A Discrete Constraint-Based Method for Pipeline Build-Up Aware Services Sales Forecasting

Peifeng Yin[(⊠)], Aly Megahed, Hamid Reza Motahari Nezhad,
and Taiga Nakamura

Cloud Services Analytics Group, IBM Almaden Research Center, San Jose, CA, USA
{peifengy,aly.megahed,motahari,taiga}@us.ibm.com

Abstract. Services organizations maintain a pipeline of sales opportunities with different maturity level (belonging to progressive sales stages), lifespan (time to close) and contract values at any time point. As time goes, some opportunities close (contract signed, or lost) and new opportunities are added to the pipeline. Accurate forecasting of contract signing by the end of a time period (e.g., quarterly) is highly desirable to make appropriate sales activity management with respect to the projected revenue. While the problem of sales forecasting has been investigated in general, two specific aspects of sales engagement for services organizations, which entail additional complexity, have not been thoroughly investigated: (i) capturing the growth trend of current pipeline, and (ii) incorporating current pipeline build-up in updating the prediction model. We formulate these two issues as a dynamic curve-fitting problem in which we build a sales forecasting model by balancing the effect of current pipeline data and the model trained based on historical data. There are two challenges in doing so, (i) how to mathematically define such a balance and (ii) how to dynamically update the balance as more new data become available. To address these two issues, we propose a novel *discrete-constraint method (DCM)*. It achieves the balance via fixing the value of certain model parameters and applying a *leave-one-out* algorithm to determine an optimal free parameter number. By conducting experiments on real business data, we demonstrate the superiority of DCM in sales pipeline forecasting.

1 Introduction

IT services organizations manage a pipeline of *sales opportunities*. Such sales opportunities/deals go through an elaborative process that may take three months to two years of negotiations and working until a contract is signed or the opportunity is lost. During this process, deals move from one sales stage (e.g. qualified) to another (e.g., conditional agreement). In order for an IT services organization to manage its sales effectively, they need to have the ability to forecast services sales revenue one quarter, two quarters, or sometimes up to a year in advance to harvest new opportunities or make sign/no-sign decisions on others.

Existing works for services sales forecasting (e.g., [10,17,24]) fit in one of two main categories. The first category includes those operating at the opportunity-level, i.e., predict the revenue by considering which opportunities currently present in the pipeline may turn as won. Then, projected win values are aggregated to get the estimated sales revenues. More technically, this can be described as learning/applying a sale conversion factors from historical data to compute the won value of the pipeline at the end of the target period. The next category operates at the aggregate pipeline level; i.e., the total value of opportunities at a given sales stage. They predict the sales revenue for the target quarter mainly by learning a model from historical aggregated values of pipeline. However, both categories are concerned only with the current pipeline information, and do not consider future opportunities that may still end up being won within the remaining time of the target period. In this context, an critical factor is predicting how much the pipeline may grow from that point in time, referred as "growth factor". This factor may be high in the beginning of the time period, where more opportunities are constantly added to the sales pipeline. From a given point of time, that varies depending to the type of business, the pipeline does not monotonically grow throughout the target period. Indeed, after some more time (e.g., after mid-period), this value may get smaller. This is because no new sales opportunities with a target closure date of that period are added anymore, and existing opportunities are identified as loss or their projected closure dates are transferred to a next target period. Therefore, it is important to consider the growth factor for a more accurate prediction.

In this paper, we focus on the problem of dynamic modeling of services sales forecasting in which both current pipeline build-up and the pipeline growth are considered in making sales forecast for a given target time period. As a modeling option for this problem, one may think of treating it as a time series problem [1,3,20,22]. However, the typical assumption of time series does not hold in this context. The classic time-series model, e.g., ARIMA, focuses on one single continuous timeline and the goal is to predict future value (e.g., $t + k$) based on historical data. In our problem, the sales forecasting data consists of multiple series of data with varying time lengths, in which sales opportunities get added with a target closing date at different points in time (e.g., few quarters in advance), each representing the change of pipeline value during the time period. These periods can not be simply concatenated to a single one because the life span of any two periods may overlap with each other. Also, compared to what a typical time series model expects, the number of data points in the series is extremely low. For example, in a typical real-world sales data set, it is not uncommon to have few data points. For instance, in our real-world data set used in the experiments of this paper, there are only 28 points per period. This sparse quantity makes it rather hard to apply models such as ARIMA that assumes large sets of data points in a time series.

Given the above context, we formulate the growth forecast as the process of a dynamic polynomial curve-fitting. With all historical records, we propose learning of a polynomial growth curve. It serves as a basis of the forecast.

The other issue with the curve-based methods is that the learned model is based on historical data. However, for any target period, as time goes, more estimated win contribution values will be available. In this case, the problem is then how to dynamically update the prediction accordingly. To solve this problem, we propose a *Discrete Constraint Method (DCM)*. The general idea is to enforce the similarity between the historical curve and the one that is fitted only on the current pipeline data. The key challenge lies in two-folds: (i) how to formulate such similarity degree and (ii) how to find an optimal similarity degree. Our model addresses the first one by limiting the number of free parameters, mainly according to their impact on objective function. With such design, the second issue is reduced to determining the optimal number of free parameters. To solve this latter problem, we apply a *leave-one-out algorithm*. This algorithm takes turns to mask one historical pipeline record as a pseudo target period. The optimal number is the one that achieves lowest prediction error in the validation.

Thus, in summary, the contributions of this work are as follows:

- We formulate the problem of services sales growth prediction as a dynamic curve-fitting problem, considering both the historical curve and current available pipeline data points.
- We solve the problem by designing models that control the degree of similarity between the historical curve and the one learned on current pipeline data. Particularly, we propose a Discrete Constraint Model which formulates the degree of similarity by fixing some of the model parameters to have the same value with the historical one, according to the absolute value of the gradient.
- We propose a leave-one-out algorithm that dynamically determines the constraint degree of similarity according to the amount of available data.
- We report the results of experiments on real business data to evaluate the performance of the model, which shows the superiority of the proposed method over existing works in terms of prediction accuracy.

The rest of the paper is organized as follows. Section 2 gives a literature review. Section 3 illustrates the curve-fitting problem. Sects. 4 and 5 give details of the solution and experiment results, respectively. Finally Sect. 6 concludes the whole work and describes potential future research.

2   Related Work

The related work can be studied in two broad categories: (i) the opportunity-level prediction and (ii) the time-series driven sales analysis. The former one relates to works that train different types of models according to information about each opportunity (or deals as mentioned in other works). The goal is to predict the outcome [6,10,23], the health status [18] or monitor the progress [24]. For further details on services in IT service contracts, we refer the reader to our previous works in [9,16].

From the first category, Greenia et al. [10] presented a quantitative-based approach for the prediction of deal outcomes and identified a number of key

factors that are highly correlated to the deal's final outcome. The prediction approach is based on a Naïve Bayesian classifier that is trained on these factors (features). In [6], Carman et al. investigated the problem of comment-based opportunity outcome prediction. Nezhad et al. [18] presented an text analytics-based deal outcome prediction approach that integrates a concept clustering approach, and sentiment analysis, and a set of semantics-based features for deal outcome prediction. These new features are used to train classifiers to predict the outcome of each deal. In [23], Yan et al. apply the multi-dimension Hawkes process to model the probability of deal outcome. In that work, the deal information determines a base probability, which is then updated based on the history of interaction of sellers with the client. Such impact exponentially decays in terms of the time interval when those interactions happened. In our prior work [24], we applied multinomial and Dirichlet process to model the progress of a deal. Given the deal information and update records, the model is capable of predicting the next event type (new update, win or loss) as well as the time interval. These works provide methods and tools for sales manager on opportunity-level control. However, they focus on opportunities in the current pipeline and does not consider the future pipeline growth by the end of the target period.

There have been a few decades for research in time-series analysis. For a detailed review of this field, we refer readers to the survey by Gooijer and Hyndman [7]. As early as 1960, Winters [22] presented an exponential forecasting model for sales based on moving averages. ARIMA [4] (or its variant) is another popular method used in time-series sales prediction [2,5,13,21] and multiple works [8,19] have demonstrated its good performance in "mediate and short-term forecasts" [3]. Furthermore, artificial neural network (ANN) is a relative new method that is studied and compared with conventional ones [3,11,12,14]. Finally, a few works [1,3,15] propose to use hybrid methods of conventional ARIMA and ANN in time-series prediction. These methods work on a single long time-series data and requires a sufficiently large quantity of data points for training. In our problem, however, data consists of multiple overlapping time periods. These periods can not be concatenated to a single one because the life span of two periods may overlap with each other, and the target time period may not have the same time scale of the historical ones. Also, the length of each period is quite short compared to conventional time-series data. For example, in our experimental data set there are only 28 points for a given period. This small quantity makes it hard to apply complex model such as ARIMA. Finally, the number of periods is also relatively small as historical data from far back in the history may not be relevant to today's business due to the change of business strategy, business lines or products or services.

3   Problem Illustration

Services pipeline is building up as time passes and as new opportunities are arriving with an estimated closing date within the target time period. At any given point in time, there is a pipeline build up, and also future growth potential

for the pipeline. Forecasting the period-end sales revenue needs to consider both current pipeline from the lens of historical model as well as the future growth in order to capture the full picture of pipeline sales by the end of the target period.

Formally, let $\mathbb{D}$ denote historical data and $\mathbf{Y}_n = \{y_1, \cdots, y_n\}$ represents the current data with only n available data points. Suppose T is the total number of time points for full period, the problem is then formulated as below. Given $\mathbb{D}$ and $\mathbf{Y}_n$, find a function $\mathbf{f} : |\mathbb{D}| \times n \to 1$ to make estimation $\hat{y}_T = \mathbf{f}(\mathbb{D}, \mathbf{Y}_n)$ such that $|\hat{y}_T - y_T|$ is minimized.
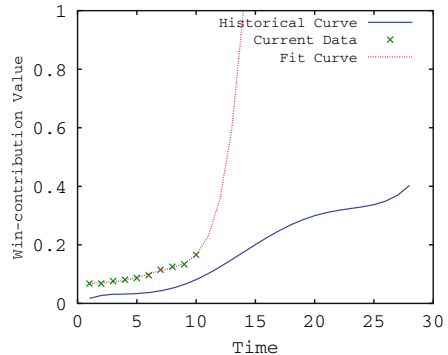
Although the goal is to predict the period-end pipeline value, we can model the change of the pipeline build up during the whole life cycle. At any given time point, there is a record of the pipeline. Connecting all these records models the growth of the pipeline as a curve and capturing the growth trend can thus be solved via curve fitting. Consider Fig. 1 as an illustration. In this figure, the blue line represents the curve fitted on historical data while the green crossing stands for the estimated win-contribution value of target period. The goal is to predict where the final green crossing will locate.

On one extreme, we may totally discard the current data and only rely on the historical curve, i.e., $\mathbf{f}(\mathbb{D}, \mathbf{Y}_n) \to \mathbf{f}'(\mathbb{D})$. Reflecting this method in Fig. 1, it is equivalent to use the end of blue line to estimate the value. This method is problematic. As can be seen, the green crosses does not lie closely to the curve, indicating that the target period growth may have a different pattern. Relying only on the history may suffer inaccuracy. On the other extreme, we may simply discard the historical model and fit a completely new curve based on the current data, i.e., $\mathbf{f}(\mathbb{D}, \mathbf{Y}_n) \to \mathbf{f}''(\mathbf{Y}_n)$. This method is illustrated in Fig. 2, which demonstrates a high possibility of overfit. Thus this method is also problematic.

We argue that a good strategy should combine both history and present, using current data to fit a new curve while at the mean time applying historical curve to avoid overfit. In another word, we need to find a balance point between
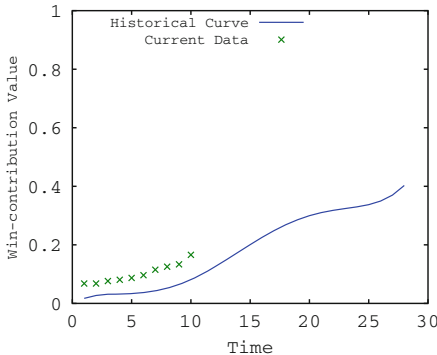
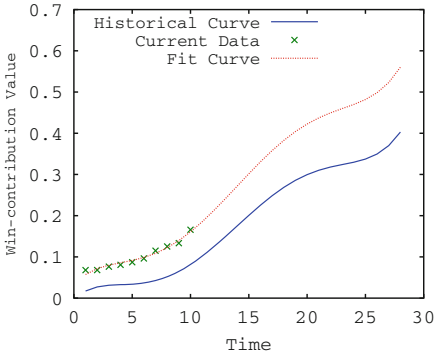**Fig. 1.** Problem illustration   **Fig. 2.** Risk of overfit
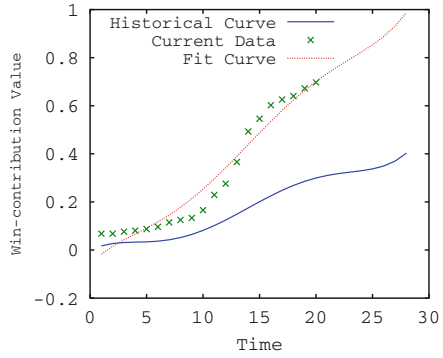
Fig. 3. A reasonable fit

Fig. 4. Dynamic update on curve

current data and historical model. Figure 3 shows an example of fitting considering these two factors. Finally, as time goes forward, more and more current data will be available. In this case, the balance point needs to be dynamically updated accordingly. Figure 4 displays an example of fitting when more data points are available. As can be seen, the curve is different from what it used to be in an earlier time as in Fig. 3.

As can be seen from the example above, there are two key challenges in the dynamic curve-fitting problem: (i) definition of balance point in combining current data and historical model and (ii) mechanism of dynamically updating such balance point as new data becomes available.

One intuitive method is to simply add the current data into historical one and fit a curve on this combined data. Figure 5 shows the result of this method. As can be seen, the first part of the fitted curve try to find a tradeoff between current and historical data. The latter part of the curve overlaps with the historical one.
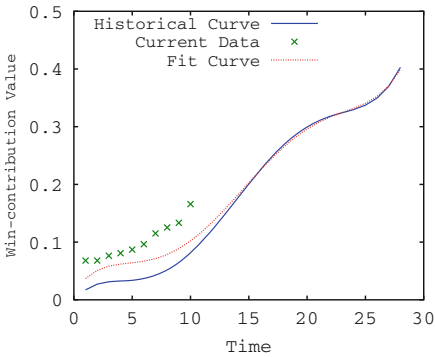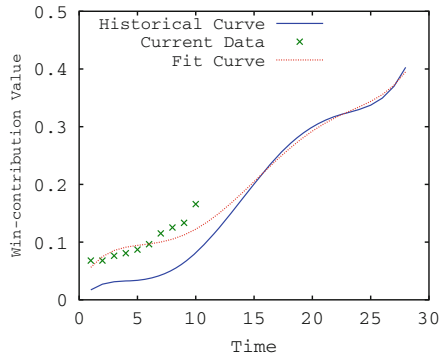

Fig. 5. Combine current data

Fig. 6. Replace history

Since we use the end of the curve as the prediction of period-end value, this method makes no difference with the one using history data only.

An alternative method is to replace part of historical data with current one and connect with other historical record. Then a curve can be fit on such hybrid data set, as shown in Fig. 6. It can be seen that the resulted curve tries fitting the current data and in later part goes closely to the historical data. Again the current data does not affect the final prediction in this method.

As can be seen from the two possible methods above, a straightforward way of combining historical and current data does not help much in forecast prediction. What is needed is a methodological approach to define the balancing mechanism in order to achieve a reasonable fit as shown in Fig. 3.

4   Pipeline Build-Up Aware Sales Forecasting

In this section we first introduces the curve-fitting approach. Then, we present the discrete constraint model. It formulates the problem as the determination of free and fixed parameters in the model. Also, we introduce a leave-one-out validation mechanism as a method to determine how many parameters should be fixed. For easy reference, we list all symbols and their meanings in Table 1.

Table 1. Summary of symbols

Symbol	Meaning
$\mathbf{w}$	Parameters of the curve
$\mathbf{w_0}$	Parameters of the curve learned on historical data
$\mathbf{Y}$	The current data
$\mathbf{C}$	The constraint matrix
l	The number of non-zero diagonal values in $\mathbf{C}$
M	The number of historical periods
$\mathbf{X_1} \cdots \mathbf{X_M}$	Historical data

4.1   Curve-Fitting Methodology

Curve-fitting is a special case of regression, one classic machine learning problem. Given a series of pair data, the goal is to fit a curve that can map the independent variable value to the dependent one. Particularly in this work we consider the polynomial curve-fitting where the independent variable is time and the dependent one is the corresponding sales win value in pipeline.

Formally, let $\mathbf{w}$ denote the vector of model parameter, which satisfies a zero-mean Gaussian distribution. Moreover, suppose the parameters are independent

of each other but shares the same standard deviation σ, we may construct the objective function as in Eq. (1).

$$\mathfrak{L}(\mathbf{w}|\mathbf{Y}) = \mathfrak{R}(\mathbf{Y}, \mathbf{w}) + \log \mathcal{N}(\mathbf{w}; 0, \sigma^2)$$
$$= \mathfrak{R}(\mathbf{Y}, \mathbf{w}) + \log \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\mathbf{w}^T\mathbf{w}}{2\sigma^2}} \right) \propto \mathfrak{R}(\mathbf{Y}, \mathbf{w}) + \frac{\mathbf{w}^T\mathbf{w}}{2\sigma^2} \qquad (1)$$

where $\mathbf{w}^T$ is the transpose of a vector, $\mathbf{Y} = \{(t_1, y_1), \cdots, (t_n, y_n)\}$ represents the current data and $\mathfrak{R}$ is an objective function evaluating the error of model fitness with $\mathbf{w}$. Particularly in our problem scenario, this objective function is a k-order polynomial curve-fitting where $\mathbf{w} = \{w_1, \cdots, w_k\}$, as shown in Eq. (2).

$$\mathfrak{R}(\mathbf{Y}, \mathbf{w}) = \sum_{i=1}^{n} (y_i - \sum_{j=0}^{k} t_i^j w_j)^2 \qquad (2)$$

The second part in Eq. (1) is the penalty of large difference between the model and zero value. This term is widely used to avoid overfit, especially when the polynomial order is bigger than the number of available data points.

4.2 Discrete Constraint Method

Given a model trained on historical data and partial data of target period, the goal is to produce a new model that fits current data and at the same time keeps consistent with historical model.

Recall that the use of curve for prediction is based on such assumption that the shape of the pipeline growth is correlated with the final value. In the regression model, the shape of a polynomial curve is jointly controlled by the parameter vector $\mathbf{w} = \{w_1, \cdots, w_k\}$. The change of its value affects the curve. However, such impact is different among elements w_i. For instance, consider a parabola represented as $y = w_1 + w_2 t + w_3 t^2$, the value change of w_1 will only move the curve vertically and has no impact on the shape. On the other hand, the change of w_2 and w_3 have different impacts, depending on the value of t. Specifically, if $|t|$ is bigger than 1, w_3 has larger impact. Otherwise, w_2 changer curve more than w_3. More generally, given an objective function $F(\mathbf{w})$, the impact of parameter on objective value is proportional to the absolute value its corresponding gradient, i.e., $\frac{\partial F}{\partial \mathbf{w}}$.

In last section, the enforcement term in Eq. (1) prefers small parameter value, i.e., close to zero unless there is strong evidence. Here we adopt a similar form. Instead of a continuous parameter σ, we put a binary constraint on each parameter. Specifically, parameters that have big impact on the curve shape are forced to be equal to historical model while those with small impact are free. The rationale behind is that historical data (or curve) determines the general shape of growth curve, while the current partial data makes small modifications.

To formulate this principle mathematically, we first define a constraint matrix, denoted as $\mathbf{C}$. It is a diagonal square matrix whose dimension is the

same with the order of polynomial curve. The element value has only two types, i.e., zero or positive infinity. Then given current data $\mathbf{Y}$ and historical curve $\mathbf{w}_0$, the objective function to combine the two can be written as below:

$$
\begin{aligned}
\mathfrak{L}_{dc}(\mathbf{w}|\mathbf{Y}, \mathbf{w}_0, \mathbf{C}) &= \mathfrak{R}(\mathbf{Y}, \mathbf{w}) + (\mathbf{w} - \mathbf{w}_0)^T \mathbf{C} (\mathbf{w} - \mathbf{w}_0) \\
&= \mathfrak{R}(\mathbf{Y}, \mathbf{w}) + (\mathbf{w} - \mathbf{w}_0)^T \begin{pmatrix} c_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & c_k \end{pmatrix} (\mathbf{w} - \mathbf{w}_0) \\
&= \mathfrak{R}(\mathbf{Y}, \mathbf{w}) + \sum_{i=1}^{k} c_i (w_i - w_{o_i})^2 \quad \text{where } c_i \in \{0, +\infty\}
\end{aligned}
\tag{3}
$$

As can be seen, when $c_i = 0$, the corresponding w_i goes with no constraint and is free for any value. On the other hand, when $c_i = +\infty$, there is such a strong constraint that w_i can only be equal to w_{0_i}.

The key challenge for this method is how to determine the value of c_i in constraint matrix $\mathbf{C}$. Firstly, without loss of generality, we assume elements of the parameter vector $\mathbf{w}$ are sorted in a non-descending order according to its impact on curve shape. Now the question is transformed to how to set the proper value for l. Heuristically if all data are available, then we do not need historical model ($l = 0$). On the other hand, if no data is available, we can only rely on the historical model ($l = |\mathbf{w}|$). However, it is unclear how the value to be set for a common case between these two scenarios. To solve this problem, we formulate it to an optimization task on historical data. Let $\mathbf{C}_l$ denote that the first l diagonal values are 0 while the other $k - l$ are set to $+\infty$. The objective function to determine the optimal l is defined as in Eq. (4).

$$
\Gamma(l|\mathbf{X}_1, \cdots, \mathbf{X}_M) = \frac{1}{M} \sum_{m=1}^{M} \{\mathfrak{R}(\mathbf{X}_m, \mathbf{w}^*)\}
$$
$$
\text{where } \mathbf{w}^* = \arg\min_{\mathbf{w}} \mathfrak{L}_{dc}(\mathbf{w}|\mathbf{X}_m^n, \mathbf{w}_0^{m-}, \mathbf{C}_l)
\tag{4}
$$
$$
\text{where } \mathbf{w}_0^{m-} = \arg\min_{\mathbf{w}} \frac{1}{M-1} \sum_{j=1, j\neq m}^{M} \mathfrak{R}(\mathbf{X}_j, \mathbf{w})
$$

Equation (4) defines a discrete search space for l and the space volume is equal to the number of parameters in the model. Specifically for polynomial curve fitting, it is an integer ranges from 1 to the polynomial order plus 1.

Algorithm 1 shows the specific steps. Given M historical data $\{\mathbf{X}_1, \cdots, \mathbf{X}_M\}$, number of available points n and polynomial order k, we take turns to remove one from the historical records and treat it as a pseudo "current data" with only n data points known (line 6). Then a historical model $\mathbf{w}_0$ is trained on the subset of historical data (line 7). After that, a new model $\mathbf{w}^*$ is obtained on partial data and $\boldsymbol{w}_0$ (line 8). Prediction error is evaluated on the full pseudo "current data" and is associated with corresponding parameter number (line 9).

Algorithm 1. Leave-one-out

1: **procedure** FINDL($\{\mathbf{X}_i | 1 \leq i \leq M\}, n, k$)
2:     $l^* \leftarrow 0, err^* \leftarrow +\infty$
3:     **for** $l \leftarrow 1$ to $k+1$ **do**
4:         $err \leftarrow 0$
5:         **for** $i \leftarrow 1$ to M **do**
6:             $\mathbb{D} \leftarrow \{\mathbf{X}_j | 1 \leq j \leq M \wedge j \neq i\}$
7:             $\mathbf{w}_0 \leftarrow \arg\min_{\mathbf{w}} E_{\mathbf{X}_j \in \mathbb{D}}(\mathfrak{R}(\mathbf{X}_j, \mathbf{w}))$
8:             $\mathbf{w}^* \leftarrow \arg\min_{\mathbf{w}} \mathfrak{L}_{dc}(\mathbf{w} | \mathbf{X}_i^n, \mathbf{w}_0, \mathbf{C}_l)$
9:             $err \leftarrow err + \frac{\mathfrak{R}(\mathbf{X}_i, \mathbf{w}^*)}{M}$
10:         **end for**
11:         **if** $err^* > err$ **then**
12:             $l^* \leftarrow l, err^* \leftarrow err$
13:         **end if**
14:     **end for**
15:     **return** l^*
16: **end procedure**

Finally the proper l is the one with the minimum error of prediction (line 12). We name this method *leave-one-out algorithm*.

5   Evaluation

In this section, we present the results of some experiments that we performed on real-world business data set to evaluate the performance of our proposed method. The evaluation is done based on the effectiveness metric; that is, the prediction accuracy on the target-period's predicted revenue. We next describe our data set and experimental setup.

5.1   Data Set Description

The data set we use consists of 11 periods of real business pipeline data. The period here refers to a fiscal quarter. For each period, we are given aggregated deal records whose target closing date is the end of the particular quarter. Each period spans 13 weeks before the target quarter starts and goes forward until 2 weeks after the quarter ends, making the total of 28 weekly data for each period.

    As mentioned earlier, a deal may go through several stages throughout its lifecycle in the pipeline. In the given data, each weekly record consists of total deal values for each stages (denoted as *current pipeline value*) as well as the value that turns out to be in won (denoted as *win contribution value*) by the end of the quarter. The prediction task is to predict, for any given week given this week's current pipeline value, the win stage's pipeline value of the final week.

5.2   Experimental Setup

From the 11 available quarters, we use the first 9 historical ones as training data and the remaining two as testing data. Testing is conducted for each weekly record, where only the current value of the pipeline is given. Note that the method we propose aims at predicting the growth curve of each week's win contribution value to the final win value. Thus, we apply two methods to estimate the win contribution value from the current pipeline value.

The first method we use is to compute the average rate of training data and then apply it to the testing data for prediction. Specifically, for a particular week, we check the training period data of the same week. For each stage, a conversion rate can be calculated via dividing the contribution value by the current pipeline value. Then the average rate is used to convert the testing period's pipeline current value to win contribution value. Finally, the DCM is applied to predict the period-end win value. The combination of this average ratio and our two models are denoted as *Avg-DCM*.

The second method, instead of computing the average, uses regression to predict the conversion rate. Specifically, to predict the conversion rate of a particular week, we fit a curve on the same week of all historical periods. Then, this curve is used to predict the next point, which is the conversion rate of the target period. This method combined with ours is denoted as *CF-DCM*.

Lastly, instead of using our models, we can purely apply curve-fitting in predicting the final win value. Firstly, we use curve fitting to predict the conversion rate as described above. Then, for each week, a growth rate can be calculated by dividing the final win value by that week's win contribution value. Now, applying the same curve-fitting technique, we can predict the growth rate of the target period. Finally, the sales win value is estimated by multiplying the current pipeline value by estimated conversion rates and growth rates. This method is denoted as *CurveFit*.

In evaluation, we use the relative error rate as our metric. Formally, let y denote the real value while y' denote the predicted value. The relative error is defined as $e_{\mathrm{rel}} = \frac{|y - y'|}{y}$.

5.3   Results

In this section, we present the experiment results. Particularly, two sets of experiments are conducted and reported. In first one, we aim at evaluating the performance of the proposed models in a scenario where the actual win-contribution value (with 100 % accuracy) is provided. In second case, the goal is to evaluate the performance in a scenario where the win-contribution value is predicted by some other method. We next discuss the two scenarios in more detail.

Scenario I: Win Rate Is Given. As mentioned in Sect. 1, the pipeline value forecast consists of two steps: (i) win-contribution prediction (conversion prediction) and (ii) win-value growth prediction. Our model focuses on the second part.

In this experiment, we use the real win-contributions and testify its performance in growth prediction.

There are four baselines used for comparison. The first one uses only the historical data and completely ignores the current data, denoted as *History-only (HO)*. The second baseline, denoted as *History-combine (HC)*, simply adds the new data into the historical one, and fits a curve on this merged data set. The third one, denoted as *History-replace (HR)*, replaces the old data of the same time unit with the newly available ones. Thus, the combined life cycle consists of new data by current week and historical data for unknown weeks. Then a polynomial curve is fitted. Finally, we apply the curve-fitting method, denoted as *Ideal-CurveFit (ICF)*.

Table 2 shows the average error rate for different departments. Because of data confidentiality, we anonymize the real department name, and call departments *Dept X* instead, where *X* is a number. As can be seen, our DCM method outperforms all other methods in all departments except for Dept 3. In dept 3, all history-based methods achieve the top-3 performance. In this department, although the target period has a similar final win value, its early data points are different, which misleads the DCM and let it give a wrong prediction. Nevertheless, the difference is quite small. Apart from the department's impact, we next explore the impact of time in the prediction accuracy. Figure 7 shows the error rate of all evaluated methods with regards to different times. As can be seen, the error of DCM decreases as time proceeds. This scenario can be expected and justified for two reasons. Firstly, when the time is closer to the period end, there is less fluctuations in the data, and thus it is easier to predict. Secondly, as time moves forward, more data is available, and thus the prediction can be improved. Also, we observe that the curve-fitting has an error curve far above other methods, and the error fluctuates a lot. This scenario suggests the vulnerability of curve fitting to data noise. Another observation is that the three history-based methods (HO, HC, HR) display a stable performance that is independent of time. As shown earlier in Sect. 3, the latter part of the curve is based on history only and therefore the prediction does not change too much as time goes forward.

Scenario II: Win Rate Is Predicted. In this experiment, we aim at evaluating its performance in the real scenario, i.e., do the forecast with estimated win-contribution/conversion values. We apply two methods for estimating the win-contribution values. First, we compute the average rate of training data and then apply it to the test data for prediction. Specifically, for a particular week, we check the training period data of the same week. For each stage, a conversion rate can be calculated via dividing the contribution value by the current pipeline value. Then, the average rate is used to convert the testing period's pipeline current value to win contribution value. The second method applies the curve-fitting to estimate the current one. Particularly, we denote the first method as *Avg-DCM* and the second one as *CF-DCM*.
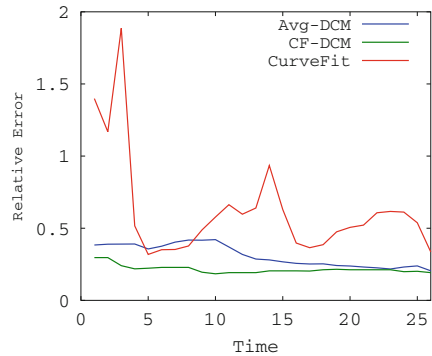
Table 2. Evaluation results in Win-Rate-Given Scenario

Department	Method				
	DCM	HO	HC	HR	ICF
Dept 1	**0.162**	0.292	0.292	0.293	0.241
Dept 2	**0.256**	0.423	0.423	0.422	1.120
Dept 3	0.421	**0.402**	0.403	0.408	0.835
Dept 4	**0.335**	0.384	0.384	0.387	0.804
Dept 5	**0.089**	0.227	0.227	0.226	0.132
Dept 6	**0.114**	0.200	0.200	0.201	0.313
Dept 7	**0.114**	0.143	0.143	0.142	0.487
Dept 8	**0.732**	1.094	1.094	1.100	0.837
Dept 9	**0.115**	0.240	0.240	0.239	0.395
Dept 10	**0.194**	0.416	0.416	0.416	0.231
Dept 11	**0.305**	0.689	0.690	0.692	0.305
Dept 12	**0.171**	0.284	0.284	0.285	0.396
Total	**0.251**	0.399	0.400	0.401	0.508

Table 3 displays the experimental results. For the sake of completeness, we also report the performance of DCM in Scenario I where the win rate is given (denoted as *WRG-DCM*) here. As can be observed, the CF-DCM method achieves the lowest error rate in all departments among all methods in the real scenario. The better performance of CF-DCM compared to Avg-DCM suggests the contribution of curve fitting in predicting conversion rates. Averaging over all departments, the Avg-DCM has a better performance than CurveFit, indicating the higher importance of growth prediction than conversion-rate estimates
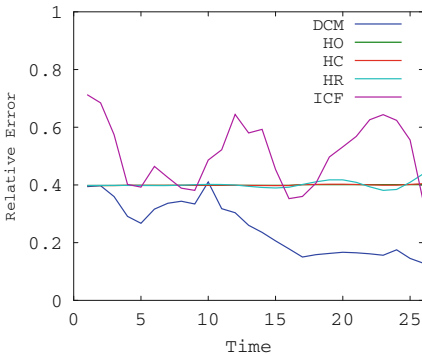
Fig. 7. Time impact in Win-Rate-Given Scenario

Fig. 8. Time impact in Win-Rate-Predicted Scenario

Table 3. Evaluation results in Win-Rate-Predicted Scenario

Department	Avg-DCM	CF-DCM	CurveFit	WRG-DCM
Dept 1	0.186	**0.105**	0.192	0.162
Dept 2	0.347	0.287	1.370	**0.256**
Dept 3	0.539	**0.315**	0.850	0.421
Dept 4	0.423	0.413	2.213	**0.335**
Dept 5	0.191	0.134	0.097	**0.088**
Dept 6	0.159	0.248	0.389	**0.114**
Dept 7	0.115	0.117	0.439	**0.114**
Dept 8	0.863	**0.352**	0.710	0.732
Dept 9	0.151	0.140	0.350	**0.115**
Dept 10	0.259	**0.150**	0.207	0.193
Dept 11	0.382	**0.224**	0.416	0.305
Dept 12	0.105	**0.101**	0.270	0.171
Total	0.310	**0.215**	0.625	0.251

in forecasting pipeline value. Another interesting observation is that the WRG-DCM outperforms the CF-DCM in half of our departments, and only achieves the second lowest error in all departments. A possible reason is the "noisy bump" in the data. In the records, the real win-contribution value may suddenly increase in some time and then decrease for the next time period. The can be attributed to either a data entry error or to the change of target closing date for some big opportunity/opportunities. In either case, such sudden jump misleads the ideal DCM, resulting in a higher prediction value and thus lets it end ends with a worse performance on average.

We also show the time impact on the forecasted performance in Fig. 8. Again, we can see the decreasing trend of all three methods with the proceeding of time. The CF-DCM shows the lowest error in the whole time line. We do observe that the CurveFit's relatively high error, especially at early stages. This demonstrates the robustness of our method in handling highly uncertain data.

6   Conclusion and Future Work

Services sales forecasting is different from traditional one as the pipeline is dynamic. At any given time point, some opportunities in the pipeline may reach an outcome (win or loss) and new ones may be identified and added. For more accurate prediction, a forecasting method should not only consider predicting the conversion/wining of current opportunities, but also need to predict the future pipeline growth. In this paper, we formulate the sales growth prediction as a dynamic curve fitting problem that combines historical data with the currently available data. The key challenge lies in how to combine the two aforementioned

data sets to capture future pipeline growth. We introduced a discrete-constraint method (DCM) that enforces similarity of the new model to the historical one by keeping some parameters fixed during learning. As our experiments showed, the DCM achieved best performance among multiple methods.

The current method treats all historical pipeline data equally. For future work, we may add different importance factors to different records simulating the possible seasonality. Alternatively, it may be a good approach to collect more data about the context and use the contextual similarity to determine such importance factors.

References

1. Arunraj, N.S., Ahrens, D.: A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting. IJPE **170**, 321–335 (2015)
2. Arunraj, N.S., Ahrens, D., Fernandes, M.: Application of SARIMAX model to forecast daily sales in food retail industry. IJORIS **7**(2), 1–21 (2016)
3. Aye, G.C., Balcilar, M., Gupta, R., Majumdar, A.: Forecasting aggregate retail sales: the case of South Africa. IJPE **160**, 66–79 (2015)
4. Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M.: Time Series Analysis: Forecasting and Control. Wiley, Hoboken (2015)
5. Bratina, D., Faganel, A.: Forecasting the primary demand for a beer brand using time series analysis. Organizacija **41**(3), 116–124 (2008)
6. Carman, S., Strong, R., Chandra, A., Oh, S., Spangler, S., Anderson, L., Bernard, J.J.: Predictive value of comments in the service engagement process. ASIST **49**, 1–6 (2012)
7. De Gooijer, J.G., Hyndman, R.J.: 25 years of time series forecasting. IJF **22**(3), 443–473 (2006)
8. Dugan, M.T., Shriver, K.A., Silhan, P.A.: How to forecast income statement items for auditing purposes. JBF **13**(2), 22 (1994)
9. Gajananan, K., Megahed, A., N.T., Abe, M., Smith, M.: A top-down pricing algorithm for IT service contracts using lower level service data. In: SCC. IEEE (2016, to appear)
10. Greenia, D.B., Qiao, M., Akkiraju, R.: A win prediction model for it outsourcing bids. In: SRII, pp. 39–42. IEEE (2014)
11. Hasin, M.A.A., Ghosh, S., Shareef, M.A.: An ANN approach to demand forecasting in retail trade in Bangladesh. IJTEF **2**(2), 154 (2011)
12. Kong, J., Martin, G.: A backpropagation neural network for sales forecasting. In: Neural Networks, vol. 2, pp. 1007–1011. IEEE (1995)
13. Lee, M.H., Hamzah, N.A.: Calendar variation model based on ARIMAX for forecasting sales data with Ramadhan effect. In: Proceedings of the RCSS, pp. 349–361 (2010)
14. Lee, W.I., Chen, C.W., Chen, K.H., Chen, T.H., Liu, C.C.: A comparative study on the forecast of fresh food sales using logistic regression, moving average and BPNN methods. JMST **20**(2), 142–152 (2012)
15. Luxhøj, J.T., Riis, J.O., Stensballe, B.: A hybrid econometric–neural network modeling approach for sales forecasting. IJPE **43**(2), 175–192 (1996)

16. Megahed, A., Gajananan, K., Abe, M., Jiang, S., Smith, M., Nakamura, T.: Pricing it services deals: a more agile top-down approach. In: Barros, A., Grigori, D., Narendra, N.C., Dam, H.K. (eds.) ICSOC 2015. LNCS, vol. 9435, pp. 461–473. Springer, Heidelberg (2015)
17. Megahed, A., Ren, G.J., Firth, M.: Modeling business insights into predictive analytics for the outcome of it service contracts. In: SCC, pp. 515–521. IEEE (2015)
18. Nezhad, H.R., Greenia, D.B., Nakamura, T., Akkiraju, R.: Health identification and outcome prediction for outsourcing services based on textual comments. In: SCC, pp. 155–162 (2014)
19. O'Donovan, T.M.: Short Term Forecasting: An Introduction to the Box-Jenkins Approach. Wiley, New York (1983)
20. Ramos, P., Santos, N., Rebelo, R.: Performance of state space and ARIMA models for consumer retail sales forecasting. RCIM **34**, 151–163 (2015)
21. Shukla, M., Jharkharia, S.: Applicability of ARIMA models in wholesale vegetable market: an investigation. IJISSCM **6**(3), 105–119 (2013)
22. Winters, P.R.: Forecasting sales by exponentially weighted moving averages. Manag. Sci. **6**(3), 324–342 (1960)
23. Yan, J., Zhang, C., Zha, H., Gong, M., Sun, C., Huang, J., Chu, S., Yang, X.: On machine learning towards predictive sales pipeline analytics. In: AAAI (2015)
24. Yin, P., Motahari Nezhad, H.R., Megahed, A., Nakamura, T.: A progress advisor for IT service engagements. In: SCC, pp. 592–599. IEEE (2015)