*https://dx.doi.org/10.22161/ijaems.4.5.8* *ISSN: 2454-1311*

# Applying Soft Computing Techniques in Information Retrieval

Namrata Nagpal

Amity Institute of Information technology, Amity University, India
Email: nnagpal@lko.amity.edu

***Abstract—*** *There is plethora of information available over the internet on daily basis and to retrieve meaningful effective information using usual IR methods is becoming a cumbersome task. Hence this paper summarizes the different soft computing techniques available that can be applied to information retrieval systems to improve its efficiency in acquiring knowledge related to a user's query.*

***Keywords— fuzzy logic, information retrieval, IR models, neural networks, soft computing.***

## I. INTRODUCTION

There is an urgent requirement of effective Information Retrieval Systems(IRS) has aroused with the quick growing development of the Internet, and plethora of availability of online text based information. The goal of an IR System is to retrieve relevant information regarding a user's query.

The efficiency of an Information Retrieval System can be calculated using parameters which are prime in order to meet the requirement of the system and accomplish its goal instantly. The keywords put by the user while forming a query as per his requirement is mostly vague and uncertain. Also the document representation as well as the process by which the query is matched to the document is also uncertain. Hence the effectiveness of an Information Retrieval System crucially depends upon the system's capability to deal with the vague and uncertain retrieval process.

There are various techniques utilized for above efficiency; we majorly focus on the significance of soft computing related to artificial neural networks, fuzzy logic, genetic algorithms, or rough sets, to name a few.

The paper states about the significance of information retrieval and various IR models in section 2. Section 3 deals with soft computing techniques available whereas Section 4 specifies the application of soft techniques to improve the information retrieval process.

## II. INFORMATION RETRIEVAL

Information retrieval refers to a task of finding the relevant documents from given a set of documents and applying a user query. Information retrieval applications also require characteristics like accuracy, speed, consistency and ease of use in recovering relevant documents that satisfy user queries. Some required characteristics for efficient information retrieval are as under [2]:

- *Accuracy:* It refers to the relevant correct response of information retrieved in recall i.e., the percentage of relevant information retrieved with high precision.
- *Consistency:* Consistency of data retrieval should be maintained via indexing of the text by the groups of indexers or by the authors.
- *Ease of use:* Information retrieval has become the blood in the vein of the users in current scenario using Internet on daily basis. Hence, it is more of a responsibility now of information retrieval systems to cater user requests with utmost efficiency and effectiveness.
- *Speed:* It refers to the instant time taken in return to searching the given information using fast search techniques like NLP or run as "batch" indexing.

### 2.1 IR MODELS

Information Retrieval (IR) is by far the most sought after process used nowadays with the netizens. It is the process which shows that a data collection is represented, stored, then searched for the purpose of discovering knowledge in response to user's query. Information Retrieval (IR) came into existence in 1950s out of sheer necessity of acquiring information from users' perspective. Since then, there has been a constant development in this field in search of the betterment of IR systems. Several IR systems like google.com have become an everyday commodity and are used by wide variety of users. Thus, Information retrieval has become an integral part of our lives and an ever growing research area too. The quality of information retrieval can be measured using two main features:

- *Precision:* This is the percentage of documents retrieved related to the user's query.

*https://dx.doi.org/10.22161/ijaems.4.5.8* *ISSN: 2454-1311*

- *Recall:* This is the percentage of documents that are query related and are actually retrieved.

The fundamental IR models can be classified into Boolean, vector & probabilistic models. Each model has been introduced in its basic format as under:

1. *The Boolean model:* The Boolean model is the most primitive information retrieval method and one of the most criticized one too. This model takes a query term as an unambiguous definition of a set of documents. The Boolean model makes use of Boolean algebra operators like AND, OR, NOT in formulating a query. The major disadvantage found in this Boolean system is that the model is not able to give ranking to the returned list of documents [4]. This model associates a document and its set of keywords as a part of query separated by AND, OR, NOT. The retrieval function in the end determines if the document is relevant or not.

2. *Vector space model:* In the Vector Space Model, documents and query are represented as a Vector and the angle between the two vectors are computed using the similarity cosine function as:

   Sim(dj,q) = dj.q / | dj|. |q|

   The Vector Space Model has introduced a weight scheme called *if-idf weighting*. The new scheme has weights defined as:
   - Term frequency (tf): This factor determines the the number of times a term has occurred in the document/ query.
   - Inverse document frequency (idf): This factor measures the inverse of the number of documents that holds a given query / document term [4].

3. *Probabilistic Model:* The probabilistic model attempts to rank the documents by their probability of relevance of a given query. Binary vectors ~d and ~q are used to denote the document queries. Each binary vector indicates if a given term or document attribute occurs in the document/ query.

## III. SOFT COMPUTING TECHNIQUES

Soft Computing is a field in CS/IT dealing with the fusion of methodologies that are designed to model and provide solutions to real world problems mathematically. It aims to exploit the uncertainty, imprecision and approximate reasoning so as to achieve low-cost solutions that are robust and tractable [1]. Its solutions are majorly unpredictable, uncertain and between 0 and 1.

Soft computing is but different from hard computing; it can tolerate imprecision, approximation uncertainty, partial truth. The human mind holds the pivotal position in soft computing. In fact soft computing deals in implementation of optimization techniques to find probable solutions of hard core problems.

Various techniques used under soft computing are genetic algorithms (GA), fuzzy logic (FL), neural networks (ANN), Bayesian networks, machine learning, etc. some of which are described as under:

- *Neural networks:* Artificial neural network (ANN) is a very demanding field of computer science research in today's era. ANN is an information system whose working and thinking is based on the working of the brain. It is similar to brain in two possible ways:
  a) ANN acquires knowledge via a set learning process.
  b) Knowledge is stored in interconnected links called synaptic weights.

  ANN is mostly used in areas related to weather prediction, pattern recognition, data recognition, stock market prediction ,image processing, image compression, to name a few. ANN works best in areas which do not have set algorithms. ANN architecture comprises of three layers: input, hidden and output layer; each layer having many nodes. Back propagation algorithm is the most common method used in ANN networks [5].
- *Fuzzy Logic:* Fuzzy logic is a rule-based system that relies on the practical experience of an operator. Fuzzy logic was first devised by Dr. Zadeh of University of California at Berkeley during 1960s. Fuzzy logic deals with the concepts which can't be exactly explained in terms of binary 0 or 1. It defines the way of reasoning the Boolean values by giving results between 0 and 1. It is quite similar to the way our brain functions. Fuzzy systems are a part of developing human like capabilities. It reflects to those areas where our solution is not absolute but something fuzzy. Fuzzy logic is a form of artificial intelligence software; therefore, it would be considered to be a subset of A.
- *Genetic algorithms(GA) :* Genetic algorithm is a subset of AI and fuzzy logic. It is majorly used to deduce different optimization problems related to real-life applications. The basic idea of any GA is to copy the normal selection of any user the way he would to find a suitable solution for given application. Genetic algorithm is basically machine learning model motivated by the biological evolution model. It solves both the types of problems –

*https://dx.doi.org/10.22161/ijaems.4.5.8* *ISSN: 2454-1311*

constrained and unconstrained and repeatedly modifies series of individual solutions [6].

The utility of genetic algorithms can be seen in varied fields like climatology, control engineering, automated manufacturing & design, biomedical engineering, games theory, code-breaking, and electronic design.

## IV. APPLYING SOFT COMPUTING TO INFORMATION RETRIEVAL

A lot of effort has been made to improvise the performance of Information retrieval systems in recent years. Till now the researchers are trying hard to explore the information retrieval systems further by applying new methodologies. Applying soft computing techniques is giving positive results in increasing the efficiency of IR systems. Using the artificial neural networks and fuzzy logic concepts with information retrieval generate altogether a new field called soft information retrieval [7].

Fuzzy set theory is hence applied to increase the flexibility if IR systems. The main vagueness of IR system can be tapped well with the use of fuzzy logic. The main concern of applying fuzzy set theory to IR is:

- How to define the Boolean model representing the documents and the query language.
- How to define the associative mechanism like fuzzy clustering or fuzzy thesaurus.

To solve above mentioned problems, Boolean models have been extended to represent a document as a fuzzy set of terms. Each term has a numeric weight specified which describes the association of keywords to the document‘s content.

Various other fuzzy methods devised to improve the IR systems can be applying numeric query weights, linguistic query weights, and aggregation operators, fuzzy thesauri of terms or fuzzy clustering of documents.

Another successful method of implementing soft information retrieval is to apply ANN to the IR systems. The efficiency of information retrieval can be improved by applying the Supervised and Unsupervised learning method of neural networks.

Supervised learning method of ANN incorporates an “external teacher" [7]. This teacher then specifies the required output of the Neural Networks. During the initial learning phase, the ANN takes the values of the weights to obtain the required output.

An unsupervised learning procedure is not bothered about any learning or teaching feedback. It aims at applying self-learning. This method is also called “self-organization" because the system relies only upon local information and internal control by using the input patterns. Hence unsupervised learning has become more popular in IR specifically for documents or terms classification & clustering.

## V. CONCLUSION

Soft computing techniques have become the talk of the town in recent years and a promising area for the researchers to explore. The implementation of soft computing on IR systems to improvise its efficiency has become more important as the power of computer processing devices has increased while the cost has reduced.

The soft computing techniques have become an integral part of our daily life in the form of applying fuzzy logic, expert systems and artificial neural networks to the appliances like cookers, washing machines and refrigerators. Many commercial and industrial applications of soft computing are widely in use and is expected to grow exponentially over the years to come..

The application of soft computing techniques on information retrieval should increase drastically when combined with IoT devices in years to come.

## REFERENCES

[1] http://shodhganga.inflibnet.ac.in/bitstream/10603/10161/11/11_chapter%203.pdf
[2] http://www.doc.ic.ac.uk/~nd/surprise_97/journal/vol2/hks/infor_ret.html
[3] A. Roshdi, A. Roohparvar, “Review: Information Retrieval Techniques and Applications”, International Journal of Computer Networks and Communications Security, VOL. 3, NO. 9, SEPTEMBER 2015
[4] D.Hiemstra,P. de Vries, “Relating the new language models of information retrieval to the traditional retrieval models”, published as CTIT technical report TR-CTIT-00-09, May 2000.
[5] https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html#Introduction%20to%20neural%20networks
[6] Dogan Ibrahim, “An overview of soft computing”, 12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS 2016, 29-30 August 2016, Vienna, Austria.
[7] Fabio Crestani, Gabriella Pasi, “Soft Information Retrieval: Applications of Fuzzy Set Theory and Neural Networks”.