Chapter 6

6.1 Summary

Web services are currently the most promising Service-oriented Computing (SoC) based technology. They use the Internet as the communication medium and open Internet-based standards, including the Simple Object Access Protocol (SOAP) for transmitting data, the Web Services Description Language (WSDL) for defining services, and the Business Process Execution Language for Web Services (BPEL4WS) for orchestrating services. Web service can operate under three architectures: single, composite, and community. In all these architecture, the potential of the Web services are often challenged by the presence of malicious services. However due to the lack of perfect information about the quality of the services they provide, there is always a question to which extent the client may trust the Web services to provide the required functionality, prior to their accesses. Traditional security mechanisms based on cryptographic approach do not go beyond the boundary of identity security. Trust and reputation mechanism based on social approach provide an alternative solution. Central to any trust model, prediction or forecasting of the trust value for the future time is of paramount importance.

In this study, based on the well-known social theory and properties of trust we propose a categorization of intermediaries. We also propose an effective method of prediction based on the sound mathematical foundation of Markov model, Gaussian process, non-linear time series principle and clustering. Using Markov chain and Hidden Markov Model, the job of trust prediction is reduced to a local regression problem. Regression is done by using Gaussian process regression framework. Using time series embedding principle and clustering technique, a trust or reputation time series is first reduced to regime transition network. Each regime represents a transient behavior pattern of the service provider. The Markov chain or the hidden Markov models then

identifies the most possible behavior pattern that the service provider might follow in the next interaction. Using the vectors defining this possible future regime, a local regression model in the form of a GP is trained and is then used to estimate the future value of trust or reputation.

The models' prediction accuracy has been assessed against some of the existing models using both synthetic and real life data. The prediction frameworks also have the ability to recover the trust value from the associated QoS attributes of Cloud Service dataset. Therefore, the models can be used in the prediction of missing values in in trust such datasets.

The study has also shown the advantage of the local model over the global model. Local model is found to have better ability to capture dynamic behaviour of a Web service. This fact is experimentally shown by the better accuracy rate of our Markov chain Local GPR and Coarse grained Hidden Markov Model GPR over the Markov chain global GPR. Further it was shown that Hidden Markov Models are more powerful tool to model time variant dynamic processes such as trust and reputation.

Gaussian process is a flexible non parametric tool for regression. However proper selection of its kernel is important for the accuracy. We proposed a method of ensemble of GP models based on model selection approach. Further by the simple mechanism of "Sleep and Recovery", we are able to reduce the computation by allowing the weaker model to sleep and recover only when the data patterns suit them. All proposed models are extended to better models using this ensemble framework. Using a common simulated environment, all the models are compared against existing state of the art model. The results have shown that ensemble based model works better in the prediction of trust values.

Existing reputation evaluation models are far from the concept of heterogeneous data reliabilities in which the reputation feedbacks are corrupted by different noise sources associated with the varying trustworthiness of the recommenders who supply the feedbacks. To address

this limitation, a method of handling variable noise infected feedbacks in the prediction of reputation trust is proposed. Using heteroscedastic Gaussian process, an effective model to deal heterogeneous data reliabilities in reputation evaluation is developed. In the model, the noise associated with a reputation feedback is related to a precision parameter. Due to this, reputation feedbacks from malicious sources are scaled down by high value of precision generally reported by them. Second, trust parameter i.e. trustworthiness is used to scale the feedbacks for the recommenders. This scaling has the ability to flexibly increase the noise around reports associated with untrustworthy users. Further, the model also takes care of the time varying nature of reputation using exponential decay principle.

A method based on hypothesis testing to filter malicious feedback reports has also been tested in the model.

Finally, we show that our method is more accurate than other exiting model with an extensive experimental evaluation on real-world dataset.

6.2 Future Work

Limitation of Gaussian process is that computing the matrix inverse in its prediction equations takes $O(n^3)$ time, making exact inference prohibitively slow for more than a few thousand data points. However, this problem can be addressed by cost effective approximations techniques of GP[157]. Extending our proposed models using some of these techniques is our future work.

The predictive distribution of a standard GP model is Gaussian. But the input data may not follow a Gaussian distribution, we may need to use non-Gaussian predictive likelihoods. Therefore experimentation with the non-Gaussian likelihood is another future work.

Our implementations are sequential implementations. Parallelization of the prediction from multiple GPs in ensemble prediction will be able to reduce the computational time. We aim to investigate this aspect in future.

The feasibility of our models proposed in Chapter 3-4 lay on the availability of the trust values for the selected time horizon. Such trust values can be retrieved by using a number of data collection models [80] – Publisher subscriber model, Community broadcast model and Credibility based model. Using these models, the data required to train our GPR can be obtained. However, runtime overhead involved in using any of these data collection models needs to be investigated in order to see how well our proposed models are scalable. This is another line of work in future.