Paper ID: 218

Improving Web Service Recommendation using Clustering and Model-Based Methods

LRMAM Weerasinghe[1#] and RAHM Rupasingha[2]

[1] *Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka,*
[2]*Faculty of Social Sciences and Languages, Sabaragamuwa University of Sri Lanka*

[#] lrmamweerasinghe@gmail.com

Abstract: With the development of the world wide web (WWW), the number of people who can deal with their work through the Internet, is increasing and it helps to do their tasks effectively and efficiently. In this case, a very important task is fulfilled by Web services. But the main problem is users struggling to select their favourite Web services quickly and accurately among available Web services. Web service recommendations help to solve this problem successfully. In this paper, we used collaborative filtering (CF)-based recommendation technique, but it suffers from the data sparsity and cold-start problem. Therefore, we applied an ontology-based clustering approach to overcome these problems. It effectively increased the data density by assuming the missing user preferences comparing the history of user favoured domains. Then, user ratings are predicted based on the model-based approach such as singular value decomposition (SVD). The result showed that the clustering approach can overcome the CF problems effectively and the SVD method can predict user ratings with lower prediction error compared with existing approaches.

Keywords: Web services, Recommendation, Collaborative filtering, Singular value decomposition, Sparsity, Cold-start

Introduction

Web services are software components that help to communicate between one computer to another computer through the Internet for data and information exchanging. In day by day, a large number of new Web services are coming to the Internet and that services support to develop some products or services. But most developers or users are struggling to find matching Web services. Therefore, the recommendation has a highly important place within the Internet users that discover services effectively and efficiently. In Web service recommendation that also suggests the best recommended services for users.

In the classification of recommendation (Adomavicius and Tuzhilin, 2005), there are three types of recommendation approaches such as content-based (Shoval, Maidel and Shapira, 2008), CF (Han et al., 2012),(Rupasingha and Paik, 2019) and hybrid filtering (Melville, Mooney and Nagarajan, 2002), (Li et al., 2010). The content-based approach contains some drawbacks (Li et al., 2010) and one of the main problems is that it needed users' personal information. And content-based techniques require the specification of a complete profile for each user/item. In CF core assumption is that users who have expressed similar interests in the past will share common interests in the future. The hybrid filtering is a combination of content-based and CF. Today one of the most successful and widely used methods in recommender systems is CF. The proposed approach is a CF and CF methods are divided into two types of recommendation techniques such as memory-based and model-based methods. Memory-based approaches use a database of past knowledge to infer the active user's preference for an item. We selected our CF approach as a model-based approach (Jia, Feng and Yu, 2010) that operate in two phases: the rating matrix is used in the offline phase to learn a compact configuration model for each user; the model is then used in an online phase to predict the user's level of interest in candidate items. Always use statistical and machine learning methods to learn the model from a dataset. Such algorithms can generate a recommendation that has the promise of being highly scalable while producing good predictive accuracy. The proposed approach is a model-based CF algorithm it presents using matrix factorization.

In any case, CF faced some major problems. Those are data sparsity, cold start, system scalability, synonymy, and shilling attack. In this approach, we focus on data sparsity and cold start problems. These problems arise because of insufficient prior transactions and available feedback data which makes it difficult to recognize similar users. There were many efforts to reduce the issues of sparsity and cold-starting. Transitive association-based methods (Shoval, Maidel and Shapira, 2008), clustering-based methods (Rupasingha and Paik, 2018a), (Rupasingha and Paik, 2019), (Melville, Mooney and Nagarajan, 2002) that decrease dimensionality by latent semantic indexing, binary preference-based methods (Ghauth and Abdullah, 2010), and correlation and cosine-based techniques (Han et al., 2012), give good performance for recommendations while reducing sparsity. In our approach, we used to resolve these recommendation issues by applying a clustering-based method that efficiently and effectively decreases the user rating dataset's data sparsity. Although there are several clustering methods (Rupasingha and Paik, 2015), (Rupasingha et al., 2016) available, we used a specificity-based novel ontology generation method as a clustering approach to classifying the service cluster groups (Rupasingha and Paik, 2019), (Rupasingha, Paik and Kumara, 2017). Because this approach showed the best performance of high precision, recall & F-measure rather than other existing clustering approaches and it successfully reduces the sparsity problem in the user service data-set.

After alleviating the sparsity using the clustering method, we used a model-based method to predict the user ratings. In model-based recommendation, we applied SVD for the CF process since it shows better performance when comparing with other model-based approaches. The SVD can capture the latent relationship between users and services. That allows us to compute the likeliness of a certain service by a user. Otherwise, SVD can produce a low dimensional representation of the original user-service space. Those predicted ratings from the SVD used for the recommendation.

We used mean absolute error (MAE) and root mean square error (RMSE) for calculating the accuracy of the predicted dataset and compared it with three types of sparsity levels. The proposed recommendation mechanism shows a lower error rate in MAE and RMSE comparing with existing methods.

This paper has been organized as follows. The second section is related work, the third section is the motivation, the fourth section is the overview of the proposed recommendation approach, the fifth section is experiments and evaluations, the sixth section is the conclusion and future work.

Related Work

A. Collaborative Filtering

CF (Han et al., 2012), (Rupasingha and Paik, 2019), is used to recommend a new service of interest for a particular user based on other users' opinions. This is depending on the user rating matrix analyzing. The assumption is that "users who adopted the same behavior in the past will tend to agree also in the future" (Nicola Barbieri, Giuseppe Manco, 2014).

CF approaches (Nicola Barbieri, Giuseppe Manco, 2014) may be classified specifically memory-based and model-based methods. Each class considers the preference matrix. However, memory-based methods infer the preference of the active user for an item by using a database of past knowledge. Among memory-based approaches, an outstanding role is contended by neighborhood-based methods (Hu, Peng and Hu, 2014), that are based on the definition of similarity between user or item. Memory-based approaches (Chen et al., 2010) are intuitive, as they directly rework hold on preference knowledge into predictions. The drawback is that they have access to the entire dataset to make recommendations, and, thus, they require specific indexing techniques, particularly when the scale of the data will increase. In contrast, model-based approaches (Chen et al., 2010) operate in two phases: within the off-line phase, the rating matrix is used to learn a compact personalization model for every user; then, the model is used in an on-line phase to predict the degree of interest of the user on candidate items. That approach (Nicola Barbieri, Giuseppe Manco, 2014) needs access to only a compact representation of the data. However, the recommendation provided to the user might not be simply explicable. A fundamental distinction also depends on the type of relationships among users and items that they're ready to exploit. Model-based approaches usually use dimensionality reduction techniques and therefore concentrate on the estimation of weak however global relationships.

Since the drawback of memory-based, we selected Probabilistic methods (Koren, Bell and Volinsky, 2009) which are the main focus of this manuscript, represent a refinement of the model-based approach, that relies on probabilistic modelling each in the learning phase and the prediction phase. In (Ramesh, Rao and Govardhan, 2017), show that model-based have different techniques Bayesian network, artificial neural network, and matrix factorization. In (Koren, Bell and Volinsky, 2009), visualized these techniques according to the MAE and RMSE. Through this result, it shows SVD generate high accurate result quickly. Therefore, we selected the SVD recommendation technique to continue our research.

B. Challenges of Collaborative Filtering

The CF used to data recommendations based on user-service ratings. Increasing the number of users and services help for some problems. If the dataset filled with null values, it becomes a reason for all users to do not interact with the rating mechanism. Therefore, the sparsity problem occurs. Otherwise, new users or items come to the system CF cannot recommend that. Therefore, cold-start problems occur. Because of these problems, it is challenging to find relations between users and services. Therefore, we want to make an effective recommendation.

1) Data Sparsity Problem:

In CF, sparsity is the main problem. That reduces the performance of the CF. To overcome this problem there are some existing researches. In (Rupasingha and Paik, 2019), address sparsity problem using the specificity aware ontology-based clustering and it improves recommendation performance. In (Chen, 2011), they fulfilled the sparsity problem using association retrieval technology and improve the recommendation result. In (Li et al., 2017), using the simplified method, that improved the performance and reduce the sparsity of the recommendation.

2) Cold Start Problem:

In cold start problem, there are no any user-service available ratings in their profiles. In (Bobadilla et al., 2012), they used CF for alleviating to the cold start problem. And in

(Ahn, 2008) they used heuristic similarity measure and then alleviate the cold start problem.

Users	Web Services							
	S1	S2	S3	S4	S5	S6	S7	S8
U1	1	0	1	0	0	0	0	0
U2	0	0	0	0	0	0	1	0
U3	1	0	0	1	0	0	0	1
U4	0	0	0	0	0	0	0	0
U5	0	0	0	1	0	0	0	0
U6	0	0	0	0	0	1	0	0
U7	0	1	0	0	0	0	0	1

Figure 20. User-Service rating matrix

Motivation

As discussed in the previous section, Web service recommendations face some problems such as data sparsity because users are not rating for services and cold-start problems because of new users and/or services coming for the recommendation. Since that case recommendation systems occur wrong predictions. Therefore, that system has low accuracy.

In Figure 1 explains 7 users and 8 Web services with their ratings. And it's all users do not interact with the rating system. So, if we have a large number of user-service ratings it is not easy to generate good accurate prediction results.

A. Example of Data Sparsity Problem

In Figure 1, let active U6. Its interesting service is S6. Then if we find who likes to S6, but no one likes it. Therefore, the system struggles to generate prediction results. This happens because of the data sparsity problem.

B. Example of Cold Start Problem

Let get U4 and S5, that are not active yet. Since that have not information about their relations the system cannot generate prediction results. It called a cold-start problem.

In Web service recommendation has above sparsity and cold start problems. So, it motivates us to find a method to overcome those problems. Because those problems occur wrong predictions and therefore, users waste their valuable time for searching their needs.

We found by comparing the existing recommendation research papers that SVD shows better results. Therefore, we wanted to resolve the above problems using clustering and then apply SVD for improving recommendation results

Overview of The Proposed Recommendation Approach

The architecture of the proposed approach is shown in Figure 2. In there, firstly collect the user-service data set and then get new data set by reducing sparsity using the ontology-based clustering method (Rupasingha and Paik, 2019). Then SVD is applied to the new data set for users' rating prediction. The recommendation is based on these prediction results.

A. Sparsity Alleviating

As shown in Figure 2, sparsity alleviates using the clustering is the first part. We used

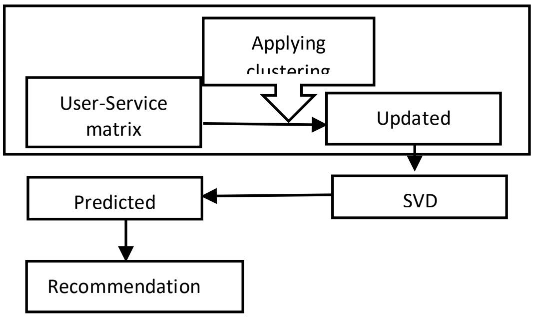

Figure 2. Architecture of the proposed method for recommending web services

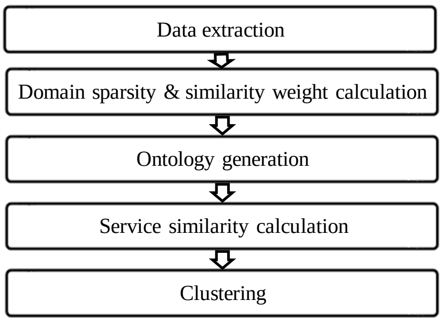

Figure 3. Ontology-based clustering method

a specificity-aware ontology generation-based clustering approach for (Rupasingha, Paik and Kumara, 2018) sparsity alleviating. Because it shows better clustering results rather than other approaches. In Figure 3, shows the process of ontology-based clustering.

For ontology generation, they used real Web services from the Web services repositories and Web Ontology Languages for Services (OWL-S) (http://projects.semwebcentral.org/projects/owls- tc/). It has Web Service Description Language (WSDL) documents and selected the five domains related to Book, Food, Film, Vehicle, and Medical. See Figure 3. Firstly data extracted as service name, operation

$$\hat{r}_{us} = \mu + b_u + b_s + q_s^T . p_u \qquad (2)$$

name, port name, input, and output messages form WSDL documents. Then do the domain specificity and similarity weight calculations and through this calculation generates the ontology. After calculating service similarity using this ontology finally got cluster results and grouped Web services into the above five clusters.

Using this clustering approach, we alleviate the sparsity (Rupasingha and Paik, 2019) as shown in Figure 4. Figure 4 explains the process of sparsity alleviating.

In the user-service rating dataset, we can see that 1 to 5 possible ratings and 0 means that user u not previously invoke with Web service s. then rus = 0.

$$r_{us} \begin{cases} r, & r = 1,2,3,4,5 \text{ if user u rated service s,} \\ 0, & \text{otherwise} \end{cases}$$

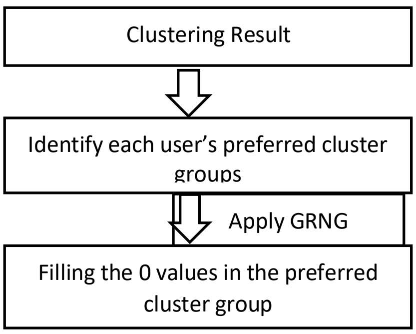

Figure 4. Sparsity alleviating process

We used this clustering result with Gaussian random number generation (GRNG) for filling the 0 values. Then, we used this result for rating prediction with the SVD method.

B. Rating Prediction

After sparsity alleviating then we improved our result using the SVD. The SVD technique is a simple matrix factorization method. We have selected it because the SVD (Solvang, 2017) is a good recommendation algorithm compare with other algorithms and it easily improves recommendation results.

In our SVD calculation the prediction $\hat{r}_{us}$ is set as:

μ – mean normalization

bu , bs - Observed deviation of user u, service s

If user u is unknown, then the bias bu and the factors pu are assumed to be zero. The same applies for service s with bs and qs .

To estimate all the unknown, we minimize the following regularized squared error

$$\sum_{r_{us} \in R_{Train}} (r_{us} - \hat{r}_{us})^2 + \lambda(b_s^2 + b_u^2 + \|q_u\|^2 + \|p_u\|^2)$$

The minimization is performed by a very straightforward stochastic gradient descent:

$$\begin{aligned} b_u &\leftarrow b_u + \gamma(e_{us} - \lambda b_u) \\ b_s &\leftarrow b_s + \gamma(e_{us} - \lambda b_s) \\ p_u &\leftarrow p_u + \gamma(e_{us}.q_s - \lambda p_u) \\ q_s &\leftarrow q_s + \gamma(e_{us}.p_u - \lambda q_s) \end{aligned} \qquad (1)$$

In there;
γ = learning rate
λ = regularization term
r_{us} = training dataset
e_{us} = error of training and predicted dataset
R_{Train} = set of the (u,s) pairs for which r_{us} is known

All the user ratings are predicted by the SVD and top ratings are used for service recommendation.

Experiments and Evaluations

We used Python as a language for SVD-based recommendation. For the implementation, we used the Surprise package (Hug, 2017).

First, we collect input user-service ratings. In this dataset, we simulated 200 users' ratings using 400 real Web services from real Web service repositories and OWL-S (http://projects.semwebcentral.org/project

$$Sparsity\ level = \frac{Number\ of\ Non\ specified\ ratings}{Number\ of\ all\ possible\ ratings} \quad (5)$$

s/owls- tc/). For this real Web services, we couldn't find real recommendation data. Therefore, we used a simulated dataset, it is already used in (Rupasingha, Paik and Kumara, 2018), (Rupasingha and Paik, 2018b). We could trust this dataset since, in their observation, they proved this simulated dataset have high accuracy with their evaluations. In their experimentation, they applied GRNG to generate the simulated dataset. That helps to easily evaluate our results. Evaluation based on MAE and RMSE. For sparsity alleviation and rating prediction, we used three sparsity levels that are 55%, 70%, and 85%.

MAE is measuring the deviation of the prediction and RMSE is the square root of the average of squared differences between prediction and actual observation. Better predictions visualized by the smaller MAE and RMSE values.

$r_{u,s}$ = training dataset user u, on service s
$\hat{r}_{u,s}$ = predicted dataset user u, on service s
s = Web services

$$MAE = \frac{1}{N}\sum_{s=1}^{N}\left|r_{u,s} - \hat{r}_{u,s}\right| \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{s=1}^{N}(r_{u,s} - \hat{r}_{u,s})^2} \quad (7)$$

N = number of predicted values

A. Evaluation Based on Sparsity Alleviating

We measure our recommendation performance using three sparsity levels (55%, 70%, and 85%). In Figure 5, that visualized number of non-rated values before and after alleviating sparsity.

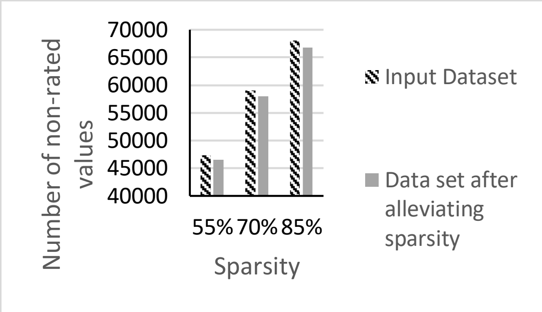

Figure 5. Number of non-rated values before and after alleviating sparsity

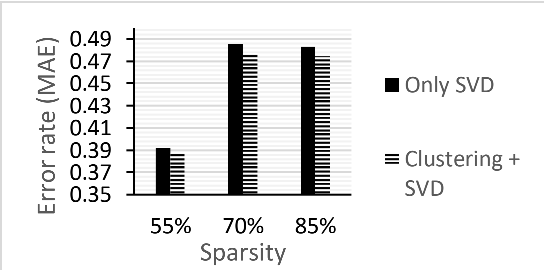

Figure 6. MAE for with and without clustering

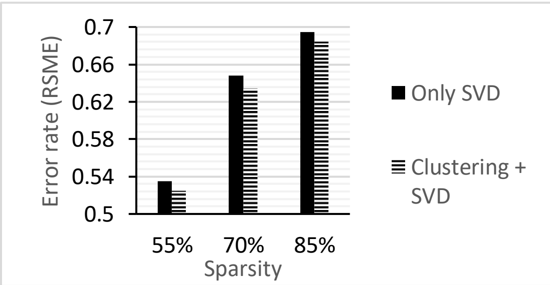

Figure 7. RMSE for with and without clustering

After alleviating the sparsity, we used SVD for rating prediction. Figure 6, and Figure 7, show that MAE, and RMSE values for prediction results with and without alleviating sparsity through the clustering method. That results also compared with three sparsity levels. However, it presents a lower error rate in SVD with a clustering method in a 55% sparsity level than other results. And if the dataset has a lower sparsity level, that generates a highly accurate result. Through this result we can suggest sparsity alleviating by clustering is successfully improve the recommendation performance.

B. Evaluation Based on Comparison with Existing Methods

We compare our method with the following three existing methods that are applied sparsity alleviation methods in recommendation.

Ontology-based clustering with Pearson correlation coefficient (PCC) (Rupasingha and Paik, 2019), They divided their Web services into five clusters and then applied GRNG for reducing to the sparsity. Then used PCC to improve to recommendation result. Association retrieval method (Chen, 2011), used the user's feedback data and calculate the relative distance between users' rating and similarity matrix then combined these and alleviating the sparsity problem. The binary method (Li et al., 2017), used a simplified similarity measure (SSM) method for fulfilling to sparsity problem. They convert user-item rating values into binary values, it helps to find similar users and therefore SSM, that speed up the process.

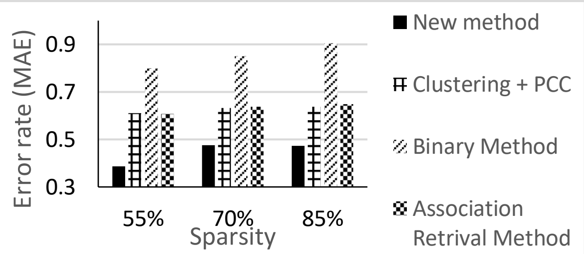

Figure 8. Evaluation based on comparison with existing methods (MAE)

We compared these methods and that shown in Figure 8 and Figure 9. According to the results, our proposed method shows better performance with minimum MAE and RMSE values. According to the evaluation result, using the ontology-based clustering result,

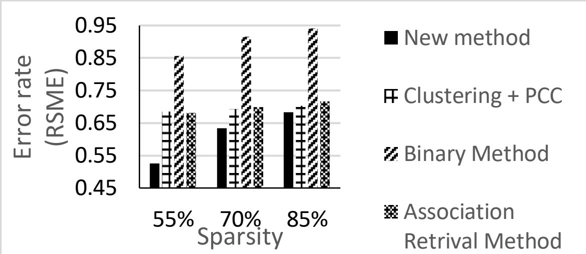

Figure 9. Evaluation based on comparison with existing methods (RSME)

we could successfully address the data sparsity problem. And in the cold start problem Web services identified by the clustering domain and if less than five ratings from users, we consider it as a cold start problem. We could overcome those problems successfully and finally compare with the existing methods we can show our approach improves the recommendation performance.

Conclusion and Future Work

Reason for the data sparsity and cold-start problems, existing Web service recommendation approaches have low performance. Therefore, our main objective is improving the recommendation result to reduce the lack of quality on the recommendation. In our proposed method we deal with these problems. We used ontology-based clustering for reducing the sparsity of the user-service rating matrix. Then we applied the SVD to rating predictions and it helps to improve the recommendation results. Through the clustering that reduces data sparsity and cold start problems successfully, and the lowest error rate verifies that our approach has high accuracy and best recommendation performance. In our future research, we aim to apply SVD++ for Web service recommendation and improve the recommendation performance.

References

Adomavicius, G. and Tuzhilin, A. (2005) 'Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions', IEEE Transactions on Knowledge and

Data Engineering, 17(6), pp. 734–749. doi: 10.1109/TKDE.2005.99.

Ahn, H. J. (2008) 'A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem', Information Sciences, 178(1), pp. 37–51. doi: 10.1016/j.ins.2007.07.024.

Bobadilla, J. et al. (2012) 'A collaborative filtering approach to mitigate the new user cold start problem', Knowledge-Based Systems, 26, pp. 225–238. doi: 10.1016/j.knosys.2011.07.021.

Chen, X. et al. (2010) 'RegionKNN: A scalable hybrid collaborative filtering algorithm for personalized web service recommendation', ICWS 2010 - 2010 IEEE 8th International Conference on Web Services, pp. 9–16. doi: 10.1109/ICWS.2010.27.

Chen, Y. (2011) 'Solving the Sparsity Problem in Recommender Systems Using Association Retrieval', 6(9), pp. 1896–1902. doi: 10.4304/jcp.6.9.1896-1902.

Ghauth, K. I. and Abdullah, N. A. (2010) 'Learning materials recommendation using good learners' ratings and content-based filtering', Educational Technology Research and Development, 58(6), pp. 711–727. doi: 10.1007/s11423-010-9155-4.

Han, X. et al. (2012) 'A Big Data Model Supporting Information Recommendation in Social Networks', in 2012 Second International Conference on Cloud and Green Computing. IEEE, pp. 810–813. doi: 10.1109/CGC.2012.125.

Hu, Y., Peng, Q. and Hu, X. (2014) 'A time-aware and data sparsity tolerant approach for Web service recommendation', Proceedings - 2014 IEEE International Conference on Web Services, ICWS 2014, pp. 33–40. doi: 10.1109/ICWS.2014.18.

Hug, N. (2017) Surprise, a Python library for recommender systems. Available at: surpriselib.com.

Jia, Z., Feng, Q. and Yu, J. (2010) 'A service model based on recommendation trust in pervasive computing environment', ICPCA10 - 5th International Conference on Pervasive Computing and Applications, pp. 393–397. doi: 10.1109/ICPCA.2010.5704134.

Koren, Y., Bell, R. and Volinsky, C. (2009) 'C OV ER F E AT U RE MATRIX TECHNIQUES FOR', pp. 30–37.

Li, L. et al. (2010) 'A contextual-bandit approach to personalized news article recommendation', Proceedings of the 19th International Conference on World Wide Web, WWW '10, pp. 661–670. doi: 10.1145/1772690.1772758.

Li, L. et al. (2017) 'A Simplified Method for Improving the Performance of Product Recommendation with Sparse Data', (iCAST), pp. 318–323.

Melville, P., Mooney, R. J. and Nagarajan, R. (2002) 'Content-boosted collaborative filtering for improved recommendations', Proceedings of the National Conference on Artificial Intelligence, pp. 187–192.

Nicola Barbieri, Giuseppe Manco, E. R. (2014) Probabilistic Approaches to Recommendations Synthesis Lectures on Data Mining and Knowledge Discovery. Available at: http://www.morganclaypool.com/doi/abs/10.2200/S00574ED1V01Y201403DMK009.

Ramesh, C., Rao, K. V. C. and Govardhan, A. (2017) 'Ontology based web usage mining model', Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2017, (Icicct), pp. 356–362. doi: 10.1109/ICICCT.2017.7975219.

Rupasingha, R. A. H. M. et al. (2016) 'Domain-aware web service clustering based on ontology generation by text mining', 7th IEEE Annual Information Technology, Electronics and Mobile Communication Conference, IEEE IEMCON 2016, (October). doi: 10.1109/IEMCON.2016.7746301.

Rupasingha, R. A. H. M. and Paik, I. (2015) 'Calculating Web Service Similarity using Ontology Learning with Machine Learning', 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC) (pp. 1-8). IEEE.

Rupasingha, R. A. H. M. and Paik, I. (2018a) 'Evaluation of Web Service Recommendation Performance via Sparsity Alleviating by Specificity-Aware Ontology-Based Clustering', 2018 9th International Conference on Awareness Science and Technology, iCAST 2018. IEEE, (September 2018), pp. 279–284. doi: 10.1109/ICAwST.2018.8517251.

Rupasingha, R. A. H. M. and Paik, I. (2018b) 'Improving Service Recommendation by Alleviating the Sparsity with a Novel Ontology-Based Clustering', Proceedings - 2018 IEEE International Conference on Web Services, ICWS 2018 - Part of the 2018 IEEE World Congress on Services. IEEE, pp. 351–354. doi: 10.1109/ICWS.2018.00059.

Rupasingha, R. A. H. M. and Paik, I. (2019) 'Alleviating sparsity by specificity-aware ontology-based clustering for improving web service recommendation', IEEJ Transactions on Electrical and Electronic Engineering, 14(10), pp. 1507–1517. doi: 10.1002/tee.22970.

Rupasingha, R. A. H. M., Paik, I. and Kumara, B. T. G. S. (2017) 'Improving Web Service Clustering through a Novel Ontology Generation Method by Domain Specificity', Proceedings - 2017 IEEE 24th International Conference on Web Services, ICWS 2017, pp. 744–751. doi: 10.1109/ICWS.2017.134.

RUPASINGHA, R. A. H. M., PAIK, I. and KUMARA, B. T. G. S. (2018) 'Specificity-Aware Ontology Generation for Improving Web Service Clustering', IEICE Transactions on Information and Systems, E101.D(8), pp. 2035–2043. doi: 10.1587/transinf.2017EDP7395.

Shoval, P., Maidel, V. and Shapira, B. (2008) 'An Ontology- Content-based Filtering Method', 15, pp. 303–314.

Solvang, M. L. (2017) 'Video Recommendation Systems Finding a Suitable Recommendation Approach for an Application Without Sufficient Data', (August).

Acknowledgment

This study was conducted under the Department of Computing and Information Systems, Faculty of Applied Sciences at the Sabaragamuwa University of Sri Lanka. We acknowledge all the people who were directly and indirectly help us to complete the research work successfully.

Author Biographies

L.R.M.A.M Weerasinghe is an undergraduate student at the Sabaragamuwa University of Sri Lanka. His research interest in recommendation systems.

R.A.H.M Rupasingha received her bachelor's degree in 2013 from Sabaragamuwa University of Sri Lanka. She obtained her master's and the PhD degree in 2016 and 2019, respectively from the School of Computer Science and Engineering, the University of Aizu, Japan. Her research interests include Web service clustering, ontology learning, and recommendation systems.