CHAPTER 6

CONCLUSION, FINDING AND SUGGESTION FOR FUTURE RESEARCH

6.1 INTRODUCTION

Major contribution and the important findings are brought out first. The performances exhibited through various experiments are summarized. Section 6.2 narrates the major contribution and important findings of this research work. The comparative study performed by the proposed works with existing works is summarized and the final conclusions have been drawn and presented in section 6.3. At the end, section 6.4 brings out the possible suggestions that can be worked on/extended from the present works are hinted out.

6.2 MAJOR CONTRIBUTION AND IMPORTANT FINDINGS

6.2.1 Major Contribution

As per the exhaustive literature studied and a part reported, it is found that data preprocessing is specific to the application and accordingly preprocessing is done on input Weblog file. The main contribution of this research is the clustering-based MART1 algorithm for Web pre-fetching; and MLFU and MPLFU algorithms for Web caching with pre-fetching using clustering-based MART1 algorithm. The proposed works have attempted to establish and endeavored to determine the apt replacement policies after experimenting for Web caching with Web pre-fetching. However, the
integrated Web caching and pre-fetching assures optimized page retrieval speed and reduced page access latency.

6.2.2 Major Findings

The delimitations of the proposed work as per the research objectives have clearly stated the need for considering server load. However, the server load has to be addressed separately for the proposed work. When the cache size increases, the hit rate also increases. The experiments show that the algorithms would work better for huge data.

Other conclusions and findings are narrated in the following sections.

6.3 CONCLUSION

The proposed work provides a complete approach for analyzing proxy server log file and extracting useful patterns in the form of access pattern. Using this pattern, frequently requested Web pages by the users were identified. The extracted pattern is then used in many applications, and one such application is enhancing the proxy-based Web caching system. Thus, the proposed system optimizes the existing Web caching system for better performance.

Since Web caching provides spatial locality and Web pre-fetching provides temporal locality, integration of these two would perform better. Therefore, this work integrates Web caching system with clustering-based Web pre-fetching technique to optimize the proxy-based Web caching system further. It also introduces a novel approach called MART1 for clustering-based pre-fetching technique. Empirical analyses were performed to show that the proposed system would outperform the traditional ART1 by using real
time datasets. From the results, it is observed that there is performance improvement in terms of the hit rate which ultimately reduces the user’s perceived latency. It is also inferred, that the accuracy of the prediction system could also be improved in terms of precision and recall when using MART1. From the obtained results, it is inferred that MART1 produces highly homogenous clusters by yielding higher average inter-cluster distance that is 0.991 while ART1 produces only 0.976. It is also observed that ART1 pre-fetching system only improves hit rate unto 22.5% whereas MART1 produces higher hit rate of 62.5%.

Finally, it also proposes different cache replacement policies namely MLFU and MPLFU to address the issues while integrating Web caching system with clustering-based Web pre-fetching system. It is evident that the proposed cache replacement policies suit the proposed system for improved network performance and reduced latency. From the results, it is observed that the proposed policies perform better than traditional policies like FIFO, LFU and LRU and it also produces higher hit rate and byte hit rate. Hence it saves bandwidth which ultimately improves the network performance and reduces congestion on server i.e. server load is controlled.

6.4 SUGGESTION FOR FUTURE STUDY

1. As the user datasets containing the privacy information should not be exposed to the outside world, privacy preserving data mining techniques can be applied in order to hide personal information about the users.

2. Evolutionary optimization techniques can be applied in order to optimize the system further.
3. A session-based approach can also be adopted to improve bandwidth utilization.

4. A hybrid approach out of the existing algorithms can be tried out for cache replacement process.

5. This system can also be implemented to extend the performance of Content Delivery Network (CDN) server and Enterprise Resource Planning (ERP) system for effective content distribution and decision making process.

6. This system can also be used in the application areas where Web search, access and retrieval are involved, such as Predicting user purchase pattern of commodities in E-Commerce Website, Redesigning of a site according to user interest.

7. Instead of static HTML and XML files considered offline might be extended to other Web related file formats also.