PREDICTIONS IN TIME SERIES WITH REPEATED PATTERNS, USING PIECEWISE LINEAR REGRESSION

Arash Niknafs, Bo Sun, Michael M. Richter, Günther Ruhe

REPORT NUMBER 094/2011

SOFTWARE ENGINEERING DECISION SUPPORT LABORATORY
UNIVERSITY OF CALGARY

April, 2011
# Table of Contents

1. Abstract .................................................................................................................. 3

2. Introduction .............................................................................................................. 3

3. RELATED WORK .................................................................................................... 4

4. BACKGROUND ...................................................................................................... 4

   4.1 Piecewise Linear Regression (PLR) .................................................................. 5
   4.2 Support Vector Machine (SVM) .......................................................................... 5
   4.3 Artificial Neural Networks (ANN) ....................................................................... 6

5. PROBLEM STATEMENT .......................................................................................... 7

6. SOLUTION APPROACH ......................................................................................... 7

   6.1 Dataset .............................................................................................................. 7
   6.2 Data pre-processing ............................................................................................ 8
   6.3 Split points detection .......................................................................................... 8
   6.4 Performance measures ...................................................................................... 10

7. EVALUATION (CASE STUDY) ................................................................................ 10

   7.1 Applying PLR .................................................................................................... 10
   7.2 Experimental results .......................................................................................... 11

8. CONCLUSION .......................................................................................................... 12

9. Acknowledgements .................................................................................................. 13

10. References ............................................................................................................. 13
1. ABSTRACT

Modelling nonlinear patterns is possible through using regression (curve fitting) methods. However, they can be modelled by linear regression (LR) methods, too. This kind of modelling is usually used to depict and study trends and it is not used for prediction purposes. Our goal is to study the applicability and accuracy of piecewise linear regression in predicting a target variable in different time spans (where a pattern is being repeated).

Using Moving Average, we identified the split points and then tested our approach on a real world case study. The dataset of the amount of recycling material in Blue Carts in Calgary (including more than 31,000 records) was taken as a case study for evaluating the performance of the proposed approach. Root mean square error (RMSE) and Spearman rho were used to evaluate and prove the applicability of this prediction approach and evaluate its performance. A comparison between the performances of Support Vector Machine (SVM), Artificial Neural Networks (ANN), and the proposed LR-based prediction approach is also presented. The results show that the proposed approach works very well for such prediction purposes. It outperforms SVM and is a powerful competitor for ANN.

2. INTRODUCTION

Parameterized models such as regression models, used to explain time-series, vary over time. This means that they are not constant models. In many regression problems where the functional relationship between the independent variable and the target variable changes over time, we cannot fit one uniform regression function to that data because the functional relationship between those variables changes at certain points of the domain (Kuchenhof, 1996); these points are usually referred to as split points.

Regression can be performed with one of the following goals in mind (Cherkassky and Lari-Najafi, 1991):

- Data reduction: The linear function which is an “easily storable analytical expression” represents a large number of data points.
- Data interpolation: An approximation for a number of data points by means of an analytic expression.
- Data prediction or estimation of an underlying natural law: An accurate estimate for a finite number of possibly inaccurate data points.

Linear regression (LR) is one of the regression models but a single LR model is not appropriate for long-term forecast of highly varying and deviant time series because it assumes stable increase or decrease of the target variable over time. This is why the family of LR models are usually used for data interpolation and data reduction. For example piecewise linear regression (PLR) has been used in many works such as (Loesch et al., 2006) for modeling, observation and then description of medical facts.

PLR is generally used in applications where the modeling of the target variable by a single line is too inaccurate and not representative. Such variables are a function of different inputs and follow a non-linear pattern. In this paper we intend to study the applicability of PLR in prediction problems where such non-linear patterns are being repeated periodically over time.

The structure of the rest of the paper is as follows. In Section 2 we talk about the related work. Section 3 discusses the background and methods used. The problem statement and the proposed solution are illustrated in section 4 and section 5, respectively. Then the solution is evaluated in a case study in section 6 –where the experimental results are also presented. Finally, a few conclusions are drawn in section 7.
3. RELATED WORK

Some papers such as (Brown et al., 1975) question and study the time-constancy of linear regression models. (Kalaba et al., 1989) suggests that the regression coefficients evolve slowly over time. That is why it is very common in forecasting time series, to consider basing the forecasts only on the most recent version of a time series model rather than on a model built from the entire series; and that’s because the recent past contains more information about the immediate future than the distant past (Guthery, 1974).

There are two concerns regarding PLR that we are going to address in this paper. First, PLR is generally used to observe patterns and describe trends in datasets (e.g. trends in natural phenomena in medical or biomedical fields of research) and not used for predictions. Second, in common piecewise regression approaches, it is only the information about the input space that is used for partitioning the data (identifying the split points) (Nusser et al., 2008). (Nusser et al., 2008) for the first time, did the partitioning using the target variable. The authors in that paper suggest that ignoring the target variable and clustering the input data for partitioning the space is an insufficient strategy in two cases: (a) “where the data points cannot be distinguished within the input space”; (b) in regions of high data density in real-world application problems. These regions usually correspond to operating points of the system which are not necessarily appropriate for partitioning the input space; the reason is that “the underlying function of the system might not change within the operating points but in regions with low data density”.

There are some researches that address the partitioning of the data space by considering the values of the target variable. For example, (Hathaway and Bezdek, 1993, Ferrari-Trecate, 2002, Höppner and Klawonn, 2003) use clustering for this purpose.

From an application point of view were found out that there has been a similar work by (Jahandideh et al. 2009) but their PLR model was completely different from that of ours. In (Jahandideh et al. 2009) the goal was devoted to offer a suitable model to predict the quantity (rate) of medical waste generation. They used ANNs and (Multiple Linear Regression) MLR models. Their results indicate that the ANNs model show absolutely lower error measure compared to the MLR model. Based on those results, ANNs indicated superiority in assessing the quality in term of accuracy.

In this paper, we incorporate the information about the target variable into the process of split (break) point identification in piecewise linear regression modeling and use the PLR for long-range forecasting of time series. An important factor in many fields of application where transparency and understandability are highly required in order to evaluate the proposed solutions by the experts of that field is the interpretability of the solution (Nusser et al., 2008). Using Moving Average, our proposed solution is very easy to understand and interpret by the experts of other fields where little mathematical knowledge is required. The need for this kind of prediction has been always there in the City of Calgary but has been mainly addressed intuitively.

4. BACKGROUND

Different machine learning (ML) approaches are applied to different fields of research and the common agreement indicates that not all of them perform well on a specific dataset. Depending on the application, data type and the goal of learning, different ML approaches are suggested.

In this section we present two of the commonly used methods -the performance of which have been usually compared to each other in the literature (Fischer, 2008, Sikka et al., 2010)- to model and forecast time series. These methods are PLR and SVM. In this paper, we propose an approach to time series predictions using PLR and then compare its performance with SVM. The results are also compared to the results of a Perceptron NN predictions (ANN is popular in forecasting time series).
We chose SVM because it is well known that SVM approach works well for pattern recognition problems and for estimating real-valued function (regression problems) from noisy sparse training data (Cherkassky and Ma, 2005). Therefore it is a good competitor for PLR. This means that it allows for a good evaluation of the proposed PLR prediction method through comparing its performance with that of SVM.

### 4.1 Piecewise Linear Regression (PLR)

According to (Ferrari-Trecate, 2002) PLR can be formulated as follows. Suppose X is the whole input space in the n-dimensional space \( \mathbb{R}^n \). Suppose \( \{X_i\}^n_{i=1} \) are disjoint regions of X, where \( X_i \cap X_j = \emptyset \) (\( i,j = 1,2,...,s \)) and \( \bigcup_{i=1}^n X_i = X \). Then the generation of a piecewise linear regression function is to determine \( f: X \to \mathbb{R} \) with a linear behavior in each \( X_i \) (Figure 1). The function in each partition looks as the following:

\[
f(x) = \omega_{i0} + \sum_{j=1}^n \omega_{ij}x_j \quad \text{where} \; x \in X_i \quad (1)
\]

Suppose each point in partition \( i \) is represented by \( (Z_i,Y_i) \) where \( Z_i \) represents a point in a n-dimensional input space and \( Y_i \) is the actual target variable. Suppose \( \hat{Y}_i \) is the target variable calculated by the linear modeler. Then the following is true: \( Y_i = \hat{Y}_i + \varepsilon_i \) where \( \hat{Y}_i = \beta_0 + \beta_1 Z_i \) and \( \varepsilon_i \) shows the deviations (errors) from the actual function values. The problem in each region is now reduced to estimating the intercept (\( \beta_0 \)) and the slope (\( \beta_1 \)) by minimizing residual sum of squares (RSS).

\[
RSS = \sum \varepsilon_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \beta_0 - \beta_1 Z_i)^2 \quad (2)
\]

The following figure (Figure 1) shows an example of PLR in a case where the target variable is modeled with different straight lines in different time spans (the red lines in figure 1). In this figure the x-axis shows the time (years) and the y-axis shows the values of the target variable.

![Figure 1. An example of PLR in time series.](http://processrends.com/loc_trend_analysis_with_excel.png)

### 4.2 Support Vector Machine (SVM)

The idea of SVM was first developed in (Cotres and Vapnik, 1995). It is a useful technique for data analysis and usually used for classification and regression analysis.

As illustrated in Figure 2, SVM maps the original input \( x \in \chi \) onto a higher dimensional feature space \( \mathcal{H} \) by a (potentially non-linear) function \( \phi(\cdot): \chi \to \mathcal{H} \), where \( x \) is a point in the input space \( \chi \). SVMs classify a new pattern like \( x \), using the output \( y \) that is given by the following formula (Culpepper and Gondree, 2008):
\[ y = b + \langle \phi(x), \phi(w) \rangle = \omega_0 + \text{Ker}(x, w) \quad (3) \]

\( \text{Ker}(.) \) is a kernel function that returns the dot product of the image of the two inputs in the higher dimensional feature space \( H \). The existence of a kernel function means each input vector \( x \) does not need to be mapped to \( \phi(x) \). In fact, given a kernel function; we do not even need to know \( \phi(.) \) (Culpepper and Gondree, 2008). The weights \( w \) are related to the \( m \) training vectors in the set \( \text{Training} = \{x_j\} \) by the following (Culpepper and Gondree, 2008):

\[ w = \sum_{j=1}^{m} \alpha_j \phi(x_j) \quad (4) \]

Figure 2. General mapping idea in SVM.

The SVM learning algorithm chooses the one with the maximum margin around it. Only those training vectors which are on the margin of the discriminating hyper plane are significant (Scholkopf et al., 1999). These are called the support vectors. As these are discovered, they populate a set of support vectors (SV):

\[ SV = \{x_i\} \subseteq \text{Training} \quad (5) \]

### 4.3 ARTIFICIAL NEURAL NETWORKS (ANN)

We use a two layer feed-forward Perceptron neural network. This network changes the weight by an amount proportional to the difference between the desired output and the actual one.

\[ \Delta W_i = \eta \ast (D - Y) \cdot I_i \quad (6) \]

Where \( \eta \) is the learning rate, \( D \) is the desired output, and \( Y \) is the actual output.

Our network has a single output and 6 real-valued inputs. The idea is to generate the output using a linear combination of inputs according to the input weights and then possibly putting the output through some nonlinear activation function; mathematically this can be written as (Honkela, 2001):

\[ y = \varphi(\sum_{i=1}^{n} \omega_i x_i + b) = \varphi(W^T X + b) \quad (7) \]

Where \( W \) denotes the vector of weights, \( X \) is the vector of inputs; \( b \) is the bias and \( \varphi \) is the activation function. In the original Perceptron a Heaviside step function was used. But now the activation function is often chosen to be the logistic sigmoid \( 1/(1 + e^{-x}) \) or the hyperbolic tangent \( \tanh(x) \).
The characteristics of this type of network are: no connections within a layer; no direct connections between input and output layers; fully connected between layers; number of output units need not equal number of input units; number of hidden units per layer can be more or less than input or output units.

In our case study we set both the learning rate and the momentum to 0.5. Error epsilon is 1.0E-5 and there are 500 training cycles in which the learning rate decreases. This is called decay is meant to avoid over fitting.

5. PROBLEM STATEMENT

Imagine a nonlinear pattern in a time series where the values of the target variable follow a periodical pattern. If in each period the pattern can be divided into smaller time intervals where it can be modeled by a linear function, then it can be said that this nonlinear function can be modeled using a set of linear functions. An example of such time series can be seen in figure 4 and figure 1. As we mentioned before, another advantage of this type of modeling is that it provides easier interpretability for the experts in different areas. But our goal is not just modelling the data pattern but also using PLR to make predictions in both near and distant future.

As we mentioned in section 1, LR is usually preferred for data reduction, data modelling, and data interpolation. Predicting the closest future points are also another application of LR. However, we intend to use PLR for both modelling and predicting the near and distant future points in a periodical time series - which can be modeled by LR in smaller time spans in each period.

Beside a solution for applying PLR in this type of prediction, one has to think of a way of finding the split points in each of the periods too.

In other words, we are looking for the set of disjoining points or split points like \( y_d \in \mathbb{R} \), \((d = 1,2,\ldots,S-1)\) among the values of the target variable. Between every two consecutive split points, the representing points in the input space can be classified into one class like \( X_i \). Therefore we will have \( S \) classes \( \{X_{i_1}\}_{i=1}^{S} \) in the input space, where \( X_i \cap X_j = \emptyset \) and \( \bigcup_{i=1}^{S} X_i = X \).

At each of the classes in \( X \), a line can represent the behaviour of the target variable: \( f_i: X_i \rightarrow \mathbb{R} \). We are studying the special case where one or more of these lines are being repeated in the time series with a time period of \( \tau \), \( \tau \in \mathbb{N} \) and \( \tau < S/2 \) with slight differences in the slope. Knowing the \( \tau \) we will be able to predict a point in future located in \( n \tau \), by using the most recent corresponding line in \( (n-1)\tau \) where \( n \in \mathbb{N} \). The following section (section 5) describes our solution approach.

6. SOLUTION APPROACH

In this section we start by giving a description about our dataset and then we talk about the different stages that we have gone through towards solving the problem.

6.1 DATASET

We received and processed the daily transaction records (since April, 2009) of recyclables collected from Blue Carts -carts in residential places where residents put the recyclable wastes- from the WRS at the City of Calgary. This dataset contains the details about the amount of different recyclables delivered at the landfills by each truck in each day.
From some pre-evaluations and after discussion with the domain experts, we found out that the amount of recyclables is greatly influenced by the changes in the seasons as well as changes in weekdays. Thus, in the dataset we split the Date attributes into three attributes: month, week number and weekday. Finally, the dataset was structured as shown in Figure 3. The attributes are Month, Weknunb, Weekday, Date and different Weights, where B1 to B3 refer to the weights of Blue Cart recyclables. The numbers 1, 2, and 3, following the alphabet B correspond to the weight of the full truck as it enters the landfill, the load weight and the truck weight, respectively.

<table>
<thead>
<tr>
<th>Month</th>
<th>Weknunb</th>
<th>Weekday</th>
<th>Date</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>17</td>
<td>3</td>
<td>1</td>
<td>947.41</td>
<td>500.60</td>
<td>146.81</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>4</td>
<td>2</td>
<td>513.58</td>
<td>555.40</td>
<td>0.0040</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>5</td>
<td>3</td>
<td>967.28</td>
<td>729.00</td>
<td>138.28</td>
</tr>
</tbody>
</table>

Figure 3. Sample data after data pre-processing.

6.2 DATA PRE-PROCESSING

The raw data contained lots of noises, empty records and unrelated columns and it was in different tables. First, data cleaning was performed to remove noises and empty records. Then a review of the attributes was done to delete those obviously unrelated attributes, e.g. Truck No. (plate) and specific transaction hours -because the minimum prediction granularity is for each day. The next stage was data aggregation. It was performed to merge multiple records in one day into one record containing the total amount of material in that day. Finally, we aggregated the original 31,898 hourly records into 700+ daily records.

We used Microsoft SQL Server to combine data tables and aggregate records. But most of our work was done with RapidMiner, which covers most of the machine learning techniques. It follows a modular operator concept that allows the design of nested chains of operators for huge learning problems; so we integrated the implementations of most of our system components (including data pre-processing, model training, testing, and prediction and evaluation) into its environment.

6.3 SPLIT POINTS DETECTION

By visualizing the Blue Cart data, after pre-processing phase, lots of outliers could be easily observed at the bottom of the plot as shown in Figure 4. Further investigation indicated that those outliers mostly occur during weekends (Saturdays, Sundays) and Mondays, and that is because of certain operational regulations at the City of Calgary. Most of these data had values equal or very close to zero, so we eliminated those days from our dataset.

Using a time window of 20 days, the Moving Average of the target variable values is calculated. This time window of twenty days moves one day by one day and the process of averaging repeats until all data points are met. A curve will be fitted on the data points resulted from the Moving Average and then the slope of the curve is calculated. Using the slope, the extreme points of the curve are identified. These extreme points are the split points (y_d) in this periodical time series, if they are repeated periodically.

The points between these extreme points in the original dataset are modeled by linear regression. For predicting the value of the target variable in a point in future, all we need is to identify in which corresponding time slot of the most recent period this point is located. Then we can choose the line that models that time slot in the time series history data and use it to predict.

To calculate the Moving Average of a variable, a window of a fixed size is moved over the time (time unit by time unit) and at each position, the values that lie within the window are aggregated according to an average function. The aggregated average value is the Moving Average value for that window at that position.
As depicted in Figure 4, starting from late May until early days of July, there is an instant increase in the amount of recyclables in Blue Carts. This is because the Blue Cart was introduced to different parts of the city, step by step and in different time slots. In our approach, we do not take into account this portion of time because it happens just once for all and it is not representing the natural behavior of the target variable (because not all parts of the city had Blue Carts at that time).

![Figure 4. Visualization of Blue Cart actual data including outliers.](image1)

As depicted in Figure 5, there is a pattern in Blue Carts data. This pattern repeats every six months. Those six months can be divided into two consecutive three months periods. In each of those three months periods the data points are almost moving along a single line which is why we used linear regression to model each of them separately.

![Figure 5. The pattern in the Blue Cart data (Moving Average) in 2009 and 2010.](image2)
6.4 PERFORMANCE MEASURES

Using two popular measures (RMSE and Spearman rho) we were able to evaluate the performance of the proposed approach and also compare it with the performance of SVM and ANN.

The first evaluation measure is the Spearman rho ($\rho$), a non-parametric measure of correlation, which was first introduced in (Spearman, 1904). The Spearman rho has two advantages. First, a normal distribution is not necessary because it is a nonparametric measure, and second, it is less affected by outliers (Gauthier, 2001). $R$ (rho) is computed by the following formula:

\[
R = 1 - \frac{6 \times \sum D_a^2}{A (A^2 - 1)}
\]  

(6)

$D_a = U_a - V_a$ where $U_a$ and $V_a$ are the actual value and the predicted value by the regression model ($a \in \{1, \ldots, n\}$, where $n$ is the number of the target variable points in the time series). In this context the mean squared error (MSE) is defined as follows:

\[
\text{MSE}(Y) = E\left(\hat{Y} - Y\right)^2
\]

(7)

Where $\hat{Y}$ and $Y$ are: the estimated value by the regression method and the actual value at a certain point, respectively. Taking the square root of MSE yields the root mean squared error or RMSE. RMSE has the same units as the quantity being estimated which makes it easier to judge the regression method’s error.

7. EVALUATION (CASE STUDY)

In order to study the feasibility and the accuracy of the proposed approach we took the case of the Waste and Recycling Services (WRS) unit at the City of Calgary (CoC) in Alberta, Canada. Prediction of recyclable material collected throughout Calgary is part of a project under the umbrella of the collaborative work between the City of Calgary and the University of Calgary (Urban Alliance Collaboration (The City of Calgary, 2010)). In this case study, we focused on predicting the amount of recyclables in the residential Blue Carts.

7.1 APPLYING PLR

Cross-validation was used to train the linear regression model for each time span (between two consecutive split points) and estimate how well the model would perform if it is applied to other data which have not been used during the training process. Here, we divided the example set into ten folds for a 10-folds cross-validation. Then we used all folds but one to train the learning algorithm. The excluded fold is used to evaluate (test) the learnt model. This process repeats until all folds have been used once for testing.

Each linear regression model calculated by RapidMiner is based on the Akaike criterion for model selection. RapidMiner provides two types of feature selection methods for a linear regression model, M5 and Greedy. In this case study, both of them yielded into almost the same prediction results (their relative error differed by 0.12%). In general, the Greedy approach that we used for all experiments performed slightly better than M5 method.
7.2 EXPERIMENTAL RESULTS

We applied the proposed PLR, as well as SVM and NN, to predict the daily amount of recyclables for two different months (July, 2010 and September, 2010). We chose these two months because they are in two different parts of the repeated pattern, where the pattern can be modeled by two different lines (figure 5). For each of the above mentioned cases, the history data prior to the target month was used to train the model.

Using the performance evaluation metrics that we discussed in section 5, a comparison between the performance of SVM, NN, and the proposed PLR approach is made and presented in Table 1 and Table 2.

Figure 6 and Figure 7 show the prediction results for the two months. For both figures, the blue diagram depicts the predicted amount of wastes in Blue Carts, using piecewise linear regression; whereas the red diagram shows the actual amount. Table 3 and Table 4 also provide some other evaluation measures for the proposed PLR approach and give a broader evaluation view for both predictions in the two months.

Table 1. Evaluation parameters comparing SVM to the proposed approach on predictions for September 2010.

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
<th>Relative error</th>
<th>Spearman rho</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>10.568</td>
<td>2.67%</td>
<td>0.940</td>
</tr>
<tr>
<td>SVM</td>
<td>19.730</td>
<td>5.01%</td>
<td>0.886</td>
</tr>
<tr>
<td>NN</td>
<td>10.212</td>
<td>2.56%</td>
<td>0.939</td>
</tr>
</tbody>
</table>

Figure 6. Amount of material in Blue Cart in September 2010; the actual amount (in red) and predicted amount (in blue) – PLR Prediction.

Table 2. Evaluation parameters comparing SVM to the proposed approach on predictions for July 2010.

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
<th>Relative error</th>
<th>Spearman rho</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>7.590</td>
<td>1.95%</td>
<td>≅1.000</td>
</tr>
<tr>
<td>SVM</td>
<td>15.575</td>
<td>4.31%</td>
<td>0.905</td>
</tr>
<tr>
<td>NN</td>
<td>9.611</td>
<td>2.43%</td>
<td>0.969</td>
</tr>
</tbody>
</table>
8. CONCLUSION

We incorporated the information about the target variable into the process of split point identification in PLR. The proposed approach is an easily interpretable method which makes it very convenient for the experts of different fields of research to use PLR, interpret the patterns and make conclusions out of the highly accurate forecasts.

In this paper, the application of PLR in seasonal forecasting in time series with nonlinear patterns is newly introduced. The identification of the split points (break points) using Moving Average is also new. The applicability and accuracy of the proposed approach are demonstrated in a case study at the City of Calgary.

The results show that the proposed approach is a close competitor of the NN. This close performance could be attributed to the NN’s non-linear nature which provides the opportunity to relate different variables to a target variable.

Table 3. Evaluation measures of the proposed approach on predictions for September 2010.

<table>
<thead>
<tr>
<th>Performance parameters</th>
<th>Blue Cart (PLR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Root mean squared error</td>
<td>10.568 +/- 0.000</td>
</tr>
<tr>
<td>Absolute error</td>
<td>0.749 +/- 5.928</td>
</tr>
<tr>
<td>Relative error</td>
<td>2.67% +/- 1.73%</td>
</tr>
<tr>
<td>Relative error lenient</td>
<td>2.67% +/- 1.73%</td>
</tr>
<tr>
<td>Relative error strict</td>
<td>2.77% +/- 1.83%</td>
</tr>
<tr>
<td>Normalized absolute error</td>
<td>0.626</td>
</tr>
<tr>
<td>Root relative squared error</td>
<td>0.662</td>
</tr>
<tr>
<td>Squared error</td>
<td>111.681 +/- 112.955</td>
</tr>
<tr>
<td>Prediction average</td>
<td>318.862 +/- 15.969</td>
</tr>
<tr>
<td>Spearman rho</td>
<td>0.940</td>
</tr>
</tbody>
</table>

Figure 7. Amount of material in Blue Cart in July 2010; the actual amount (in red) and predicted amount (in blue) – PLR Prediction.
Table 4. Evaluation measures of the proposed approach on predictions for July 2010.

<table>
<thead>
<tr>
<th>Performance parameters</th>
<th>Blue Cart (PLR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Root mean squared error</td>
<td>7.590 +/- 0.000</td>
</tr>
<tr>
<td>Absolute error</td>
<td>5.511 +/- 5.219</td>
</tr>
<tr>
<td>Relative error</td>
<td>1.95% +/- 1.72%</td>
</tr>
<tr>
<td>Relative error lenient</td>
<td>1.91% +/- 1.67%</td>
</tr>
<tr>
<td>Relative error strict</td>
<td>1.98% +/- 1.78%</td>
</tr>
<tr>
<td>Normalized absolute error</td>
<td>0.370</td>
</tr>
<tr>
<td>Root relative squared error</td>
<td>0.346</td>
</tr>
<tr>
<td>Squared error</td>
<td>57.606 +/- 110.227</td>
</tr>
<tr>
<td>Prediction average</td>
<td>281.252 +/- 21.923</td>
</tr>
<tr>
<td>Spearman rho</td>
<td>≅ 1.000</td>
</tr>
</tbody>
</table>

The seasonal pattern covering a six month period of time is one of our observations that might be of interest to the CoC. This pattern in the amount of recyclables in Blue Carts reaches its extreme points almost every three months. This information can be of help to optimizing the human resources as well as trucks deployments in the CoC.

In this paper, we just based the forecasts on the most recent linear pattern that corresponds to the prediction point rather than considering the similar patterns happening earlier than that. And that’s because, as we mentioned in section 1, “the recent past contains more information about the immediate future than the distant past”. However, this does not take into account the possible changes on the patterns (e.g. changes in the width or position of time slots). In addition to this limitation of this study, there is a limitation related to our dataset. The available history data is limited to almost 2 years, and that is because the residential Blue Cart program was just launched in April 2009.

In our future work we intend to study the effect of integrating a competitive learning method - similar to the one used in learning the synopsis weight in competitive neural networks- into the proposed forecasting piecewise linear regression approach in this paper. In this way, we will be able to take into account the variation of the coefficients – which we talked about in section 1- in different time slots.

9. ACKNOWLEDGEMENTS

This research is part of the collaboration project between the University of Calgary and the Waste and Recycling Services unit at the City of Calgary and it is under the Urban Alliance project umbrella. The project (NSERC CRD #386808-09) is supported by NSERC (Natural Sciences and Engineering Research Council) and the City of Calgary. We would also like to thank Scott Banack from WRS at CoC for his time and consideration.

10. REFERENCES


LOESCH, D. Z., BUI, Q. M., HUGGINS, R. M., MITCHELL, R. J., HAGERMAN, R. J. & TASSONE, F. 2006. Transcript levels of the intermediate size or grey zone fragile X mental retardation 1 alleles are raised, and correlate with the number of CGG repeats. Journal of medical genetics, 44.


