CHARACTERIZATION AND PREDICTION OF WELDING DROPLET RELEASE USING TIME SERIES DATA MINING

RICHARD J. POVINELLI AND XIN FENG
Department of Electrical and Computer Engineering
Marquette University, P.O. Box 1881, Milwaukee, WI 53201-1881, USA
E-mail: Richard.Povinelli@Marquette.edu
Web: http://povinelli.eece.mu.edu
Phone: 414.288.7088

ABSTRACT
This paper presents the results from characterizing and predicting the release of droplets of metal from a welder. The welding process joins two pieces of metal into one by making a joint between them. An arcing current melts the tip of a wire, forming a metal droplet that elongates until it releases. The goal is to predict the moment when a droplet will release, which can improve the quality of the joint by allowing the droplet releases to be monitored and controlled. Because of the irregular, chaotic, and event nature of the droplet release, prediction is impossible using traditional time series methods. Using Time Series Data Mining techniques allows the droplet releases to be predicted with a high degree of accuracy.

The Time Series Data Mining (TSDM) framework (Povinelli 1999; Povinelli and Feng 1998; Povinelli and Feng 1999) is applied to the prediction of welding droplet releases. Methods based on the TSDM framework are able to successfully characterize and predict complex, nonperiodic, irregular, and chaotic time series such as the release of metal droplets from a welder.

This paper, which is divided into three sections, presents the results of applying the TSDM framework to this problem. The first section discusses the welding problem[1]. The second section reviews the key TSDM concepts and an extension of the TSDM framework to multiple temporal patterns. The third section presents the prediction results.

PROBLEM STATEMENT

Welding joins two pieces of metal by forming a joint between them. As illustrated in Figure 1, a current arc is created between the welder and the metal to be joined. Wire is pushed out of the welder. The tip of the wire melts, forming a metal droplet that elongates (sticks out) until it releases. Predicting when a droplet of metal will release from a welder allows the quality of the metal joint to be monitored and controlled. The problem is to predict the releases $Y = \{y_t, t = 1, \ldots N\}$, where t is a time index, and N is the number of observations, using the stickout $X = \{x_t, t = 1, \ldots N\}$ time series. This time series is a

[1] Drs. C. Tolle, E. Larsen, D. Pace, and D. Iosty of INEEL gathered the data used in this paper. Their work was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Division of Materials and Materials Engineering, under DOE Idaho Operations Office Contract DE-AC07-94ID13223.

measure of the droplet elongation. Because of the irregular, chaotic, and noisy nature of the droplet release, prediction is impossible using traditional time series methods.

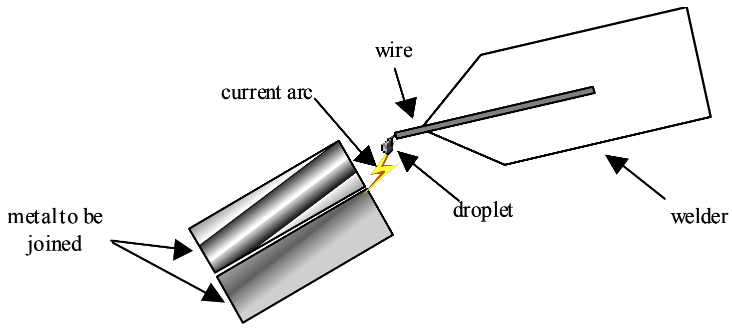

Figure 1 – Welding Process

A sample of the stickout time series is illustrated in Figure 2. An electronic camera on the welding station measures the droplet stickout in pixels. It is sampled at 1kHz and comprised of approximately 5,000 observations. The release time series, also illustrated in Figure 2, indicates the release of a droplet (event) with a one and a non-release (non-event) with a zero. It is synchronized with the stickout time series.

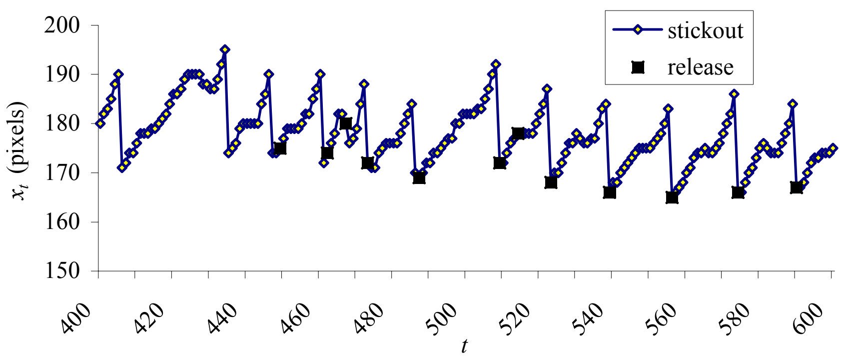

Figure 2 – Welding Stickout and Release Time Series

TIME SERIES DATA MINING METHOD

Previous work (Povinelli 1999; Povinelli and Feng 1998; Povinelli and Feng 1999) presented the TSDM framework. Here the TSDM method for identifying multiple temporal pattern clusters is discussed.

The TSDM method discussed here discovers hidden temporal patterns (vectors of length Q) characteristic of events (important occurrences) by time-delay embedding (Abarbanel 1996; Iwanski and Bradley 1998; Tolle 1997; Tolle and Gundersen 1998) an observed time series X into a reconstructed phase space, here simply called phase space. An event characterization function g is used to represent the eventness of a temporal pattern. An augmented phase space is formed by extending the phase space with g. The augmented phase space is searched for a collection $\mathcal{C}$ of temporal pattern clusters P that best characterizes the desired events. The temporal pattern clusters are then used to predict events in a testing time series. Figure 3 presents a block diagram of the TSDM method.

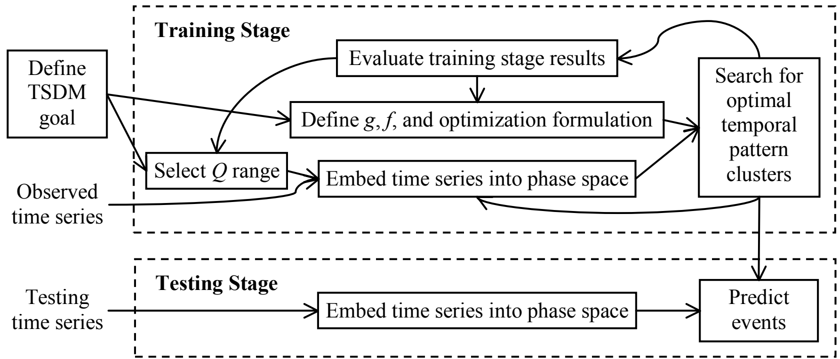

Figure 3 – Block Diagram of TSDM Method

Given a TSDM goal (predict droplet releases), observed time series X with events Y to be characterized, and a testing time series $Z = \{x_t, t = R, \ldots, S\}$ $N < R < S$ with events $W = \{y_t, t = R, \ldots, S\}$ to be predicted, the steps in the TSDM method are:

I. Training Stage (Batch Process)

1. Frame the TSDM goal in terms of the event characterization function, objective function, and optimization formulation.
   a. Define the event characterization function, g. For this problem $g(t) = y_{t+1}$, which allows characterization of droplet releases one time step in advance.
   b. Define the objective function, f_1. For this problem $f_1(\mathcal{C}) = (t_p + t_n)/(t_p + t_n + f_p + f_n)$, which has an optimal value when every event is correctly predicted. The values $t_p, t_n, f_p, \text{and} f_n$ are described in Table 1.

Table 1 – Event Categorization

	Actually an event	Actually a non-event
Categorized as an event	True positive, t_p	False positive, f_p
Categorized as a non-event	False negative, f_n	True negative, t_n

   c. Define the criteria to accept a temporal pattern cluster. For this problem $f_2(P) = t_p/(t_p + f_p)$, called the positive accuracy, must be greater than a threshold. It defines how well a temporal pattern cluster is at avoiding false positives.
   d. Define the optimization formulation. The optimization formulation for the whole training stage is $\max f(\mathcal{C})$ subject to $\min c(\mathcal{C})$. The optimization formulation for the intermediate steps is $\max f(P)$.
2. Determine the range of Q's, i.e., the dimensions of the phase spaces and the length of the temporal patterns.

3. Embed the observed time series into the phase space using the time-delayed embedding process.
4. Associate with each time index in the phase space an eventness represented by the event characterization function. Form the augmented phase space.
5. Search, using a modified simple GA (Povinelli 2000), for the optimal collection of temporal pattern clusters in the augmented phase space using the following algorithm.
    a. if the temporal pattern cluster meets the criteria set in 1.c then, repeat step 5 after removing the clustered phase space points from the phase space.
    b. elseif the range of Q is not exceeded, increment Q and goto step 2.
    c. else goto step 6.
6. Evaluate training stage results. Repeat training stage as necessary.

II. Testing Stage (Real Time or Batch Process)
1. Embed the testing time series into the phase spaces.
2. Apply the collection of temporal pattern clusters to predict events.
3. Evaluate testing stage results.

The search in step I.5

WELDING APPLICATIONS

The stickout time series X consists of the 4,985 equally sampled observations, from t = 175 through 5,159. The observed stickout time series consists of t = 175 through 2,666, while the testing time series consists of t = 2,667 through 5,159. Figure 4 illustrates both the observed and testing stickout time series, while the previously shown Figure 2 provides a detailed view of a sample of the time series.

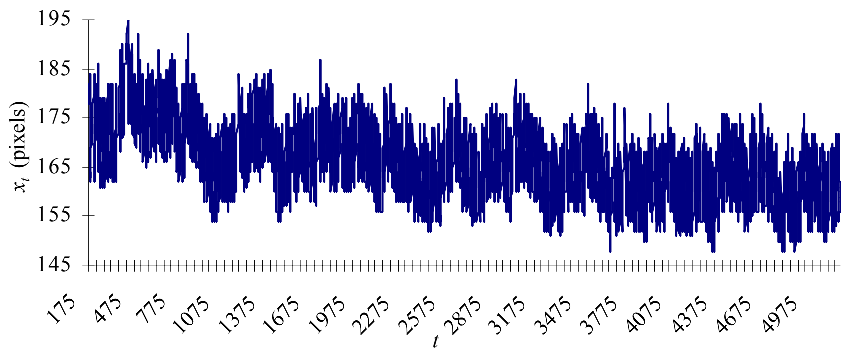

Figure 4 – Stickout Time Series

Besides the obvious nonperiodic oscillations, the stickout time series exhibits a large-scale trend. Removing the trend helps the method find the necessary temporal patterns. A first difference filter could be applied, but that would introduce a synchronization problem between the release and stickout time series. Instead, a recalibration rule is used to remove the trend: when there is a 10-pixel drop between two consecutive observations, the second observation is recalibrated to zero.

Figure 5 presents an illustrative augmented phase space. The events are inseparable from the non-events using a two-dimensional phase space. Hence, the TSDM method, which finds multiple temporal clusters of varying dimensionality, is applied.

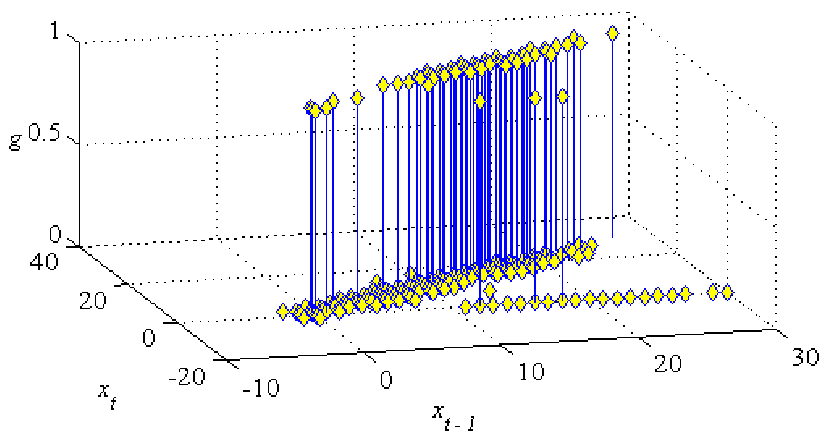

Figure 5 – Stickout and Release Augmented Phase Space (Observed)

The augmented phase space is searched using a tournament genetic algorithm. Two parameter sets are used for the genetic algorithm. For both sets, the population size was 30; the elite count was one; the gene length was eight; and the tournament size was two. The first parameter set had a mutation rate of 0% and convergence criterion of 0.65. The second set had mutation rate of 0.05% and a convergence criterion of 0.50. The results of the search are shown in Table 2. In the observed time series there are 154 droplet releases and 2,491 total observations.

Table 2 – Recalibrated Stickout and Release Results (Observed)

	Actually an event	Actually a non-event
Categorized as an event	$t_p = 101$	$f_p = 41$
Categorized as a non-event	$f_p = 53$	$t_p = 2303$

Fourteen temporal pattern clusters form the temporal pattern cluster collection employed to identify events. This collection contains temporal pattern clusters that vary in dimension from 1 to 14. The release observations are correctly characterized 96.23% overall and events are correctly categorized as events 71.13%.

The testing time series is shown in Figure 4. The test time series is recalibrated using the same process as for the observed time series.

The results of applying the temporal pattern cluster collection to the testing time series are seen in Table 3. As with the training stage results, the testing stage results have high prediction accuracy of 96.43% and positive accuracy of 73.53%. These results are better than those found in the training phase.

Table 3 – Recalibrated Stickout and Release Results (Testing)

	Actually an event	Actually a non-event
Categorized as an event	$t_p = 100$	$f_p = 36$
Categorized as a non-event	$f_p = 53$	$t_p = 2296$

Reviewing the stickout time series in Figure 4, it can be seen that these results are quite good. The results from this paper can be applied to improving welds in two ways. The first is by being able to tell how many droplets have been laid down on the welding seam. In this case, the accuracy increases since the false positives almost evenly balance out the false negatives. Such a system would report that 136 droplets have been released vs. the actual 153 droplets released. The second mechanism would be by using the prediction as an input to a control system.

Future work will involve improving the accuracy by using a system identification approach. This approach will identify rather than predict droplet releases. In addition, two other data sources are available – current and voltage – to improve accuracy. Alternative eventness functions also can be employed to improve accuracy. One example is an event function that characterizes all events one to five time steps ahead instead of in just one time step ahead.

To conclude, using the droplet stickout as measured by an electronic camera the releases of welding droplets was predicted with 96.34% total prediction accuracy and 73.53% positive prediction accuracy. These results show that the TSDM method can be used in a system to control and monitor the welding seam thereby improving the quality of the weld.

REFERENCES

Abarbanel, H. D. I. (1996). *Analysis of observed chaotic data*, Springer, New York.

Iwanski, J., and Bradley, E. (1998). "Recurrence plot analysis: To embed or not to embed?" *Chaos*, 8(4), 861-871.

Povinelli, R. J. (1999). "Time Series Data Mining: Identifying Temporal Patterns for Characterization and Prediction of Time Series Events," Ph.D. Dissertation, Marquette University, Milwaukee.

Povinelli, R. J. (2000). "Improving Computational Performance of Genetic Algorithms: A Comparison of Techniques." *Genetic and Evolutionary Computation Conference (GECCO-2000) Late Breaking Papers*, Las Vegas, Nevada, 297-302.

Povinelli, R. J., and Feng, X. (1998). "Temporal Pattern Identification of Time Series Data using Pattern Wavelets and Genetic Algorithms." *Artificial Neural Networks in Engineering*, St. Louis, Missouri, 691-696.

Povinelli, R. J., and Feng, X. (1999). "Data Mining of Multiple Nonstationary Time Series." *Artificial Neural Networks in Engineering*, St. Louis, Missouri, 511-516.

Tolle, C. R. (1997). "Time-Based Clustering and Its Application to Determining a Signal's Motivation: Deterministic Chaos or Random Disturbance," Ph.D. Dissertation, Utah State University, Logan, Utah.

Tolle, C. R., and Gundersen, R. W. (1998). "Searching for Determinism in Erratic Signals: Fuzzy Time-Based Clustering of Delay Vectors." *Artificial Neural Networks in Engineering*, St. Louis, Missouri, 43-48.