CHAPTER 7

CONCLUSION AND FUTURE WORK

This thesis presented the design and implementation of Context aware web (WWW and deep web) search engines that were enhanced using Nature Inspired Algorithms. The work also tackled the shortcomings of existing nature inspired algorithm namely shuffled frog leaping algorithm (SFLA) and presented two extensions as ISFLA and SRSFLA. This chapter highlights the contributions and presents future directions of the work.

7.1 Summary of the Thesis Contributions

This thesis presented a unified framework of Contextual Web Search using Nature Inspired Algorithms. The main contributions of the framework are divided into three phases. The first phase presented a novel model of Contextual Web Search system based on Belief, Desire and Intention (BDI) MultiAgent architecture. The model employed team of agents which cooperate and coordinate with each other to understand the users’ immediate contextual desires and provide context relevant web urls from both the surface and deep web. The presented model is neither too interactive to lose the user and nor completely automatic to overlook the users instant needs. It provides an appropriate level of interaction to identify the users’ instant context for retrieving context relevant surface and deep web urls. To fetch the contextual web (surface and deep web) urls from WWW, a distinct meta-crawler is developed using shuffled frog leaping algorithm (SFLA). This approach overcomes the limitations of conventional web crawlers to perform broad search before getting the desired contextual web urls.

The presented SFLA based surface and deep web meta-crawlers were assessed over diverse contextual domains such as Health, Computer, Travel and Consumer Electronics. Studies established that employment of nature inspired algorithm namely Shuffled frog leaping algorithm to develop the meta-crawler enhances its performance by more than 50% in all contextual domains. Analysis performed with respect to Genetic algorithm (GA) portrayed
that both the mean and maximum precision of SFLA based meta-crawler show steady improvement over the generations while precision of GA based meta-crawler fluctuates over time. SFLA based contextual meta-crawlers converge faster and attain gain in precision up to 16% and generate up to 22% more diverse web urls as compared GA.

Hence, this phase contributes a distinct methodology to understand the users instant need context and efficient techniques to populate the contextual web (surface and deep web) url repositories. The presented techniques have a better potential for serving diverse information needs of the users.

The next phase focussed on retrieval of contextual deep web content from the deep web search systems. The widespread techniques used to develop the deep web search systems include content based filtering and social context based filtering. The presented work primarily aimed to overcome the inherent limitations (cold-start, scalability and sparsity) of these systems using nature inspired algorithms. Subsequently, trust and context aware approaches were developed to further improve their accuracy.

In content based deep web search, Genetic Algorithm (GA)-Case Based Reasoning (CBR) approach was presented to establish the potential of nature inspired algorithm for solving cold-start problem. The GA-CBR hybrid approach demonstrated promising performance with respect to conventional CBR systems.

For social context based deep web search systems, work presented nature inspired algorithms namely memetic algorithm (MA) and shuffled frog leaping algorithms (SFLA) to tackle the scalability problem. Experimental studies performed over MovieLens and Epinions benchmark datasets established that MA based approach performs better than K-means and Genetic algorithm (GA), over precision, recall, f1 and mean absolute error (MAE) metrics. SFLA based social filtering technique outperforms both MA and GA over MAE metric. SFLA based approach was able to reduce the MAE by approx 23% with respect to GA and upto 19% with respect to MA.

To deal with the cold-start and sparsity problems in SFLA based social filtering systems, an approach to incorporate user’s personal ‘Web of Trust’ was presented. Studies
established that presented trust aware SFLA based approach further enhances the accuracy of social context based deep web search systems by up to 45%. Results also established gain in enhancement with the increase in number of active users. It leads to the conclusion that integration of interests and preferences of user’s true friends boosts the generation of more relevant content.

After tackling the intrinsic limitations, two distinct context aware approaches were developed. The techniques incorporate user’s personal demographic context such as gender, occupation or combination of both in SFLA based social filtering systems at different stages. It was demonstrated through experiments that SFLA based contextual two dimensional approach (SC2D) improved accuracy by up to 5% and SFLA based contextual three dimensional (SC3D) technique attained significant gain of up to 20% as compared to non-contextual SFLA based approach. These results emphasize that personal demographic context is an important factor to determine the interests and preferences of the users. Hence, presented SFLA based demographic context aware deep web search systems may provide better search experience.

In the final phase limitations of conventional shuffled frog leaping algorithm were tackled by developing the two modified forms of SFLA; Improved Shuffled Frog Leaping algorithm (ISFLA) and Self-Reformed Shuffled Frog Leaping Algorithm (SRSFLA). The algorithms were evaluated over benchmark test functions for small as well as large dimensions using a distinct toolbox, SEVO toolbox. Swarm and Evolutionary optimization (SEVO) toolbox was developed using Java platform. It supports simulation of four established algorithms (GA, MA, PSO and SFLA) and sixteen benchmark test functions. SEVO toolbox carries all the advantages of java platform and may be extended further to include more algorithms and test problems. This contribution would save substantial time and efforts of the research community for simulating established nature inspired algorithms.

The presented improved shuffled frog leaping algorithm(ISFLA) was evaluated (using SEVO toolbox) with respect to GA, MA, PSO and SFLA over continuous uni-modal and multimodal test functions for dimensions (d) = 100, 500 and 1000. Results indicated that ISFLA significantly intensifies the global minima as compared to genetic, memetic and particle swarm optimization algorithms. ANOVA test further corroborated the statistical
significance of the obtained results. To evaluate the strength of SRSFLA (self-reformed shuffled frog leaping algorithm), a comparative study was performed with respect to MSFLA-EO, MLCC, EPUS-PSO, JDEdynNP-F, DewSAcc, DMS-PSO, LSEDA-gl, UEP, ALPSEA, classical DE (differential evolution) and real coded CHC algorithms that participated in ‘CEC 2008’ competition. SRSFLA outperformed these algorithms for small as well as large dimensions. SRSFLA significantly augmented the exploration capability and speed up convergence to global optima. SRSFLA also performed better than SFLA and genetic algorithms in contextual web search. Hence, both the presented algorithms (ISFLA and SRSFLA) were found to be more effective against its contemporaries over benchmark test problems and in retrieving context relevant web documents.

Thus, the various contributions of this thesis made in contextual web search and nature inspired algorithms may serve as better choice for the users with diverse contextual information requirements and for solving high dimensional problems with multiple optimal solutions.

7.2 Directions for Future Work

The work done in this thesis has good scope for further extensions along many dimensions.

- The working of deep web content extractor agent can be extended to fetch content from other deep web portals along with digital libraries. An automated approach can be developed to understand and formulate the query structures of varied deep web systems.

- The GA-CBR hybrid approach can be enhanced further by developing SFLA-CBR approach. This approach can also be extended for multi-objective optimization. The approach can be applied in varied real world applications for solving cold-start problems.
• The developed SFLA based trust and context aware social filtering approaches can be applied to social networks for recommending friends, in e-commerce websites for suggesting products etc.

• The capabilities of SEVO toolbox can be enhanced further to make it suitable for the research community. It can be extended to include new nature inspired algorithms and other benchmark test problems.

• The work can be extended to employ presented Self-reformed shuffled frog leaping algorithm (SRSFLA) for multi-objective optimization and evaluate its efficacy over other benchmark problems e.g. TSP, QSAR etc
Research Publications

International Journals


International Conferences


