

A Prediction Algorithm for Airfare Based on Time Series

Xin Zheng¹

*School of Software Engineering, Beijing University of Posts and Telecommunications
No.10, Xitucheng Road, Haidian Dist., Beijing, 100876, China
E-mail: zxStar@bupt.edu.cn*

Kun Niu

*School of Software Engineering, Beijing University of Posts and Telecommunications
No.10 Xitucheng Road, Haidian Dist., Beijing, 100876, China
E-mail: niukun@bupt.edu.cn*

Junbang Ma^a, Qiaochu Li^b, Ziwei Zhang^c

*School of Software Engineering, Beijing University of Posts and Telecommunications
No.10 Xitucheng Road, Haidian Dist., Beijing, 100876, China
E-mail: ^abiatbang@163.com; ^b13261191060@163.com; ^c877649823@qq.com;*

Xiaoya Li

*School of Software Engineering, Beijing University of Posts and Telecommunications
No.10 Xitucheng Road, Haidian Dist., Beijing, 100876, China
E-mail: lixiaoya_01@bupt.edu.cn*

In order to solve the accuracy problem of the prediction algorithm for the lowest airfare in the future, this paper presents an United Intelligent Forecasting algorithm(UIF) consisting of two sub-algorithms, the Composite Weighted Time Series method(CWTS) and the Similarity Time Average method (STA) based on the idea of time series. CWTS calculates the sub-price on the target day from the quoted prices on the same day and a period of time in the past. STA calculates the sub-price on the target day from the prices on the similar period of time. Experiments on the real datasets show that UIF outperforms the traditional prediction algorithm and provides enhanced accuracy for airfare prediction. This airfare forecasting model based on time series can effectively solve the predictive conflict between sequences with smooth and fluctuating trends and thus a class of predictive analysis problems for the lowest airfare of air tickets at all kinds of time points are solved.

*CENet2017
22-23 JULY 2017
Shanghai, China*

¹This work was supported by Research Innovation Fund for College Students of Beijing University of Posts and Telecommunications; Special Found for Beijing Common Construction Project; The National Natural Science Foundation of China (No. 61272515, 61171102 and 61671081), National Science & Technology Pillar Program (2015BAH03F02).

1.Introduction

As the society progresses and the economy grows, the airplane is now becoming one of the most important forms of transportation[1]. While the airplane makes our life convenient, its fare always changes violently. Affected by holidays, low and peak seasons, supply and demand, *etc.*, seats of the same class in the same flight usually have different prices, which makes passengers hard to make decision when to buy; therefore, it is meaningful to analyze and predict the trend for the lowest airfare in the future. Up to now, many domestic websites provide airfare query function but cannot give a prediction. Some of foreign websites provide prediction services, such as FareCast[2], FareCompare[3] and Yapta[4], however, the prediction time is too long. Moreover, the difference between domestic and foreign air freight management models leads to the research blank in this field.

1.1Related Work

The time series is proposed as a solution to rank the airfare in a chronological sequence. It is an important concept that can reflect the developing process and direction usefully in prediction. Currently, the following time-series prediction algorithms [5,6,7] are in common use.

Weighted Moving Average method (WMA)[8,9] gives the data different weights according to the distance of time and then calculates the average. However, WMA can't keep up with the change when predicts with obvious trend of rising or falling. In addition, it will reduce WMA's sensitivity to fluctuations to increase the number of iterations. Because of these weakness, WMA only fits the situation with little change in a short-term trend.

The Exponential Smoothing method (ES) [10,11] calculates a series of exponential smoothing values to eliminate irregular changes and reflects the trend of time series. However, the result of ES is easily affected by the smoothing coefficient. That is, the sensitivity is high.

These methods above have their own shortcomings. They can only predict accurately in particular situations. There isn't a way to balance their advantages and disadvantages to reduce the sensitivity of the change of the fare. In this paper, we propose a united prediction algorithm based on WMA and ES to improve the accuracy of results by adjusting the proportional coefficient functions.

1.2 Our Contributions

Based on the existing methods, we propose an United Intelligent Forecasting algorithm (UIF), which consists of two sub-algorithms. Our main contributes include:

a. UIF is flexible and adaptive to both cases of smooth and fluctuating trend. The first sub-algorithm Composite Weighted Time Series method (CWTS) improves the stability of the prediction by adjusting the proportional coefficient function. The second sub-algorithm Similarity Time Average method (STA) added similarity time selection and special influence coefficient to improve the sensitivity of fluctuation points.

b. UIF raises the accuracy of forecasting result. Experiments on the real datasets show that UIF outperforms the traditional prediction algorithm.

c. UIF has low computational cost. It's able of dealing with massive instances synchronously.

2.United Intelligent Forecasting Algorithm

2.1 Statement of Problem

As the airfare is closely linked to time, the traditional method selects a time-series algorithm and adjusts its coefficients on the idea of time series. However, as described above in this paper, WMA is prone to "lag" for predictions of increasing or decreasing trend and ES is too sensitive to make prediction. In addition, these methods have very limited scope of application. The factors affecting the airfare are numerous and changeable but the traditional algorithms cannot be so thoughtful, which to a large extent affected the practicality of the prediction.

In order to make up for the defects, this paper proposes UIF. Firstly, we adjust the coefficient function of the WMA and modify its structure to enlarge the scale of the reference data. We propose the idea of similarity time based on the thoughts of K-Nearest Neighbor[12]. The data selected from similarity times and a special coefficient will make ES predict more accurate and reduce its sensitivity to smoothing coefficients. Finally, the two algorithms are combined to balance the advantages and disadvantages. As a result, UIF can be accurately predicted in both cases of smooth and fluctuating trend.

2.2 Composite Time Series

Assuming that the price of the next 7 days is predicted on day T, the user queries the price on day T, the composite time series used is the quoted prices for the next 7 days and the past week. As shown in the Table 1.

TD RD	+1	+2	+3	+4	+5	+6	+7
T	100	110	120	130	140	150	160
T-1	95	105	115	125	135	145	155
T-2	90	100	110	120	130	140	150
T-3	85	95	105	115	125	135	145
T-4	80	90	100	110	120	130	140
T-5	75	85	95	105	115	125	135
T-6	70	80	90	100	110	120	130
T-7	65	75	85	95	105	115	125

Table 1: Composite Time Series

TD indicates the target day. RD indicates the reference day. The figures are quoted prices of TD on RD.

2.3 Similarity Time

Assuming that the user queries the airfare of day (T+n) on Day T, the reference prices selected at similar time is: the actual minimum price of day (T+n-7), the quoted price of day (T+n) on day T and the quoted price of day (T+n+7) on day T.

2.4 Composite Weighted Time Series Method

The ticket pricing is affected by revenue management and has a lot to do with time. When the reference date is closer to the target date, the historical fares on the reference day are more associated with the fare of target day. According to this basic idea, CWTS makes prediction according to the trend of airfares in the continuous nearby past. We gave the first definition of CWTS here.

Definition 1: $\forall T$, T is the date of target day, and CWTS is defined as:

$$CWTS(T) = y'_T = \frac{w_1 y_{T-7} + \dots + w_8 y_T}{w_1 + \dots + w_8} = \frac{\sum_{i=1}^8 w_i y_{T-i+1}}{\sum_{i=1}^8 w_i} \quad (2.1)$$

y'_T is the forecast price of day T. $y_{T-7} \dots y_T$ are the quoted prices of the past week. w_n ($n=1, 2, \dots, 8$) is an increment function, that is, the closer to day T, the greater the weight.

2.5 Similarity Time Average Method

STA is proposed for the case of fluctuating trend. It makes prediction on the basis of similar times.

Definition 2: $\forall T$, T is the date of target day and STA is defined as:

$$STA(T) = y'_T = \frac{w_1 y_{T-7} + w_2 y'_T + w_3 y'_{T+7} + \Delta}{w_1 w_2 w_3} \quad (2.2)$$

y'_T is the forecast price of day T. y_{T-7} , y'_T and y'_{T+7} are the ticket prices at similar time. w_1 , w_2 and w_3 are the weights due to different time distances. Δ is the holiday (such as the Spring Festival and Mid-Autumn Festival, etc.) adjustment factor.

2.6 United Intelligent Forecasting Algorithm

CWTS and STA have the conditions of use. CWTS is a prediction method based on a continuous time, which is accurate when the fare is stable. In contrast, STA is based on the discontinuous similar time points and has high accuracy when it encounters holidays and emergencies. In order to balance these two different situations, we present UIF as follows.

Definition 3: we defined $w_1 + w_2 = 1$:

$$UIF(T) = w_1 CWTS(T) + w_2 STA(T) \quad (2.3)$$

Generally speaking, the initial value of the weights is 0.5:0.5. UIF adjusts them according to specific needs. For example, w_1 increases in the off-season, while w_2 decreases. The case is opposite on holidays.

3. Algorithm Description

3.1 Question Description

Assuming that T is the date of target day, we have known the lowest prices of the target day and the last 7 days (i.e. day (T-7) to day T) and the quoted prices to the next two weeks. Calculate the lowest forecast prices in the next 7 days (i.e. day (T+1) to day (T+7)). The key to this problem is to find a balance between stability and flexibility in the existing prediction algorithms and make our new algorithm prediction with the accuracy of 85% to 90%.

3.2 Steps and Flow Chart

The whole flow of UIF is shown as Figure 1, which consists of two parts, CWTS and STA.

a. CWTS calculates the weighted average CWTS(T) from the quoted prices in the past week according to Formula 2.1.

b. STA calculates the sum STA(T) of weighted average from similar times and holiday

adjustment factor according to Formula 2.2.

c. Forget the results of two sub-algorithms. UIF calculates the final prediction result UIF(T) according to Formula 2.3.

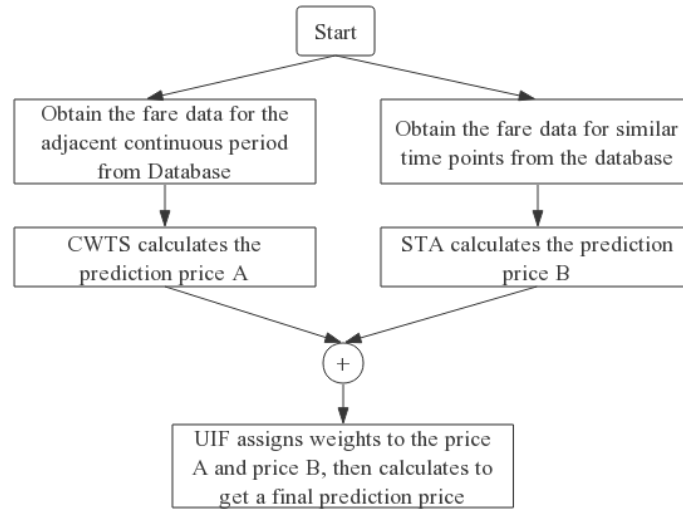


Figure 1: Flow Chart

3.3 Pseudo-code

Algorithm United Intelligent Forecasting (UIF)

1. while ($i < 7$)
 2. $CWTsum = CWTsum + price[today-i] * CWTweight[i]$
 3. $CWTweightsum = weightsum += CWTweight[i]$
 4. end while
 5. $CWT = CWTsum / CWTweightsum$
 6. $STAsum = STAweight[0] * price[today - 7] + STAweight[1] * price[today] + STAweight[2] * price[today + 7]$
 7. $STA = STAsum / (STAweight[0] + STAweight[1] + STAweight[2])$
 8. $result = CWT * CWTproportion + STA * STAproportion$
-

4. Experiences and Results

4.1 Datasets

The data used in the experiment is derived from the ticket sales website Ctrip [13]. It crawled at 11:00am every day from March 3, 2016 to March 11, 2017, including five different routes with different flights, 20 flights of Beijing-Chengdu route, 20 flights of Beijing-Xi'an route, 25 flights of Beijing-Guangzhou route, 24 flights of Beijing-Shenzhen route and 20 flights of Beijing to Qingdao route. A total of 273,154 sets of data were used, of which 218,523 sets (80%) were regarded as train set and the other 24,631 sets (20%) were regarded as the test set.

4.2 Results and Analysis

The experiment uses Root Mean Square Error (RMSE) [14] to measure the error between actual price and forecast result of UIF. The smaller the RMSE is, the less error will be.

Experiments show that UIF can get the most accurate results when CWTS uses the cubic inverse function $y = \frac{a}{x^3}$ as the coefficient function and the ratio of STA is 6: 1: 25. 'a' is a dynamic coefficient under different situations.

In order to test the prediction accuracy of UIF in both cases of smooth and fluctuating trend, we hereby show two sets of test results at different time periods: a month (13/ 1~13/2) before and after 2017 Chinese New Year (28/1) as the representative of fluctuating trends, 24/2/2017~10/3/2017 as a representative of smooth trends. We compared the RMSEs of the UIF with traditional predictive algorithms WMA and ES and RMSEs of the airfare at the same time last year (13/1 / 2016~13 / 2/2016). The results shown in Figure 2 and Figure 3.

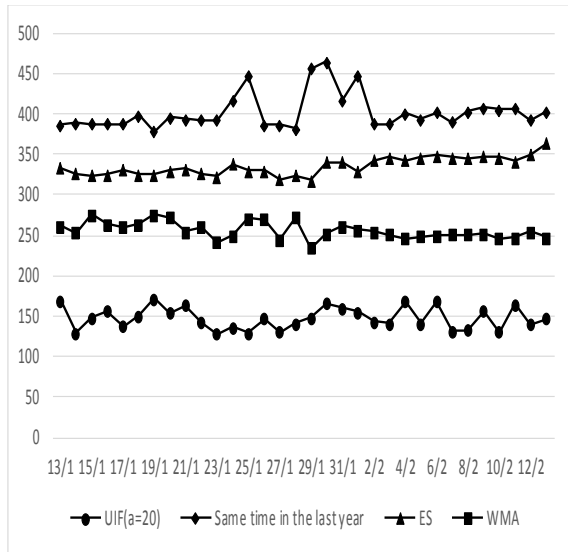


Figure 2: Comparison RMSEs in Case of Fluctuating Trend

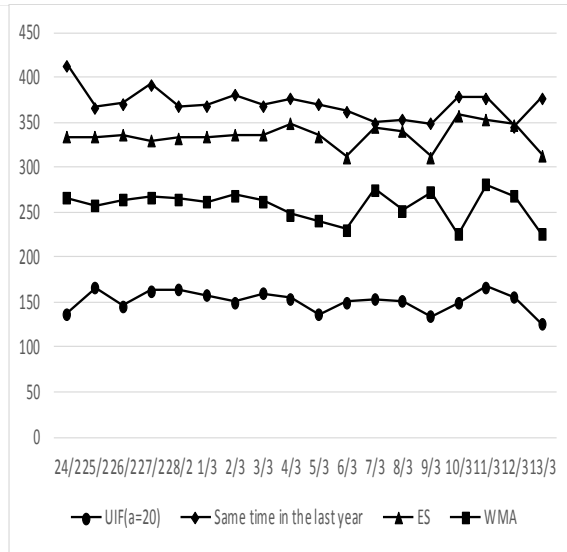


Figure 3: Comparison RMSEs in Case of Smooth Trend

Under the first situation, the differences between tickets in the adjacent dates are more than RMB106 with an average of RMB338.56, and the difference between maximum and minimum is RMB1830 for the whole time period. At this period, the minimum of UIF's RMSE is 127.95 and the maximum is 171.49. The accuracy of UIF is 42.25% higher than WMA's, 55.94% than ES and 63.26% than that in last year.

Under the second situation, the differences between tickets in the adjacent dates are less than RMB300 with an average of RMB106.88, and the difference between the maximum and the minimum is RMB510 for the whole time period. At this period, the minimum of UIF's RMSE is 125.59 and the maximum is 167.30. The accuracy of UIF is 41.21% higher than WMA's, 54.90% than ES's and 59.19% than that in last year'.

4.3 Time Complexity Analysis

The Time of this system is mostly used in traversing the database. The time complexity is $\theta(n)=n$. It's obvious that the complexity of the system is low; however, the request of hardware is high, which demands the high speed of query to handle the great amount of data.

5. Conclusion and Future Work

POS (CENet 2017) 095

This paper proposed a new algorithm UIF which consists of two sub-algorithms, CWTS and STA. UIF assigns the weights to the results of sub-algorithms and calculates the final forecast result. Experiments show that UIF improves the accuracy of the prediction. In addition, when comparing the experimental results of different flights, it shows that UIF works stably. However, like the existing majority approaches, it still features predictive defects. Taking into account more influencing factors, it is challenging the future work.

References

- [1] China.org.cn (2013.11.28) The survey shows that air plane is the preferred way to leisure travel [online] Available at: http://travel.news.cn/2013-11/28/c_125774107.htm (accessed 28 July 2016)
- [2] Picturepan2 (2014.4.9) Microsoft closed the function of ticket price forecasting in Bing Farecast [online] Available at: <http://livesino.net/archives/6787.live> (accessed 28 July 2016)
- [3] FareCompare. (2017). Cheap Flights to destinations around the world. [online] Available at: <http://www.farecompare.com> (Accessed 20 Mar. 2017)
- [4] Yapta, Inc. (2017). Airfare and Hotel Price Tracking – Save with Yapta. [online] Available at: <http://yapta.com> (Accessed 20 Mar. 2017)
- [5] Erdal K., Baris U, Okyay K.. *Grey system theory-based models in time series prediction* [J]. Expert Systems with Applications. 37(2), 1784-1789 (2014)
- [6] Shumway, R.. *Time Series Analysis and Its Applications* [M]. Springer, New York. pp,12-13 (2017)
- [7] R. Y. Zhao, Y. H. Zhu. *Discussion on Time Series Forecasting Method* [J]. Science & Technology Information,(15),192-193(2011) (In Chinese)
- [8] S. H. Qu, H. Li. *Application of Weighted Moving Average-PROMETHEE Method for Decision Making in Investing* [C]. Proceedings of the 2011 International Conference on Information ,Services and Management Engineering(ISME 2011)(Volume 2). Scientific Research Publish. pp, 231-234 (2011)
- [9] F. R. Meng, P. Zhuang, Q. Y. Yan. *Data Flow Forecasting Model Based on Weighted Moving Average* [J]. Application Research of Computers, (10),3680-3682+3686 (2009) (In Chinese)
- [10] D. H. Wu. *Dynamic exponential smoothing prediction method and its application* [J]. Journal of Systems &Management, (02), 151-155 (2008) (In Chinese)
- [11] E. Cadenas, O.A. Jaramillo, W. Rivera. *Analysis and forecasting of wind velocity in chetumal, quintana roo, using the single exponential smoothing method* [J]. Renewable Energy, 35(5), 925-930 (2010)
- [12] Mehmet Y., Seref S., Ilhami C.. *A new approach to very short term wind speed prediction using k-nearest neighbor classification* [J]. Energy Conversion and Management, 69, 77-86(2013)
- [13] Ctrip. (2017). Ctrip. [online] Available at: <http://www.ctrip.com> (accessed 25 Mar. 2017)
- [14] Chai, T., Draxler, R. R.. *Root mean square error (RMSE) or mean absolute error (MAE)?* [J]. Geoscientific Model Development Discussions, 7(1), 1525-1534 (2014)