ARIMA Model-Based Web Services Trustworthiness Evaluation and Prediction

Meng Li[1,2], Zhebang Hua[1,2], Junfeng Zhao[1,2,⋆],
Yanzhen Zou[1,2], and and Bing Xie[1,2]

[1] Software Institute, School of Electronics Engineering and Computer Science,
Peking University, Beijing 100871, China
[2] Key Laboratory of High Confidence Software Technologies, Ministry of Education,
Beijing 100871, China
{limeng09,huazb12,zhaojf,zouyz,xiebing}@sei.pku.edu.cn

Abstract. As most Web services are delivered by third parties over unreliable Internet and are late bound at run-time, it is reasonable and useful to evaluate and predict the trustworthiness of Web services. In this paper, we propose an ARIMA model-based approach to evaluate and predict Web services trustworthiness. First, we evaluate Web services trustworthiness with comprehensive trustworthy evidences collected from the Internet on a regular basis. Then, the cumulative trustworthiness evaluation records are modeled as time series. Finally, we propose an ARIMA model-based multi-step Web services trustworthiness prediction process, which can automatically and iteratively identify and optimize the model to fit the trustworthiness series data. Experiments conducted on a large-scale real-world data set show that our method can effectively evaluate and predict the trustworthiness of Web services, which helps users to reuse Web services.

1 Introduction

There are more and more reusable Web services available on the Internet. However, as most Web services are provided and/or hosted by third parties over unreliable Internet, it is useful to evaluate and predict the trustworthiness of Web services [1]. Many approaches have been proposed to evaluate the trustworthiness of Web services.

There are several problems with previous approaches. First, there are two kinds of trustworthy evidences: namely objective trustworthy evidences (e.g. QoS attributes) and subjective trustworthy evidences (e.g. reputation) [2, 3]. Most of previous approaches rely on only one kind of evidences, while it is more reasonable to take both kinds into consideration [3–5]. Second, while most previous approaches focus on the current or past trustworthiness of a service, it is reasonable and useful to evaluate and predict the trustworthiness of Web services in the (near) future [6, 7]. Third, previous prediction approaches like [8,

---

⋆ Corresponding Author

9] adopt collaborative filtering methods to make prediction, which are based on centralized portals to collect users' usage information. However, without proper incentive schemes, users are not motivated to provide feedbacks, and centralized portals like UDDI repositories are shut down or seldom updated [5].

In this paper, basing on classic time series theory, we propose an ARIMA model-based approach to evaluate and predict the trustworthiness of Web services. First, we evaluate Web services trustworthiness with comprehensive trustworthy evidences collected from the Internet on a regular basis. Then, the cumulative trustworthiness evaluation records are modeled as time series. Finally, we propose an ARIMA model-based multi-step Web services trustworthiness prediction process, which can automatically and iteratively identify and optimize the model to fit the trustworthiness series data. Experiments conducted on a large-scale real-world data set show that our method can effectively evaluate and predict the trustworthiness of Web services, which helps users to reuse Web services.

The contributions of this paper are as follows:

- We propose a novel and practical method to evaluate and predict Web services trustworthiness, which is based on classic time series analysis theories.
- The effectiveness of our approach has been validated with experiments conducted on a large-scale real-world data set. The results of our approach has been used in real systems, including CoWS[1] and TSR[2].
- The data set, including 17,832 Web services and millions of trustworthy evidences, is publicly available. This is the largest Web service trustworthiness evaluation and prediction data set to the best of our knowledge.

2   Related Work

2.1   Trustworthiness Evaluation and Prediction Methods

Many approaches have been proposed to evaluate Web services trustworthiness. E.M. Maximilien et al.[10] propose an decision theory-based approach. Z. Malik and A. Bouguettaya [11] propose a set of decentralized techniques to evaluate trustworthiness with users feedbacks. As most Web services are delivered by third parties over unreliable Internet, it is reasonable and useful to predict the trustworthiness of Web services in the future. Collaborative filtering-based prediction, e.g. [9] and [8], is a widely used prediction method. Another important category of prediction methods is based on time series analysis theories, which include [7] and [6], etc. There are several problems with previous approaches, which has been discussed in Sect. 1.

2.2   Time Series Analysis

Time series analysis have been widely applied in diverse fields including econometrics, mathematical finance, signal processing, etc. A time series can be defined

---

[1] http://www.cowebservices.com
[2] http://tsr.trustie.net

as a sequence of random variables indexed according to the order they are obtained in time [12]. Generally, the time series can be classified into two groups: stationary time series and non-stationary time series [12].

Time series prediction, also named as time series forecasting, is one of the most important parts of time series analysis. Different prediction strategies may be applied. Autoregressive moving average (ARMA) model and autoregressive integrated moving average (ARIMA) model are used to fit and forecast stationary and non-stationary time series respectively. ARIMA model can be converted to ARMA model by differencing.

3   Approach

There are three main steps in our approach: Internet-based trustworthy evidences collection, trustworthiness evaluation, and trustworthiness prediction.

3.1   Internet-Based Trustworthy Evidences Collection

We apply different strategies to collect comprehensive trustworthy evidences from the Internet. Both objective evidences (e.g. QoS attributes) and subjective (e.g. reputation) evidences are collected. We implement a third-party QoS monitor to collect objective evidences on a regular basis. To collect subjective trustworthy evidences, we apply both focused crawl and open search-based approach to collect users feedbacks including ratings, comments and relevant comment segments (RCS) [2]. These feedbacks are used to calculate a services reputation separately. For more details about trustworthiness evidences collection, we refer readers to our previous work [2, 13].

3.2   Trustworthiness Evaluation

We use a light-weight mathematical model to represent the collected evidences. All the trustworthy evidences of a service are represented as a vector. Then the trustworthy evidences of all Web services can be represented as a trustworthy evidence matrix (E):

$$E = \begin{pmatrix} e_{1,1} & e_{1,2} & \dots & e_{1,t} \\ e_{2,1} & e_{2,2} & \dots & e_{2,t} \\ \vdots & \vdots & \ddots & \vdots \\ e_{s,1} & e_{s,2} & \dots & e_{s,t} \end{pmatrix} \tag{1}$$

where each row of matrix E is a services trustworthy evidences, and each column is a kind of evidence.

The preferences of trustworthy evidences are represented as a weight vector $W = \langle w_1, w_2, \dots, w_t \rangle, \sum_1^t w_t = 1$. After normalization, the trustworthiness of each Web service can be evaluated as follows:

$$T = W^T * E^* \tag{2}$$

where E^* is the normalized trustworthy evidence matrix; W^T is the transformation of the weight vector; t_i is the i-th services trustworthiness within $[0, 1]$.

3.3   Trustworthiness Prediction

A service's cumulative history trustworthiness records are represented as a time series t_i. Then we propose an ARIMA model-based multi-step Web services trustworthiness prediction process which includes four main steps (Fig. 1).
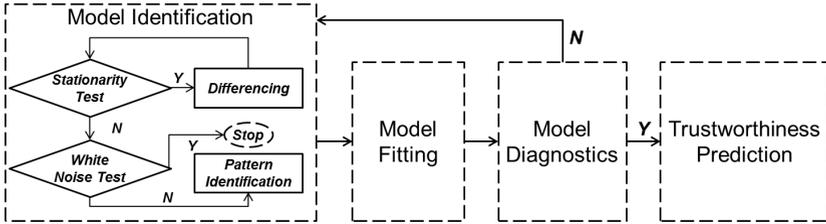

Fig. 1. Relationships of the steps of trustworthiness prediction

In **model identification**, we identify an appropriate time series model for a given trustworthiness series in the following sub steps: (1) *Stationarity Test*: Our method use sample *Autocovariance Function* (*ACF*) [12] to test whether a times series is a stationary series. (2) *Differencing*: If the series is identified as non-stationary in stationary test, it can be converted to a stationary series by differencing. (3) *White Noise Test*: We have to test whether the series is a white noise series, which is not worth any further analysis. The test is carried out using ACF. If the series is a white noise series, the ACF for any lag is always zero. (4) *Pattern Identification*: There are several specific patterns for ARMA model (shown in Table 1). By computing and comparing the ACF and PACF (Partial Autocorrelation Function) [12], it is possible to identify the exact pattern for the time series that passes the white noise test.

Table 1. Patterns for ARMA

	ACF	$PACF$
$AR(p)$	Tails off	Cuts off after lag p
$MA(q)$	Cuts off after lag q	Tails off
$ARMA(p,q)$	Tails off	Tails off

In **model fitting**, we try to find possible estimates of the unknown parameters within the identified model. For $ARIMA(p,d,q)$ model, there are three parameters to be determined: d is the number of differences to convert a non-stationary series into a stationary one; p and q can be determined by detecting the behaviors of ACF and PACF. For $ARMA(p,q)$ model, p and q are determined same as in $ARIMA(p,d,q)$ model.

In **model diagnostics**, we try to determine the best estimates of the parameters. First, the successive satisfying lags (10 at most) are used to construct

different ARMA or ARIMA models. Second, these models are applied back to the data itself. Third, *residual test* [12] is carried out to check whether a model is significant. If the residual series is a white noise series, the model is significant; otherwise, it is not. If all the models are not significant, we return to the model identification step. Finally, we use minimized sum of squares of errors to automatically determine the best estimates of the parameters from the remaining significant models.

In **trustworthiness prediction**, we use the finally accepted model to predict the trustworthiness in the near future for a given Web service. In this research, the trustworthiness values (with plus and minus one standard error) in next five days are predicted, which will be demonstrated in Sect. 4.

4   Evaluation

By applying the Internet-based trustworthy evidences collection method, we managed to collect trustworthy evidences for 17,832 public Web services. The evidences included over 2 million objective trustworthy evidences, and over 33,000 subjective trustworthy evidences, which is the largest Web services trustworthiness data set to the best of our knowledge. We conducted several experiments on this data set to validate the effectiveness of our approach.

4.1   Case Study

First, we use a real case to demonstrate the approach proposed and its effectiveness. CDYNE IP2Geo[3] Web Service is a Web service that resolves IP address to geographic information including Network Owner Name, City, State/Province, and Country. The result of trustworthiness evaluation and prediction is shown in Fig. 2. We can see that most of the history values fall in two standard errors and are very close to the predicted values, which demonstrates that our method can effectively model the Web services trustworthiness data and predict the trustworthiness values. Due to the length limitation of the paper, more details of the case study are available at CoWS[4].
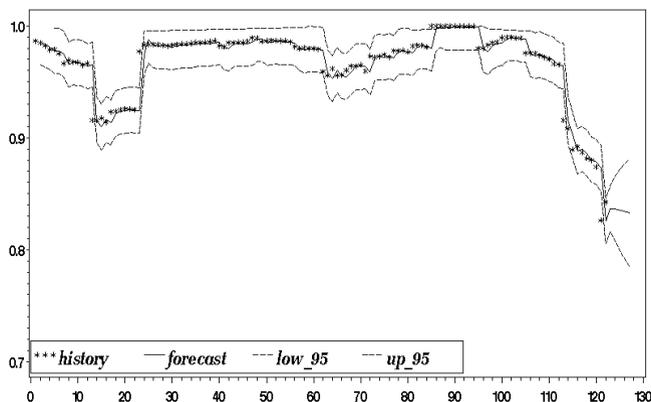
4.2   Trustworthiness Prediction Validation

In this subsection, we further studied the effectiveness of the method. We randomly selected 500 Web services from the total 17,832 collected Web services. Then we applied the prediction process on these Web services. The results showed that about 87.95% of the 500 sample Web services trustworthiness series were identified as non-stationary time series and ARIMA models were applied; the remaining 12.05% were stationary time series and ARMA models were applied. And most (about 99%) of the non-stationary trustworthiness series were converted to stationary ones by conducting differencing on the data once.

---

[3] http://ws.cdyne.com/ip2geo/ip2geo.asmx?wsdl
[4] http://www.cowebservices.com/CoWS/trust/trustIndex.jsp

Fig. 2. Case of Web service trustworthiness evaluation and prediction: abscissa axis represents the sequence of date; vertical axis represents the trustworthiness value; * marks real values; the solid (middle) line marks the predicted vales; the dashed (up and low) lines mark two standard errors.

About 96.5% of the determined prediction models were significant, and the values predicted with these models were acceptable.

4.3   Trustworthiness Application Validation

In this subsection, we validated whether the trustworthiness evaluation and prediction results were useful to help developers to reuse Web services. The latest predicted trustworthiness values were applied in CoWS to refine and rank the service candidates as follows:

$$Score_i = \gamma * sim_i + (1 - \gamma) * t_i \tag{3}$$

where sim_i is a Web services functional similarity to the users query [2]; t_i is the predicted trustworthiness value of the service candidate; $\gamma \in [0, 1]$ is the weight and assigned to 0.5 in this experiment.

We designed a web interface as shown in Fig. 3 to compare the results returned by CoWS with and without trustworthiness, and executed ten most common queries(listed in [2]) in CoWS. Fifteen graduate students with lots of experiences of reusing Web services were invited to manually explore and grade the results returned by CoWS for each query. The services were graded one of the following values: 2 for services similar to the search query and with high trustworthiness; 1 for services similar to the search query but with low trustworthiness; 0 for other services.

We used top-k Discounted Cumulative Gain (DCG_k) to measure a search engines ability to rank the good results before bad ones.

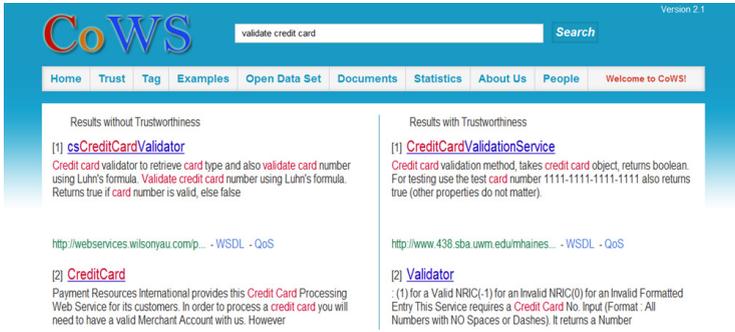$$DCG_k = gv_1 + \sum_{i=2}^{k} \frac{gv_i}{log_2 i} \tag{4}$$

Fig. 3. Web interface to comparing the results with and without trustworthiness

where gv_i is the value of a search result graded by the graduate students. Top-5 DCG, top-10, and top-15 DCG were calculated, and the results were shown in Fig. 4. From the results we can see that Web services with higher trustworthiness were ranked toper in the result list, which would help developers to select and reuse Web services.
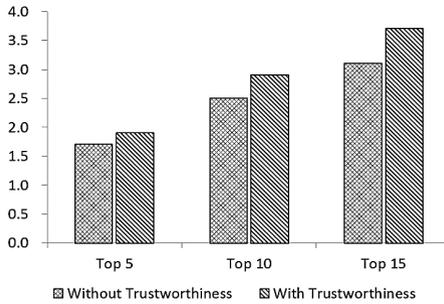

Fig. 4. DCG performance with and without trustworthiness

5   Conclusion

In this paper, we propose an ARIMA model-based approach to evaluate and predict Web services trustworthiness. Experiments conducted on a large-scale real-world data set show that our method can effectively evaluate and predict the trustworthiness of Web services, which helps users to reuse Web services.

In our future work, we will try to find and collect diverse kinds of trustworthy evidences, improve the trustworthiness evaluation and prediction method, and apply the trustworthiness in more scenarios.

References

1. Zhang, J., Zhang, L., Chung, J.: Ws-trustworthy: a framework for web services centered trustworthy computing. In: Proceedings of the 2004 IEEE International Conference on Services Computing, SCC 2004, pp. 186–193. IEEE (2004)
2. Li, M., Zhao, J., Wang, L., Cai, S., Xie, B.: Cows: An internet-enriched and quality-aware web services search engine. In: IEEE International Conference on Web Services, ICWS 2011, pp. 419–427. IEEE (2011)
3. Al-Masri, E., Mahmoud, Q.: Understanding web service discovery goals. In: IEEE International Conference on Systems, Man and Cybernetics, SMC 2009, pp. 3714–3719. IEEE (2009)
4. Mui, L., Mohtashemi, M., Halberstadt, A.: A computational model of trust and reputation. In: Proceedings of the 35th Annual Hawaii International Conference on System Sciences, HICSS 2002, pp. 2431–2439. IEEE (2002)
5. Zhang, Y., Zheng, Z., Lyu, M.: Wsexpress: A qos-aware search engine for web services. In: IEEE International Conference on Web Services, ICWS 2010, pp. 91–98. IEEE (2010)
6. Godse, M., Bellur, U., Sonar, R.: Automating qos based service selection. In: IEEE International Conference on Web Services, ICWS 2010, pp. 534–541. IEEE (2010)
7. Solomon, A., Litoiu, M.: Business process performance prediction on a tracked simulation model. In: Proceeding of the 3rd International Workshop on Principles of Engineering Service-Oriented Systems, pp. 50–56. ACM (2011)
8. Chen, L., Feng, Y., Wu, J., Zheng, Z.: An enhanced qos prediction approach for service selection. In: IEEE International Conference on Services Computing, SCC 2011, pp. 727–728. IEEE (2011)
9. Shao, L., Zhang, J., Wei, Y., Zhao, J., Xie, B., Mei, H.: Personalized qos prediction forweb services via collaborative filtering. In: IEEE International Conference on Web Services, ICWS 2007, pp. 439–446. IEEE (2007)
10. Maximilien, E., Singh, M.: Toward autonomic web services trust and selection. In: Proceedings of the 2nd International Conference on Service Oriented Computing, pp. 212–221. ACM (2004)
11. Malik, Z., Bouguettaya, A.: Rateweb: Reputation assessment for trust establishment among web services. The VLDB Journal 18(4), 885–911 (2009)
12. Shumway, R., Stoffer, D.: Time series analysis and its applications. Springer (2000)
13. Wang, L., Liu, F., Zhang, L., Li, G., Xie, B.: Enriching descriptions for public web services using information captured from related web pages on the internet. In: Fifth IEEE International Symposium on Service Oriented System Engineering, SOSE 2010, pp. 141–150. IEEE (2010)