# Swarm Intelligence Based Optimization for Web Usage Mining in Recommender System

Manisha Sajwan
DIT University
Dehradun, India

Kritika Acharya
DIT University
Dehradun, India

Sanjay Bhargava
DIT University
Dehradun, India

**Abstract**: Nowadays, the web has become one of the most effective and efficient platform for information change and retrieval .Due to heterogeneity and unstructured nature of the data available on the WWW, web mining uses various data mining techniques to discover useful knowledge from web hyperlinks, page content and usage log. This research introduces the theoretical foundations of Swarm Intelligence and Design, implementation of swarm optimization algorithm. The Swarm Intelligence optimization and data mining technique can be used together to form a method which often leads to the result. Design and implementation of a web mining system based on multi-agents technology will reduce the information overload and search depth. This is helpful to users using the web within a platform for e-commerce or e-learning.Swarm Intelligence is an efficient technology that deals with natural and artificial system. It provides an efficient way for finding optimal solution. During the past few decades researches are trying to use these techniques to solve many problems in various fields. Recommender System is the one of the most important application of e-commerce and it plays vital role in understanding the user's behaviour or interest by which it increases the profit of sales or usage of services of website. This paper describes a swarm intelligence optimization for web mining to find the optimal solution and based on that process is done.

**Keywords**: Swarm Intelligence Optimization, Natural Inspired Technique, Web Mining, Recommender System, E- Commerce, Target Marketing.

## 1.  INTRODUCTION

Swarm Intelligence is the collective behaviour of natural and artificial system. The concept is employed by Gerado, Beni and Jing Warm in 1989 in the context of cellular robotics system. Natural Inspired Computing Techniques such as swarm intelligence has the ability to solve many combinatorial problems and provide optimal solution. Swarm Intelligence deals with natural and artificial system that provides an effective and an efficient way for finding optimal solution. The combinational environment of natural and artificial system is called decentralized system. In this type of environment user has difficult to find the optimal solution so, this can be overcome by recommendation process. By using this process we can find nearest neighbour with users of similar likes and dislikes and there for we can generate an effective recommendation.

Data Mining and Swarm Intelligence may seem that they do not have many properties in common. However the combinational approach of data mining and swarm intelligence can be used to generate optimal result, even other methods or approaches would be too expensive or difficult to implement.

This paper focuses on two parts. First part describes Swarm Intelligence Based Web Usage Mining which describes a study on collective behaviour of ant, bees etc, mining techniques and Knowledge Representation and Discovery. Second part describes the user profile are selected based on neighbourhood usage of Swarm Intelligence.

**This paper is organised as follows:**
Section 2 outlines a deep study on swarm intelligence and knowledge representation and discovery. Section 3 outlines the overview of Swarm intelligence algorithms. Section 4 describes web usage mining. Section 5 describes Recommender System, E-Commerce and Target Marketing.

Section 6 outlines Web Page Recommendation using Swarm Intelligence.

## 2.  SWARM INTELLIGENCE, KNOWLEDGE REPRESENTATION AND DISCOVERY

Swarm behaviour can be seen in bird flocks, fish schools, as well as in insects like mosquitoes and midges. Swarm is an aggregate with cohesion, but a low level polarization (parallel alignment) among members. Swarm Intelligence (SI) is an innovative distributed intelligent paradigm for solving optimization problems that originally took its inspiration from the biological examples by swarming, flocking and herding phenomenon in vibrates.[7],[13]

**Main principles of collective behaviour-**

**Homogeneity-** Every bird in flock has the same behaviour model. The flock moves without a leader, even though temporary leaders seem to appear.

**Locality**- The motion of each bird is only influenced by its nearest flock mates. Vision is considered to be the most important senses for clock organization.

**Collision Avoidance-** Avoid with nearby flock mates.

**Velocity Matching-** Attempt to match velocity with nearby flock mates.

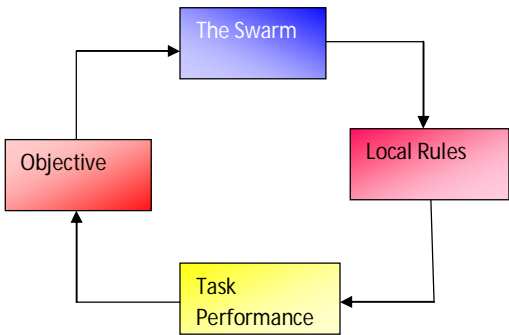**Flock Centring-** Attempt to stay close to nearby flock mates.

Figure 1. Simple Scheme of a Swarm

Since 1990, several collective behaviour inspired algorithms have been proposed. The application areas of those algorithms refer to well studied optimization problem like NP-hard problems(Travelling Salesman Problem, Quadratic Assignment Problems, Graph Problems), Network Routing, Clustering, Data Mining, Job Scheduling, etc.

## 2.1 Knowledge Representation

Knowledge Representation is necessary to solve a real word problem using large amount of data by manipulating that knowledge.

Before going through Knowledge Representation And Discovery, we have to focus on two entities.

1-Fact (Universal truth)

2-Representation of fact in some chosen form

Predicate Logic is the best method to represent knowledge. To use predicate in our application, we can implement an interface predicate in our programming like in java. This interface provides the framework for filtering of database.

## 2.2 Knowledge Discovery:

The following model as shown in figure 2 & 3 presents how to discover knowledge from the large set of information. The knowledge discovery process seeks new knowledge in same application domain. It is defined as a non-trivial process of identification of valid novel potentially useful and ultimately understandable patterns in data. It consist of many step each attempting to complete a particular discovery task and each accomplished by the application of discovery method.
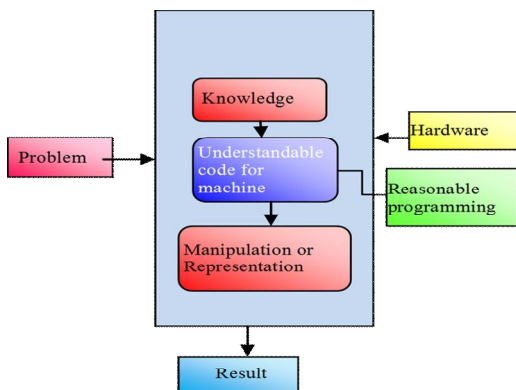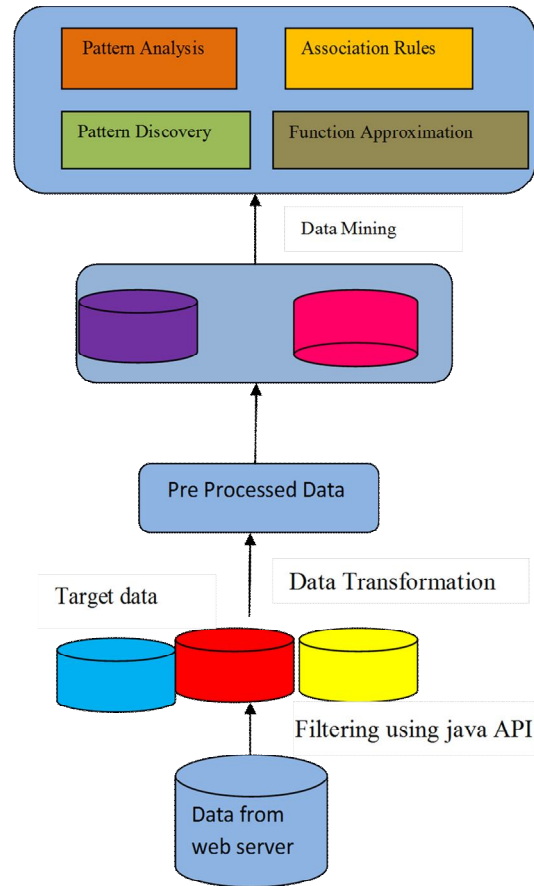


Figure 2. Knowledge Representation



Figure 3. Knowledge Discovery

## 3. SWARM INTELLIGENCE ALGORITHMS:

Several collective behaviours such as bees, ant, firefly, cuckoo, etc., inspired algorithm have been proposed. These algorithm provide problem solving ability and study on this makes human understandable in nature. The main principles of collective behaviour of swarm are homogeneity, locality, collision avoidance, velocity matching and flock clustering. Some of the popular SI algorithms are as follows:

## 3.1 Ant colony optimization

Ant colony optimization (ACO) is a behaviour of ants that finds its path between food source and colony. Ant while returning colony lay down pheromone, which directs the search of future ants on the same path. This helps to find the optimal solutions. They use the environment as a medium of communication. Some of popular variations of ant colony are edges, Max Min ant system, convergence and pheromone update. ACO has been useful to solve many combinatorial problems and optimum solutions it is also used in web usage mining to each user. ACO deals with artificial systems that inspired from the foraging behaviour of real ants, which are used to solve discrete optimization problem. The main idea is the indirect communication between the ants by means of

chemical pheromone trials, which enables them to find short paths between their nest and food. The study of ant colonies behaviour and their organizing capabilities is of interest to knowledge retrieval or management and decision support systems sciences, because it provides models of distributed adaptive organization, which are useful to solve difficult classification, clustering and distributed control problems.[1],[11],[12]

## 3.2   Bat algorithm
Bat algorithm (BA) is a Meta heuristic optimization inspired by the echolocation of micro bats. Bats are one kind of mammals that have the capacity of echolocation. These bat produce loud sound based on the echo reached it identifies the obstacles and severs even in dark. Many researchers found the behaviour of bat has the solution to many complex problems.

## 3.3   Cuckoo search
Cuckoo search (CS) is swarm intelligence inspired by the some cuckoo species they lay their eggs in the nest of another birds. The breeding behaviour is applied to many optimization problems. Each egg represents the solution. Recent studies suggest that comparison between CS searches with PSO says CS is robust results.[2]

## 3.4   Firefly algorithm-
Firefly algorithm (FA) inspired by the flashing behaviour of fireflies. Firefly algorithm consists of three rules:

1. No firefly will be attracted to another since they are unisex.

2. Attractiveness is proportional to brightness that is less brighter one is attached to brighter one.

3. If no firefly is brighter than the giver firefly then they are moved randomly.

In fact the variants of firefly algorithm are discrete firefly algorithm, multi objective FA and so on. It is applied on image processing, clustering, continuous optimization, etc.

## 3.5   Particle swarm optimization-
Particle swarm optimization (PSO) is a swarm intelligence global optimization technique. It was founded in1995 by James Kennedy and Russell Eberhart to model the convergence behaviour of a flock of birds. PSO is a population based algorithm and is initialized with a population of random solutions, called particles. Each particle is associated with a velocity. Particles fly through the search space with velocities which are dynamically adjusted according to their historical behaviours. Therefore, the particles have the tendency to fly towards the better and better search area over the course of search process. It is mainly inspired by the social behaviour of bird flock and fish school. In this if one particle identifies path for food source or protection then rest of swarm follows it automatically even if it is in opposite direction. The birds also have the capacity to smell the food so it finds the optimum solution to find the food. Using the pre-defined fitness function the performance of each particle is measured.[3], [9],[14]

## 4.   WEB USAGE MINING
Web Usage Mining aims to capture users interacting with the web. Web Usage Mining is the non-trivial process to discover valid, novel, potentially useful knowledge from web data using the data mining techniques. Data stored in usage logs can be used for solving navigational problems, improving web search, recommending queries, suggesting web-sites and enhancing performance of search engines.

Web Mining is the non-trivial process to discover valid, novel potentially useful knowledge from web data using the data mining technique. It may give information that is useful for improving the services offered by web portals. Web mining considered as the most efficient net tool in converting the meaningless information into meaningful in the internet environment , processing the data, extracting the data, and making route map b y the data acquired web 3.0.[6]

Web Usage Mining is the technique or application of data mining to the data generated by the interaction of user with the web servers, this kind of data stored in server logs, represents a valuable source of information, which can be exploited to optimize the document retrieval task or to better understand and thus satisfy user's need.

For recommendation of web page, here present a graph based approach, leverages the user browsing logs to identify early adaptors, there user discover interesting content before author and monitoring their activity we can find web pages to recommend.[10]

Usually web usage data is collected from web server which is in form of log file. The mined knowledge from log file is used by the companies to establish better customer relationship by giving them exactly what they need. The companies can find attract and retain customers which help them to improve their business performance.

## 5.   RECOMMENDER SYSTEM
An information filtering technology, company used on e-commerce website that uses a collaborative filtering to present information on items and products that are likely to be of interest to the reader. In presenting the recommendations, the recommender system will use detail of the registered user's profile and opinions and habits of their whole community of users and compare the information to reference characteristics to present the recommendations. Recommendation helps to group similar type of users together. Recommendation gives suggestions about user based on their information previously provides or currently currently visited web page.

An example of recommender System is WhatShouldireadnext.com. A site where users can enter a title of a recent book they have read and enjoyed to see recommended books that they are likely to also enjoy.

It is often necessary to make choices without sufficient personal experience of the alternatives. In everyday life, we rely on recommendations from other people either by word of mouth, recommendation letters, movies, book reviews printed in newspapers, or general surveys. In a Typical Recommender System people provide recommendations as inputs, which the system then aggregates and directs to appropriate recipients. The system's value lies in its ability to make good matches between the recommender and those seeking recommendations.

The developers of the first Recommender System, Tapestry coined the phrase "Collaborative Filtering".

The good examples of Recommender Systems are-

1. Offering news articles to on-line newspaper readers, based on a prediction of reader interests.
2. Offering customers of an on-line retailer suggestion about what they might like to buy based on their past history of purchases and for product searches.

Recommender Systems allow to learn user preferences and to make recommendations. They can be employed to recommend products (e.g. news, photos, movies, music, etc).

Recommender System can be: (i) Content based, the system recommend items similar to the ones the user preferred in past. (ii) Collaborative-Filtering based, the system recommends items that people with similar tastes liked in the past and (iii) Hybrid, the system combines content and collaborative-filtering based methods.

Target Marketing- Main goal of Target Marketing is to identify group of users or customers with similar behaviour so that one can predict the customer's interest and make proper recommendations to improve.

## 6. WEB - PAGE RECOMMENDATION USING SWARM INTELLIGENCE:

The systems for web page recommendation are based on collaborative filtering approaches. The idea is to exploit the user browsing logs (that is click information), in order to identify users with similar likes and dislikes. These recommender systems deals with very high levels of noise, since visiting a web page is not a clear indication of interest as renting a DVD. Search engine help people to search for information on the internet, but the web search is effective only when the users have a clear idea of what they want. Often, people have no specific information need. Recommender system produce suggestions and they are effective in static and relatively noise-free environment. The design of recommender system for web content poses significant challenge due the dynamic nature of webpage, and the high level of noise introduced by the analysis of the user-browsing data.[4]
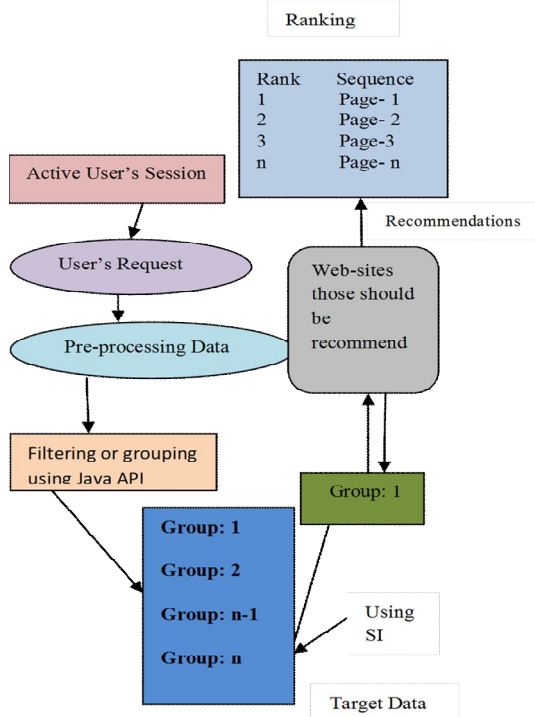


Figure 4. Block Diagram to understand web recommendations

For web page recommendation, here we will use a graph based approach. This algorithm is proposed to use this graph to identify those users who discover interesting pages before this. By tracking the browsing activity of early users we can identify new interesting pages early and recommend these pages to users who should interest with early users.

The optimal solution on diversity is found by recommendations. Based on customer's likes and dislikes they are usage profile is grouped. To generate the recommendations for active user [5], we can follow the following steps:

### 6.1 Approach
1. User Profile Management
2. Identifying the Optimal Nearest Neighbour Profile
3. Profile Matching

#### 6.1.1 User Profile Management
The Very First step is to generate or collect usage profile of each user from the given website. The profile [U, J] represents the usage profile of user U for J item. The early–user graph G is an attributed, directed, and weighted graph, where nodes correspond to user and an edge between two users, u and v, express the fact that the two different users visited the same site. We generate a graph using tuple (u, p, t, z). Where: (i) u is the identifier of the user. (ii) p is the URL of page visited by user. (iii) t is the timestamp of the visit page by u at p. (iv) z is the optional attribute, we can use this attribute to hold coordinates of topic that has been liked by user at any particular page.

Let, there be n users, $u_1,u_2,u_3......u_{n-1},u_n$ and $p_j$ is the page that has been visited by all users at different timestamp values that is $t_1,t_2,t_3......t_{n-1},t_n$. So the chronologic access will be:- $V[p_j]$: [$(u_1, t_{1j})$, $(u_2, t_{2j})$, $(u_3, t_{3j})$, ............ , $(u_{n-1}, t_{n-1j})$, $(u_n, t_{nj})$]. Where $t_{ij}$ is the timestamp value of the first visit of user i to the page j, and z denotes the co-ordinates of the topic on particular page that has been liked by users.

#### 6.1.2 Identifying the Optimal nearest neighbour Profile
The neighbourhood selection is processed to find the optimal solution the users that are most nearest user to current active user. The best nearest similarity is found nearest neighbourhood selection. The optimal nearest user profile is identified using the Swarm intelligence Optimization Techniques. By calculating relative position, time distance function, edge-weight for the given active user is selected. The Swarm intelligence can be ACO, PSO, Bat, Cuckoo Search, and Firefly.

#### 6.1.3 Profile Matching
Let R(u) be the relative position of user computed by considering the position of user in the access list. The Profile Matching is used to calculate distance between two different profiles using time distance function. We can treat the relative position and timestamp value as the coordinates of particular user like $(u_{i,tij})$, where u denotes the user and t denotes the timestamp value of user for page p that has been liked by user u. The Time distance can be defined as:

$$\text{Time Distance } [U,J] = \sqrt{[(u_{n-1}-u_n)^2 + (t_{(n-1)j}-t_{nj})^2]}$$

Using this time distance function, we can calculate similarities between user profiles and the nearest optimum is chosen. The Recommender System can easily choose the profiles whose Time Distance Function is above a certain threshold value $v_{th}$ as the neighbourhood of u.

Weight of edge from user u to v, i.e. w[u,v], represents that a page visited by u is then visited by v. The edge weight is defined by Time Distance between two different profiles.

The nearest Neighbourhood Selection process is used to find the optimal solution, the users that are most nearest to current user. Let us assume that U is the current working user, the users with similar likes are treated as the optimum solution for user U. The profiles are selected from the database. Based on threshold value profiles are grouped together. The usage profiles are best similar among them will be selected.
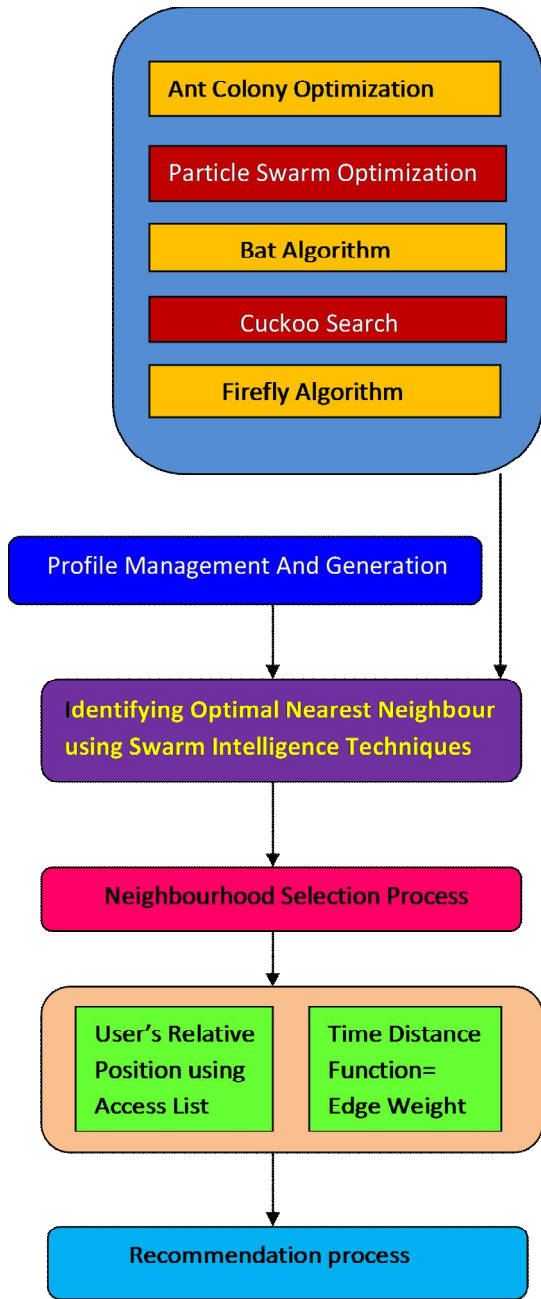


Figure 5. Web-Page recommendation Using Swarm intelligence Optimization

Web personalization implicitly or explicitly collect the data from the user. The above figure represents the framework the way recommendation is obtained. After the profile is generated obviously, it is ready for recommendation. The next step after profile generation and management is the nearest neighbourhood selection it can be processed by any one of five popular Swarm Intelligence Techniques. Next step to calculate the Relative position of user and Time Distance Function based on the threshold value. This relative position is treated as the weight between edges. The final stage is the recommendation process is generated. To improve the relevance of our recommendations, we rank recommendations by using the relative position of user u from whom the recommendations originates, as well as the edge weight w[u,v]. Additionally, we can use the page topic to improve or boost scores of page whose topics match the interests of user u. When a page is suggested by many early active users, the final recommendation from u to v is the sum of contributions of the early active users for v.

## 7. CONCLUSION

The key contribution of this paper is web usage mining using Swarm Intelligence Optimization. This work has been shown that recommendation process is carried out with Swarm Optimization Techniques. This research introduces the theoretical foundations of the biological motivation and Swarm Intelligence Optimization Techniques with a focus on Web-Page Recommendation. The interpretation of result can be used in focalized marketing strategy like direct marketing and target marketing. Concerning the early-adopter graph, we can plan to use different models to learn the edge-weights. Then we can investigate the applications of early-adopter model to other domains.

## 8. REFERENCES

[1] Abdurrahman et. al., Classification of web user in web usage mining using ant colony optimization algorithm, Doctor Dissertation, Institute of technology, 2009.

[2] Ang X. S. and Deb S. (2010a) Engineering Optimization by Cuckoo Search, Int. J. Math. Modelling &Num. Optimization, Vol. 1, 330-343.

[3] J. Kennedy, R. Eberhart, 1995. Particle Swarm Optimization, IEEE International Conference on Neural Network, Vol. 4, pp. 1942-1948.

[4] N Golovin, E. Rahm: Reinforcement learning Architecture for Web Recommendation, Proceedings on Information System.

[5] I. Mele., F. Bonchi, and A. Gionis. The early-adopter graph and its application to web-page recommendation. In CIKM, 2012.

[6] J. Srivastava , R. Cooley, M. Deshpande, and P.-N. Tan. Web Usage Mining: discovery and application of usage patterns from web data. SIGKDD Explor . Newsl. , 1(2):12-23, 2000.

[7] A. Abraham, He Guo, and Hongo Liu, Swarm Intelligence: Foundations, Perspective and Applications. Studies in Computational Intelligence (SCI), vol.26,pp.3-25.2006.

[8] A. Khosla, Shakti Kumar, K.K. Aggarwal, and Jagatpreet Singh. A Matlab Implementation of Swarm Intelligence based Methodologies for Identification of Optimized fuzzy Models, Studies in Computational Intelligence (SCI) vol.26, pp. 175-

184, 2006.(this reference is taken by me only for study point of view)

[9] J. Kennedy, R. Eberhart, 1995. Particle Swarm Optimization, IEEE International Conference on Neural Network, Vol. 4, pp. 1942-1948.

[10] Abdurrahman et al., Classification of web user in web usage mining for analyzing unique behaviour of web user, International Conference on Electrical Engineering and Informatics, 2007, pp. 356- 359.

[11] Abraham A, Ramos V (2003) Web Usage Mining Using Artificial Ant Colony Clustering and Genetic Programming, 2003 IEEE Congress on Evolutionary Computation (CEC2003), Australia, IEEE Press, ISBN 0780378040, 1384-1391.

[12] Dorigo M, Blum C (2005) Ant colony optimization theory: A survey. Theoretical Computer Science, 344(2-3), 243-278.

[13] Liu Y, Passino KM (2000) Swarm Intelligence: Literature Overview, http://www.ece.osu.edu/ passino/swarms.pdf.

[14] Pomeroy P (2003) An Introduction to Particle Swarm Optimization, http://www.adaptiveview.com/articles/ipsop1.html.