat treatment under different sets of input conditions and the impact of altering specific conditions on the resulting curve can all be observed. This can be used to draw useful inferences that corroborate and possibly augment theoretical knowledge in the domain.

### 3.8 Applications in Related Domains

AutoDomainMine can also be applied to other related domains with experimental results characterized by 2-dimensional plots denoting the functional behavior of process parameters. Although it is developed primarily for the heat treating domain, it is also likely to provide useful estimations in other areas of Materials Science. In general, it targets domains with scientific processes, e.g., Mechanical Engineering, Physics, Chemistry and so forth. In such domains performing a laboratory experiment consumes significant time and resources, hence motivating computational estimation.

## 4. Enhancements to AutoDomainMine

### 4.1 Need for Enhancement

In mining heat transfer curves with techniques such as clustering [9] the default measure for comparison is typically Euclidean distance [7]. However, using Euclidean distance for comparison often does not result in the desired outcome. See for example the case depicted in Fig. 7.

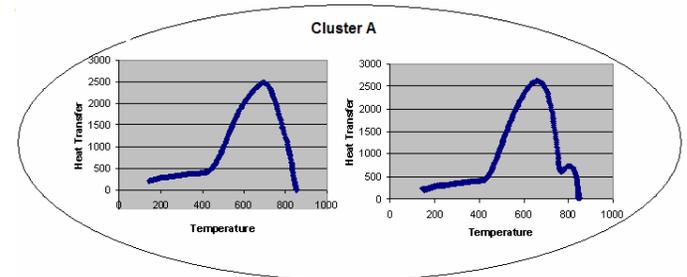


Figure 7: Example of Inaccuracy in Clustering

Euclidean distance is based on the absolute position of points in space. It is the “as-the-crow-flies” distance between the given objects [7]. Thus, if a data mining algorithm using Euclidean distance considers the two heat transfer curves shown in Fig. 7 as similar, relative to other curves, this is semantically inaccurate. The two curves depict distinctly different physical tendencies since one has a visible

Leidenfrost point while the other does not. This is confirmed by domain experts [17].

The problem is that the knowledge of various features on the curves and their relative importance may at best be available in a subjective form, but not as a metric. Though several distance metrics exist in the literature, it is often not known beforehand which of these work best in a given domain. This motivates the development of a technique to learn a distance metric that captures the semantic content of the heat transfer curves. This should enhance the clustering accuracy by addressing issues such as the domain-specific emphasis of features depicted in Fig 7. This is likely to enhance the AutoDomainMine estimation.

#### 4.2 Distance Metric Learning to Enhance Clustering

We propose a technique called LearnMet [20] to learn a distance metric for graphical plots incorporating domain semantics. We now summarize this technique in the context of the heat treating domain. Details of LearnMet can be found in [20]. The input to LearnMet is a training set with correct clusters of heat transfer curves over a subset of the experimental data in the database. The training set is provided by domain experts.

The five basic steps of LearnMet are [20]:

- (1) Guess the initial metric guided by domain knowledge.
- (2) Use that metric for clustering with an arbitrary but fixed clustering algorithm.
- (3) Evaluate the accuracy of the obtained clusters by comparing them with the correct clusters. Accuracy is measured as success rate, i.e., ratio of true positives plus true negatives over sum of true positives, true negatives, false, positives and false negatives [22].
- (4) Adjust the metric based on the error between the obtained and correct clusters. If error is below threshold or if the execution times out then terminate and go to step (5), else go to step (2).
- (5) Once terminated, output the metric giving error below threshold or minimum error so far, as the learned metric.

These steps are depicted in the flowchart in Fig. 8.

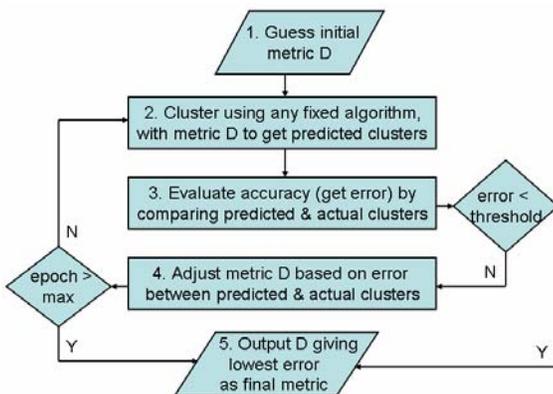


Figure 8: The LearnMet Technique

#### 4.3 Integration of Technique with AutoDomainMine

LearnMet is integrated with the AutoDomainMine system to learn domain-specific distance metrics for clustering. In order to evaluate LearnMet as a part of AutoDomainMine, the following approach is used.

The clusters obtained using the default Euclidean distance and those using the learned metrics are both sent to the same decision tree classifier such as ID3 or J4.8 [10]. The accuracies of the respective classifiers in predicting unseen data in the database are compared. Since the classifiers remain the same, the difference in prediction accuracy represents the difference in clustering accuracy. Thus it denotes the effectiveness of the learned distance metrics in clustering [20]. This is elaborated in the next section.

### 5. User Evaluation

#### 5.1 Evaluation of the AutoDomainMine System

The performance of the system built using the AutoDomainMine technique [19] is evaluated with respect to the targeted users. Thorough evaluation is conducted using experimental data in heat treating. The process of evaluation is as follows.

Test sets of performed experiments used for evaluation are distinct from the training sets used for learning in AutoDomainMine. This is to evaluate whether a generic hypothesis has been learned from the existing experiments to serve as the basis for estimation in unperformed experiments. The opinion of the domain experts is taken into account during evaluation. Examples from the evaluation are presented here.

In the example shown in Fig. 9, the user submits input conditions of a quenching experiment in order to have the system estimate the heat transfer coefficients. The estimation of heat transfer coefficient as a function of temperature, i.e., the heat transfer coefficient curve is shown in Fig. 10.

Figure 9: Given Conditions, Estimate Heat Transfer Curve

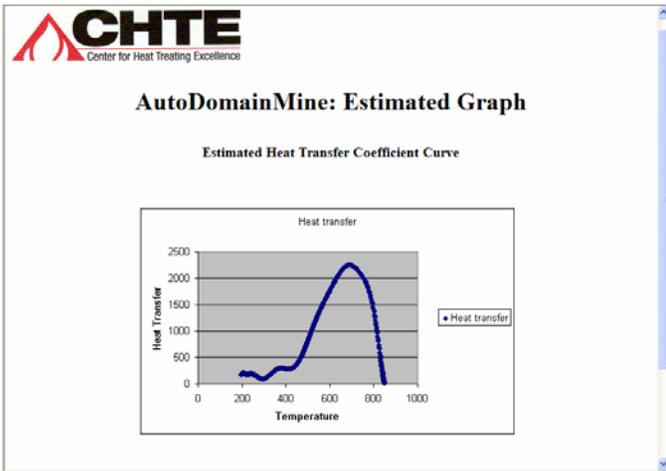


Figure 10: Estimated Heat Transfer Curve

The estimated curve is compared with the heat transfer curve obtained in the laboratory experiment performed with the same input conditions. This actual data has been stored in the test set. On comparing the two curves, the domain experts conclude that this estimation is satisfactory.

Note that AutoDomainMine estimates the heat transfer curve as a whole. This is a step ahead of QuenchMiner™ which only estimates ranges of heat transfer likely to be achieved in experiments, such as high, medium and low. Thus AutoDomainMine enhances decision support by providing more information with greater precision.

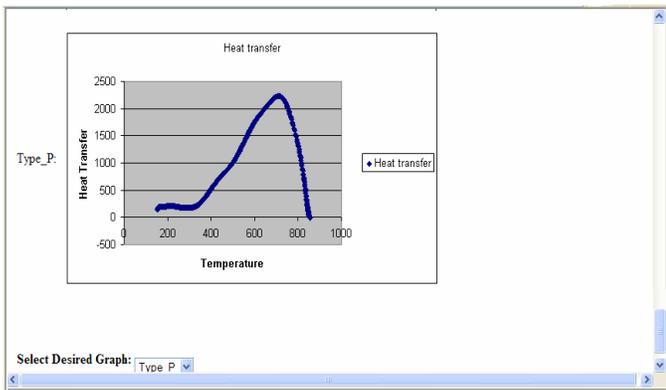


Figure 11: Given Heat Transfer Curve, Estimate Conditions

In another example shown in Fig. 11, the user submits a given heat transfer curve representing a desired nature of heat transfer. The AutoDomainMine system is requested to estimate a set of conditions that would obtain such a heat transfer curve.

The response of AutoDomainMine in terms of an estimated set of conditions to obtain the desired curve is shown in Fig 12. On comparing this with the set of conditions that yield such a heat transfer curve (as stored in the test set), it is found that the estimation given by AutoDomainMine is accurate. This is confirmed by domain experts [17].

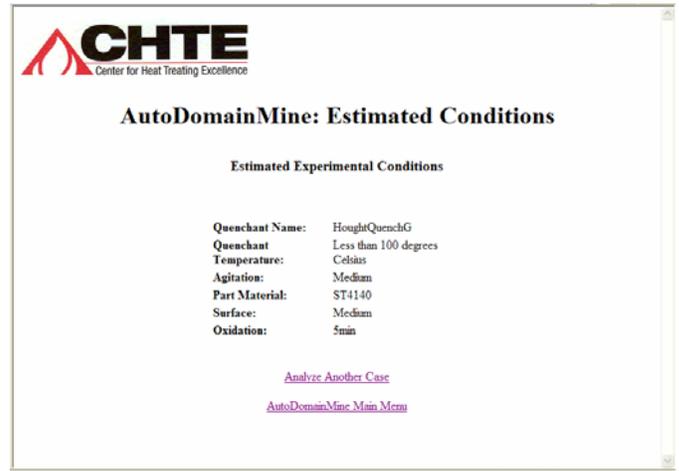


Figure 12: Estimated Conditions

Likewise on evaluating the AutoDomainMine system with over 200 experiments, it is found that the estimation accuracy is approximately 75%. The response time of the system is on an average 2 seconds [19]. Thus accuracy and efficiency are both within acceptable limits as concluded by domain experts [17].

On comparing the AutoDomainMine estimation with that provided by similarity search [7, 22] it is found that AutoDomainMine is more accurate and efficient [19]. Similarity search needs an average response time of 4 seconds and gives an accuracy of approximately 60%.

AutoDomainMine is faster because it searches over decision trees and representatives as opposed to similarity search that searches the entire database of existing experiments. AutoDomainMine is also more accurate since it incorporates domain knowledge discovered from existing experimental results while performing the estimation.

## 5.2 Evaluation of LearnMet with AutoDomainMine

LearnMet [20] has been developed mainly for the AutoDomainMine system to learn domain-specific distance metrics for clustering. The learned distance metrics can thus be used as the notion of distance in clustering heat transfer curves. Hence LearnMet is evaluated by measuring the accuracy of the AutoDomainMine estimation with and without the metrics learned from LearnMet.

The estimation accuracy obtained using different metrics for clustering in AutoDomainMine is shown in Fig. 13. Note that this denotes the accuracy of the decision tree classifier in estimating the clusters of unperformed experiments over test data. This is measured using 4-fold cross validation [22]. The observations shown here depict a summary of the rigorous experimental evaluation of LearnMet with AutoDomainMine,

The estimation obtained from clustering using the learned metrics DE1, DE2, EQU, RND1 and RND2 is compared with the estimation obtained from clustering using Euclidean distance (ED). The metrics DE1 and DE1 correspond to

experiments conducted with initial metrics given by domain experts. EQU denotes the experiment with the initial metric selected by assigning equal weights to all components. RND1 and RND2 denote the experiments with initial metrics obtained by assigning random weights to components. The corresponding learned metrics are all used for clustering in AutoDomainMine and the resulting estimation accuracies are compared with each other.

Another distance metric for performing the comparison is  $D = D_{Euclidean} + D_{LF} + D_{BP}$  which is referred to as the AutoDomainMine metric denoted as ADM [19, 20]. This metric is obtained by considering Euclidean distance over the curve as a whole, and in addition considering the critical distances corresponding to the Leidenfrost point and Boiling point. Equal weight is given to all these individual distance components in the AutoDomainMine metric.

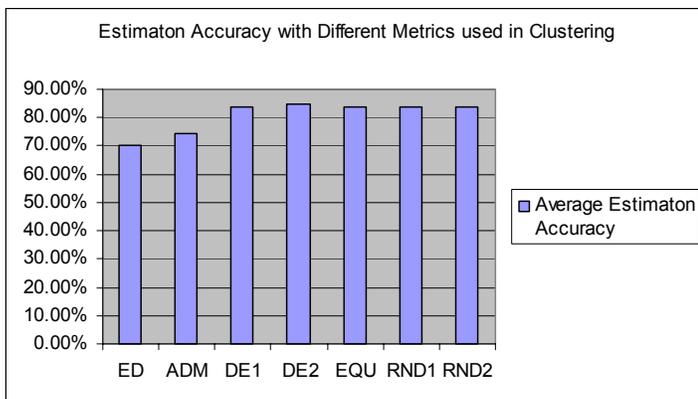


Figure 13: Estimation Accuracy in AutoDomainMine

It is observed from Fig. 13 that the accuracy with each metric output from LearnMet is higher than that with Euclidean distance. Moreover, it is also noticed that the accuracies with the learned metrics are higher than the accuracy with the AutoDomainMine metric. This proves the effectiveness of the learned metrics in capturing domain semantics during clustering in AutoDomainMine,

## 6. Related Work

### 6.1 Mathematical Modeling

In several scientific domains, mathematical modeling [14, 4] could possibly be used as estimation technique. This requires a precise representation of the concerned graphical plots in terms of numerical equations, and exact knowledge of how the inputs affect the outputs.

However, precise numerical equations and / or variables in existing models are often not known with respect to heat treating data. Hence mathematical modeling is not a popular technique in estimating heat transfer curves. For example, it has been shown that the existing models do not adequately work for multiphase heat transfer with nucleate boiling [17].

### 6.2 Naïve and Weighted Similarity Search

One intuitive estimation approach is a naive similarity search over existing data [7]. The given input conditions of a user-submitted experiment are compared with those of existing experiments to select the closest match as the number of matching conditions. However this poses the following problem. The non-matching condition(s) could be significant in the given domain. The resulting estimation could thus be inaccurate since this significance has not been incorporated.

A weighted search [22] guided by basic domain knowledge could possibly be used to overcome this problem. The relative importance of the search criteria, i.e., input conditions are coded as weights into feature vectors. The closest match is the weighted sum of the matching conditions. However these weights are not likely to be known with respect to their impact on the graph. Domain experts may at best have a subjective notion about the relative importance of a few conditions [SMM18], which is not sufficient for a weighted similarity search.

### 6.3 Case-based Reasoning Approaches

Estimation can also be performed using Case Based Reasoning (CBR) [8]. In the heat treating domain, this involves comparing conditions to retrieve the closest matching experiment, reusing its heat transfer curve as a possible estimate, performing adaptation if needed, and retaining the adapted case for future use.

However adaptation approaches in the literature [8, 13] are not feasible for us. For example, if the condition “agitation” in the new case has a higher value than in the retrieved case, then a domain-specific adaptation rule could be used to infer that high agitation implies high heat transfer coefficients. However, this is not sufficient to plot a heat transfer curve in the new case. Adaptation rules [8] are generally used when the case solution is categorical such as in medicine and law.

Case Based Adaptation [13] could possibly be applied here. In the example where agitation in the new case and retrieved case do not match, the case base could be searched to find another case that matches in terms of agitation. However this second retrieved case may not match another condition such as “Quenchant temperature”. The heat transfer curves of the two retrieved cases could then be used to build an average that forms the estimated graph in the new case. However, building such an average requires prior knowledge of the relative importance of conditions and the significant features on heat transfer curves.

Moreover, adaptation with any approach requires a computational expense for each estimation performed, which is inefficient [18]. CBR approaches without adaptation such as exemplar reasoning [1] and instance-based reasoning [15] if used in our context face the same problem as naïve and weighted similarity search respectively [18].

## 7. Conclusions and Ongoing Research

The AutoDomainMine approach has been proposed for computational estimation by integrating the data mining techniques of clustering and classification. This paper describes AutoDomainMine with specific reference to quenching heat treatment. AutoDomainMine discovers knowledge from results of existing heat treating experiments in order to predict the results of new experiments. AutoDomainMine is used to estimate the *heat transfer curve* that would be obtained in an experiment given its *input conditions* and vice versa.

The main application of AutoDomainMine is to enhance decision support in our earlier developed system QuenchMiner™. While QuenchMiner™ only estimates ranges of parameters given their input conditions, AutoDomainMine estimates the heat transfer curve, thus enhancing decision support. Thorough user evaluation has proved that the AutoDomainMine estimation is satisfactory for various applications in heat treating.

Enhancement to AutoDomainMine includes *learning* a domain-specific distance metric as the notion of similarity in clustering. This has also been rigorously evaluated and is found effective in improving clustering and hence estimation accuracy. Ongoing research includes *designing* (as opposed to selecting) domain-specific representatives for classification. This is likely to provide even better estimation.

## Acknowledgements

This work is supported by the Center for Heat Treating Excellence (CHTE) and its member companies, and by the Department of Energy - Industrial Technology Program (DOE-OIT) Award Number DE-FC-07-01ID14197. The feedback of all the CHTE Members and also of the Research Teams in the Materials Science Program at WPI is gratefully acknowledged.

The authors convey special thanks to two Research Assistants, Ms. Shuhui Ma and Ms. Olga Karabelchtchikova. This is for being volunteers at the poster session to present this work at the ASM Heat Treating Society conference in Pittsburgh, PA in September 2005. The inputs in terms of useful sources of reference given by Research Assistants Radhakrishnan Purushothaman and Virendra Warke for CHT-bf and DANTE respectively are also acknowledged.

## References

[1] A. Aamodt and E. Plaza, *Case-based reasoning: Foundational Issues, Methodological Variations and System Approaches*, In AICom, 2003, 7(1): pp 39 – 59.

[2] R. Agrawal, T. Imlinski, and A. Swami, *Mining Association Rules between Sets of Items in Large Databases*, In Proceedings of the ACM SIGMOD International Conference on Management of Data, Washington D.C., USA, May 1993, pp. 207-216.

[3] D. Bierman and P. Kamsteeg, *Elicitation of Knowledge with and for Intelligent Tutoring Systems*, Technical Report, University of Amsterdam, 1988.

[4] H. Boyer and P. Cary, *Quenching and Control of Distortion*, ASM International, 1989.

[5] ESI Group, *SYSWELD: Engineering Simulation Solution for Heat Treatment, Welding and Welding Assembly*, France, 2003.

[6] Furguson B., Petrus G. and Pattock T., *A Software Tool to Simulate the Quenching of Alloy Steel*, In Proceedings of the 3<sup>rd</sup> International Conference on Quenching and Control of Distortion, 1999, pp. 188 – 200.

[7] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, 2001.

[8] J. Koldner, *Case-Based Reasoning*, Morgan Kaufmann Publishers, 1993.

[9] L. Kaufman and P. Rouseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley, 1990.

[10] H. Kim and G. Koehler, *Theory and Practice of Decision Tree Induction*, In Omega, 1995, 23(6): pp 637 - 652.

[11] J. Kang and Y. Rong, *Numerical Simulation of Heat Transfer in Loaded Heat Treatment Furnaces*, In 2<sup>nd</sup> International Conference on Thermal Process Modeling and Simulation”, France, March 31 – April 3, 2003.

[12] Q. Lu, R. Vader, J. Kang and Y. Rong, *Development of a Computer-Aided Heat Treatment Planning System*, In Heat Treatment of Metals, 2002, 3: pp 65-70.

[13] D. Leake, “Case-Based Reasoning: Experiences, Lessons and Future Directions”, AAAI Press, 1996.

[14] A. Mills, *Heat and Mass Transfer*, Richard Irwin Inc., 1995.

[15] T. Mitchell, *Machine Learning*, WCB McGraw Hill, 1997.

[16] Scientific Forming Technologies Corporation. *DEFORM-HT*, Columbus, OH, USA, Feb 2005.

[17] R. Sisson Jr., M. Maniruzzaman and S. Ma, *Quenching: Understanding, Controlling and Optimizing the Process*, In CHTE Seminar, Columbus, OH, USA, Nov 04.

[18] A. Varde, *Automating Domain-Type-Dependent Data Mining as a Computational Estimation Technique for Decision Support in Materials Science*, Ph.D. Dissertation Proposal, WPI, MA, USA, Sep 04.

[19] Varde, A., Rundensteiner, E., Ruiz, C., Maniruzzaman, M., and Sisson Jr., R., *Data Mining over Graphical Results of Experiments with Domain Semantics*, In Proceedings of ACM SIGART's 2nd International Conference on Intelligent Computing and Information Systems ICICI16. Cairo, Egypt, Mar 2005, pp 603 – 611.

[20] Varde, A., Rundensteiner, E., Ruiz, C., Maniruzzaman, M., and Sisson Jr., R., *Learning Semantics-Preserving Distance Metrics for Clustering Graphical Data*, In Proceedings of ACM SIGKDD's MDM-05. Chicago, IL, USA, Aug 2005.

[21] A. Varde, M. Takahashi, E. Rundensteiner, M. Ward, M. Maniruzzaman and R. Sisson Jr., *Apriori Algorithm and Game-of-Life for Predictive Analysis in Materials Science*, In KES Journal, IOS Press, Netherlands, 2004, 8(4): pp 213 – 228.

[22] I. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann Publishers, 2000.