Analysis (PLSA) to extract latent factors from WSDL service descriptions after the search is narrowed down to a small cluster using a K-Means algorithm. The PLSA model represents a significant step towards probabilistic modelling of text, it is incomplete in that it provides no probabilistic model at the level of documents [4]. The Latent Dirichlet Allocation (LDA) [4] is an attempt to improve the PLSA by introducing a Dirichlet prior on document-topic distribution.

Cassar et al. [6], [7] investigated the use of probabilistic machine-learning techniques (PLSA and LDA) to extract latent factors from semantically enriched service descriptions. These latent factors provide a model which represents any type of service's descriptions in a vector form. In their approach, the authors assumed all service descriptions were written in the OWL-S. The results obtained from comparing the two methods (PLSA and LDA) showed that the LDA model provides a scalable and interoperable solution for automated service discovery in large service repositories. The LDA model assumes that the words of each document arise from a mixture of topics, each of which is a distribution over the vocabulary. A limitation of LDA is the inability to model topic correlation [5]. This limitation stems from the use of the Dirichlet distribution to model the variability among the topic proportions.

The Correlated Topic Model (CTM) has been developed to address the limitation of LDA [5]. In CTM, topic proportions exhibit correlation via the logistic normal distribution. One key difference between LDA and CTM is the independence assumption between topics in LDA, due to the Dirichlet prior on the distribution of topics (under a Dirichlet prior, the components of the distribution are independent whereas the logistic normal models correlation between the components through the covariance matrix of the normal distribution). However, in the CTM model, a topic may be consistent with the presence of other topics. In this paper, we exploit the advantages of CTM to propose an approach for web service discovery and ranking. In our approach, we utilized CTM to capture the semantics hidden behind the words in a query, and the descriptions of the services. Then, we extracted latent factors from web service descriptions. The latent factors can then be used to provide an efficient discovery and ranking mechanism for web services.

V. Conclusion

In this paper, we have used several probabilistic topic models (i.e. PLSA, LDA and CTM) to extract latent factors from web service descriptions. The learned latent factors are then used to provide an efficient Service Discovery and Ranking. We evaluated our Service Discovery and Ranking approach by calculating the precision ($Pecision@n$) and normalized discounted cumulative gain ($NDCG_n$). The comparison of Precision@n and $NDCG_n$ show that the CTM performs better than the other search methods (i.e. LDA, PLSA and Text-Search). This reflects the accuracy of the ranking mechanism used by our method. The probabilistic methods based on CTM used the information captured in the latent factors to match web services based on the conditional probability of the user query.

Future work will focus on developing a new probabilistic topic model which will able to tag web services automatically.

References

[1] Abramowicz, W., Haniewicz, K., Kaczmarek, M. and Zyskowski, D.: Architecture for Web services filtering and clustering. In ICIW'2007.

[2] Alonso, G., Casati, F., Kuno, H., Machiraju, V.: Web Services - Concepts, Architectures and Applications. Springer Verlag, Berlin Heidelberg, 2004.

[3] Atkinson, C., Bostan, P., Hummel O. and Stoll, D.: A Practical Approach to Web service Discovery and Retrieval. In ICWS'2007.

[4] Blei, D., Ng, A. Y. and Jordan, M. I.: Latent dirichlet allocation. J. Mach. Learn. Res., 3:993-1022, 2003.

[5] Blei, D., and Lafferty, John D.: A Correlated Topic model of Science, In AAS 2007. pp. 17-35.

[6] Cassar, G., Barnaghi, P. and Moessner, K.: Probabilistic methods for service clustering. In Proceeding of the 4th International Workshop on Semantic Web Service Matchmaking and Resource Retrieval, Organised in conjonction the ISWC'2010.

[7] Cassar, G.; Barnaghi, P.; Moessner, K.: A Probabilistic Latent Factor approach to service ranking. In ICCP'2011, pp.103-109.

[8] Elgazzar, K., Hassan A., Martin, P.: Clustering WSDL Documents to Bootstrap the Discovery of Web Services. In ICWS'2010, pp. 147-154.

[9] Dong, X., Halevy, A., Madhavan, J., Nemes, E., Zhang, J.: Similarity Search for Web Services. In VLDB Conference, Toronto, Canada, pp. 372-383, 2004.

[10] Hess, A. and Kushmerick, N.: Learning to Attach Semantic Metadata to Web services. In ISWC'2003, Sanibel Island, Florida, USA, 2003

[11] Hofmann, T.: Probabilistic Latent Semantic Analysis. In UAI(1999), pp. 289-296.

[12] Kokash, N.: A Comparison of Web Service Interface Similarity Measures. Frontiers in Artificial Intelligence and Applications, Vol. 142, pp.220-231, 2006.

[13] Jurman, G., Riccadonna, S., Visintainer, R., Furlanello, C., Canberra Distance on Ranked Lists. In Proceedings of Advances in Ranking NIPS'2009 Workshop, 22-27.

[14] Lausen, H. and Haselwanter, T.: Finding Web services. In European Semantic Technology Conference, Vienna, Austria,2007

[15] Liu, Wei., Wong, W.: Web service clustering using text mining techniques. In IJAOSE'2009, Vol. 3, No. 1, pp. 6-26.

[16] Ma, J., Zhang, Y. and He, .J.: Efficiently finding web services using a clustering semantic approach. In CSSSIA'08, pp 1-8. ACM, New York, NY, USA.

[17] Maximilien, E.M., Singh, M.P.: Toward Autonomic Web Services Trust and Selection. In ICSOC'2004, pp. 212-221

[18] Nayak, R. and Lee, B.: Web service Discovery with Additional Semantics and Clustering. In IEEE/WIC/ACM 2007

[19] Porter, M. F.: An Algorithm for Suffix Stripping, In: Program 1980, Vol. 14, No. 3, pp. 130-137.

[20] Platzer, C., Rosenberg F. and Dustdar, S.: Web service clustering using multidimentional angles as proximity measures. ACM Trans. Internet Technol. 9(3), pp. 1-26 (2009).

[21] Rajendran, T., Balasubramanie, P.: An Optimal Broker-Based Architecture for Web Service Discovery with QoS Characteristics. In IJWSP'2010, Vol. 5, No. 1.

[22] Salton, G.: Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA (1989).

[23] Sivashanmugam, K., Verma, A.P and Miller, J.A.: Adding Semantics to Web services Standards. In ICWS'2003, pp: 395-401.

[24] Steyvers, M. and Griffiths, T.: Probabilistic topic models. In Latent Semantic Analysis: A Road to Meaning, T. Landauer, D. Mcnamara, S. Dennis, and W. Kintsch, Eds. Laurence Erlbaum, 2007.

[25] W3C (2004). Web services architecture. Technical report, W3C Working Group Note 11 February 2004.

[26] Xu, Z., Martin, P., Powley, W. and Zulkernine, F.: Reputation Enhanced QoS-based Web services Discovery. In ICWS'2007.

[27] Yu, Q.: Place Semantics into Context: Service Community Discovery from the WSDL Corpus. In ICSOC 2011, LNCS 7084, pp. 188-203.