## Utilization of Structure and Concepts in Handling Network and Information Management Complexity

SEYED A. SHAHRESTANI School of Computing and Information Technology University of Western Sydney Penrith Campus, Locked Bag 1797 PENRITH SOUTH DC NSW 1797 AUSTRALIA

*Abstract*: - The need for access to information and services in modern societies has fuelled the rapid growth of networks and web technologies. One of the prerequisites for successful management of the resultant complex networks is the ability to handle large amounts of information. The information may contain incoherent, missing, or unreliable data that need to be filtered and processed. In this respect, the power of soft computing in handling uncertainties makes it an appropriate choice for taking up a significant role in information handling and management. This paper focuses on this topic, highlighting the role of fuzzy proximity networks and multiple distributed agents in identifying or improving solutions to network management problems. To demonstrate the effectiveness of this approach, some specific areas of functional importance, with emphasis on information retrieval and management are also discussed.

*Key-Words:* - Information management, Information retrieval, Network management, Soft computing, Fuzzy proximity network.

### 1 Introduction

The explosive growth of the Internet has provided modern societies with many new opportunities and challenges. The communication landscape has changed from a voice dominated one to a market where Internet and improved information retrieval and data services are the main drivers. Additionally, as a result of comprehensive monitoring abilities, modern systems result in an overwhelming amount of data. Conventional computer applications provide some degree of automation to process and filter the data to identify relevant information, but human interactions remain essential, as the data is often incomplete and conflicting. In principle, artificial intelligence (AI) techniques could limit the need for human intervention. In particular, several characteristics of soft computing make it an effective approach for use in an integrated network management environment.

From a business and corporate perspective, systems for information management and retrieval are also extremely important. The traditional approach of relying on paper filing systems for document storage and IR is giving way to digital document management systems. The latter can easily overcome storage and information retrieval (IR) problems at lower costs in a speedier manner. Most current IR system depends heavily on the indexing of digital records.

It can be noted that in general, it is trivial to combine multiple keywords within a query made by an individual as an aggregate fuzzy set using fuzzy operators [1]. In a similar fashion, one may propose that queries from several individuals can also be based on the simplistic approach of considering them as a single query with multiple keywords. Although in some cases, simple fuzzification of numerical terms can address this problem, the general solution needs to be based on processing of relations, structure and concepts [2]. To achieve this, more sophisticated and artificially intelligent systems are needed. The intelligent system must be built upon the collaborative nature of the queries by noting the implicit connection between the individuals. To achieve these, in this work, an IR system that is built on fuzzy proximity network (FPN) concepts [3] [4] and multiple distributed agents is proposed. The evolving system will be based on utilization of a combination of linked keywords, which can arguably result in improved search engines and IR systems.

The main objective of this work is to discuss the ways that soft computing and structural concepts can be used in enhancement of an integrated network management environment. This is achieved in the remainder of this paper by using the following structure. Section 2 presents the complexities in information and network management. Section 3 highlights how soft computing and FPN should be considered in handling such complexities. Sections 4 and 5 focus on specific and significant application areas and the proposed solutions through soft computing and FPN. The concluding remarks are given in Section 6.

## 2 Challenges in Information and Network Management

In typical cooperatively managed а environment the retrieval of information from knowledge-bases is typically achieved through utilization of keywords and unstructured phrases. Also, help desk systems are designed to provide customer support through a range of different technology and Information Retrieval (IR) tools play a fundamental role in this activity. Efficiency and effectiveness in data retrieval being crucial for the overall problem solution process heavily depend on the abstraction models. The abstraction associated with an object should capture all its easily peculiarities in an manageable Identification of relevant representation. features of achieving an object abstraction is a complex task and presence of uncertainties makes this task even harder [5].

Distributed applications are evolving towards compositions of modular software components with user interfaces based on web browsers. Each of these components provides well-defined services that interact with other components via network. The increase in the complexity of distribution makes it more difficult to manage the end-to-end Quality-of-Service (QoS). A management system deployed to diagnose QoS de-gradation should address two major issues. First, to measure the performance of applications, it needs a lowoverhead, scalable system for measuring software components. Second, the performance management system must monitor selected measurements, diagnose QoS degradation, adapt to the environment and integrate with network management systems [6].

# **3 Fuzzy Proximity Networks and Information Management**

At the highest layer of any hierarchical model for network and information management the problems can be associated with an overwhelming amount of data. Consequently, the challenge here is to process the data and present only the relevant information by acting as a decision-support tool. At this layer, the response time is important but not critical. This type of task is well suited for techniques that implement search techniques, e.g. genetic algorithm. Also, model-based expert systems can be used to hide the network complexity behind several abstraction levels. In this environment soft computing can be used to model/data uncertainties manage and ambiguities while interpolating between (possibly) several emerging models. The resultant aggregate model will also have some degree of confidence attached to it that will assist the operators in dealing with the presented information.

While in the next parts we take a closer look at some of the tasks in what may be referred to as the service layer, it can be noted that the above discussion holds for both management of the services and the network. For example, AI based network management systems that deal with the problems at network layer, are mostly based upon expert system techniques [7] [8]. From a broader point of view, the ability to handle massive amounts of information is a prerequisite for management of any complex system. The experience gained on a problem represents the knowledge that can be of value in the future. Information retrieval (IR) tools are the very bases for any process that deals with large databases. In a standard IR context, uncertainty pervades the behavior of both the system and the users.

Many current systems for retrieval of information from corporate knowledge bases have fundamental shortcomings. Their main inadequacies are related to how they interpret and utilize the keywords or search phrases used by their human clients. More specifically, the system limitations can usually be linked to either their misinterpretations arising from natural language ambiguities or their inability to manage imprecision and vague definitions. To undertake uncertainty and adaptivity problems simultaneously, soft computing offers excellent solutions. More specifically, in the next part a framework for design of a fuzzy search engine and cooperative IR system is proposed. The system will be based on expanding previous works on fuzzy proximity network (FPN) and multiple distributed agents. Each client is to be supported by an agent. The FPN will form the basis for the retrieval of several inter-related pieces of information. This will be based on (virtual) joint queries, formed by using the characterizing attributes established via the maximum spanning trees. In this way, they can be expected to account for the connection of related keywords that form a structured concept. The combination of linked keywords, rather than emphasizing on the keywords themselves, can result in improved search engines and IR systems.

## 4 Network Diagnostic Systems and Advanced Help Desk

In a competitive business environment, customer satisfaction is a vital objective for many companies: high-quality products and high-quality customer service are two strategic aspects. In this context, help desk systems play an important role providing customer support and functions like change, configuration and asset management. The two main components of a help desk system are the front-end and the back-end ones: the former manages the interaction with customers while the latter deals with IR issues.

As discussed in the previous section, the IR process is based on matching of the object descriptions (system knowledge) with an ideal description derived from the user query, while allowing for the expression of different views on the same component. It can be noted that various elements may affect the computation of the relevance of a keyword for the description of an object [5]. Using soft computing enforces the possibility to express such elements through a complete relevance distribution (through membership functions) that can be employed in the retrieval process.

The core functionality is the retrieval of data from a database whose abstraction matches the description of an ideal object, inferred from a query. Implementation issues are critical both for the overall performance of the system and the accuracy of the retrieved information. Customers usually provide data with different degrees of confidence depending on how that information has been collected. Current IR tools do not explicitly model the uncertainty associated with information but they mix the measure of relevance associated to information with the relative measure of confidence. An effective use of that information is the key to enable a process of system adaptation. The explicit management of relevance and confidence on information, integrated with an adaptivity process is the key factor for improving the retrieval precision of a help desk system [5]. Keywords remain the base of the abstraction model. But they will be enriched with information on relevance, relations, structure and concepts. As it will be discussed in the next section, this can be implemented through the use of FPN.

### 5 Information Management and Retrieval

In addition to applications in network management, search engines and information retrieval systems are also extremely important from a business and corporate perspective. This is mainly due to the fact that the traditional approach of relying on paper filing systems for document storage and IR is giving way to digital document management systems. The latter can easily overcome storage and IR problems at lower costs in a speedier manner.

Most current search engines or IR systems function through indexing, and utilization of keywords and unstructured phrases. Such approaches are based on identifying all documents that are either indexed by the used keywords or contain the word or phrase used as the search parameter. They may then rank the results by some degree of relevance according to the query made by an individual.query. In most of these he IR approaches, the presence or absence of the keywords in the query and the indexing terms of the documents, form the basis for evaluation of the relevance of a document to the query. However, basing search engines or IR systems on such approaches has fundamental shortcomings.

Among the basic deficiencies that need to be dealt with here, is the lack of ability to express the linguistic based queries made by humans in a formal way needed for machine interpretation and processing [9]. Another and probably more fundamental problem relates to identifying suitable ways for representation and inference of concepts and the context in which they appear. In machines, the concepts need to be precisely defined, leading to lack of generalization that in turn causes the number of cases that need to be dealt with increase rapidly.

For instance, consider a query with regard to *'heavy traffic'* on a link in a computer network. Obviously, there is no clear boundary that can be used to distinguish between 'heavy traffic' and let's say 'moderate traffic'. Here, the problem can be related to vagueness and lack of specificity. Additionally, the level of traffic that is considered as 'heavy' on a particular link is not necessarily considered as such on another link. Moreover, the term 'heavy traffic' may be used in contexts other than those related to the link traffic, e.g., 'the students are late due to heavy traffic'. In this case, the problem is related to the vagueness of the meaning and ambiguity in the language.

To address the lack of flexibility in representing documents and queries, fuzzy systems that deal with this type of problem for individual users have also been studied and developed by several researchers, for example see [10]. In such approaches, a fuzzy set will represent each keyword. The membership value of each piece of information or document indicates its degree of relevance to the fuzzy set denoted by the keyword. In this way, it is easy to use linguistic qualifiers for computing with words to enhance the information retrieval process. While this can help in indexing and the querying process, users can also employ it to feedback information. provide Such information can be used to evaluate the retrieval system and in turn for evaluation of the search engine.

To take the inexact matching of keywords into consideration, these ideas have been expanded to incorporate fuzzy thesaurus [11]. It is trivial to combine multiple keywords within the query made by an individual as an aggregate fuzzy set using fuzzy operators. In a similar fashion, one may propose that queries from several individuals can also be based on the simplistic approach of considering them as a single query with multiple keywords. Although in some simple cases simple fuzzification of numerical terms can address this problem, the general solution needs to be based on fuzzification and processing of *concepts*. Through computation with words and the use of linguistic variables, the solutions need to manage the inherent fuzziness in human queries, representation of concepts and coordination, properly and efficiently [1].

Within a cooperative and management environment, an intelligent system can be built upon the collaborative nature of the queries by noting the implicit connection between the individuals. The system should be capable of taking into account the individuality of different users; e.g. different people will generally use different keywords for the same query. Therefore, the emphasis should be on concepts, structure and connection of keywords as well the behavior and knowledge levels of the individuals using those keywords. One of the main advantages of such a system is therefore, related to its ability to elevate the awareness levels of the individuals by pointing to them the data set items that they might have been missing otherwise.

A scheme that is based on FPN [3] [4] can be utilized to build the required intelligent system. FPN is capable of performing the needed aggregation. The network also achieves the representation of the fuzzy engine for the implementation of the multi-agent framework. The network is capable of providing coordination services for cooperating managed agents. Conceptually, a separate network or a particular part of the network is to take on a 'coordinator' role. The coordinatorrole and its agent need to evaluate and aggregate the queries from individual agents to help the cooperating agents in achieving their common goal. One important aspect of such coordination relates to connecting the cooperating agents by pointing to them information and documents relevant to their tasks and searches, even when one agent has not explicitly asked for them. To achieve this, the system needs to be able to process queries from different cooperating users as collaborative queries[2]. In this case, each node *i* of the fuzzy proximity network represents a keyword. The weight w(i, j) represents the fuzzy relevance of the two keywords at nodes *i* and *j*. Such a scheme does emphasize the keyword structures and connections, rather than focusing on the keywords themselves. The initial relevance between the keywords is based on the co-occurrence of a keyword or the socalled Miyamoto's measure [12] [4]. Stated simply, this measure implies that the more often two keywords occur simultaneously, the



Fig.1: Formation of Fuzzy Proximity Network (FPN)

higher is their relevance to one another. In any case of practical importance, typically the pieces of information are in several documents, including a document d denoted by D(d). Let the  $k^{th}$  keyword in d be represented by K(d, k), as in the FPN depicted in its essential form in Fig.1. As a starting point, the keywords within any given document can be considered to be related to each other. For instance, keywords K $(1, 1), K(1, 2), \dots K(1, m)$  are considered to be related, as they appear within the same document, D (1). The fuzzy relevance of keywords is represented by the weight w between their respective nodes. For example, here the fuzzy relevance between the two keywords K(1, 1) and K(1, 2) is represented by the weight w (K (1, 1), K (1, 2)). In accordance with the co-occurrence concepts, if document D (1) refers to another piece of information in D(2) or is referred to by the information content of D (3), then the keywords  $K(2, 1), K(2, 2), \dots K(2, n)$  as well as the keywords K(3, 1), K(3, 2), ..., K(3, p)are also considered to be related to each other, although in a weaker sense. This type of information will establish the initial setting of weights in the network model. Obviously, after this initial stage, the weights can be updated through adaptive mechanisms and supervised learning.

For each document, its characterizing attributes are calculated based on a maximum spanning tree [3]. Here, as in several other applications, a spanning tree is the tree that covers a given set of nodes, i.e. keywords. The weight of the tree W(.), is the sum of the weights of the branches in that tree. A maximum spanning tree is established as the tree with the maximum weight for a particular set of nodes. Given a query Q(q), its maximum spanning tree weight W(Q), is used as the characterizing measure of the query. The weight of the maximum spanning tree for the keywords common between O(q) and a document D(d) divided by W(Q) is used to represent the characterizing attribute measure R(.), of document D(d) with regard to Q(q). These characterizing attributes calculated for all of the documents, are then used for ranking the documents with regard to their relevance to the query Q(q).

In summary, we have proposed a cooperative IR design framework based on multiple agents, where each human role being supported by an agent. The FPN forms the basis for the retrieval of several inter-related

pieces of information. Note that the (virtual) joint query is not formed through a union of the keywords used in the queries of the individuals. It is rather based on consideration of structure and combination of the keywords. This is achieved by using the characterizing attributes based on the maximum spanning trees. They account for the connection of keywords that are linked together to form a structured concept. Clearly, the combination of the keywords, rather than emphasizing on the keywords themselves, is highly advantageous.

#### 6 Concluding Remarks

The management system for modern complex networks must be capable of dealing with an overwhelming amount of data that may be incoherent and inconsistent or unreliable. Compared to more conventional techniques, AI approaches are more suitable for this type of task. In particular, the capabilities of soft computing in handling vague concepts or systems with uncertainties are of prime significance. This paper described several ways that soft computing and specifically fuzzy proximity networks can be used in improving the solutions to problems encountered in network and information management environments. One of the main advantages of such solutions is that they are based on processing of structural concepts. In particular, for information retrieval, the ensuing systems are based on utilization of a combination of linked keywords, which result in improved functionality.

References:

- [1] S. Ghosh, Q. Razouqi, H. Schumacher, and A. Celmins, "A survey of recent advances in fuzzy logic in telecommunications networks and new challenges," *IEEE Trans. Fuzzy Systems*, vol. 6, no. 3, pp. 443-447, 1998.
- [2] P. Ray and S. A. Shahrestani, "Fuzzy modeling of cooperative service management," *Lecture Notes in Computer Science; Trends in Distributed Systems: Towards a Universal Service Market*, vol. 1890, pp. 232-243, 2000.
- [3] S. Shenoi and A. Melton, "Proximity relations in the fuzzy relational database

model," Fuzzy Sets and Systems, vol. 31, pp. 285-296, 1989.

- [4] C. Sun, "An information retrieval model for coordination systems based on fuzzy proximity networks" *Proc. IEEE International Conference on Systems, Man and Cybernetics*, pp. 418-423, 1990.
- [5] G. Piccinelli and M. C. Mont, "Fuzzy-set based information retrieval for advanced help desk," Technical Report HPL-98-65, HP Laboratories, Bristol, 1998.
- [6] J. Martinka, J. Pruyne and M. Jain "Quality-ofservice measurements with model-based management for networked applications," Technical Report HPL-97-167, HP Laboratories, Palo Alto, 1998.
- [7] N. Nuansri, T. S. Dillon and S. Singh, "An application of neural network and rulebased system for network management," in *Proc.* 30<sup>th</sup> Hawaii International Conference on System Sciences, 1997, pp. 474-483.
- [8] S. A. Shahrestani, H. Yee, and J. Ypsilantis, "Adaptive recognition by specialized grouping of classes," in *Proc.* 4<sup>th</sup> IEEE Conference on Control Applications, Albany, New York, 1995, pp. 637-642.
- [9] L. A. Zadeh, "The role of fuzzy logic in the management of uncertainty in expert systems," *Fuzzy Sets and Systems*, vol. 11, pp. 199-228, 1983.
- [10] V. Tahani, "A conceptual framework for fuzzy query processing," *Information processing and management*, vol. 13, pp. 289-304, 1977.
- [11] Bezdek, G. Biswas, and L. Huang, "Transitive closures of fuzzy thesauri for information retrieval systems," *Int. J. of Man Machine Studies*, vol. 25, pp. 343-455, 1986.
- [12] S. Miyamoto, T. Miyake, and K. Nakayama, "Generation of a pseudothesaurus for information retrieval based on co-occurrences and fuzzy set operation," *IEEE Trans. On Systems, Man, and Cybernetics,* vol. 13, no 1, pp. 62-70, 1983.