Winter 12-5-2004

Study of Information Supply Chain and Artificial Neural Network's Related Application

Qian Li

Yongxian Wang

Youqin Zhu

Study of Information Supply Chain and Artificial Neural Network's Related Application

Qian Li, Yongxian Wang, Youqin Zhu

School of Economics and Management, Tsinghua University, Beijing 100084, China
{Liq1, Wangyx}@em.tsinghua.edu.cn, Zyq@mail.dzdaily.com.cn

ABSTRACT

Users become less and less patient with huge useless data today. One of the great challenges now most net searching engines meet is how to get valuable information from lots of data sets. Aiming to satisfy every user's special demand, we need to integrate and optimize the whole course of data searching, including adjusting the users' input keywords, searching original results from network, and further processing of these results. Learning from the idea of Supply Chain Management, we put forward the concept of Information Supply Chain (ISC) in this paper to generalize the course above .For ISC's optimization, artificial neural network is chosen as a tool to find out the relationships between different keywords and paper categories, which are summarized and stored in knowledge base. Based on it, the process of selecting proper keywords and searching news information could be more efficient. A pruning method named MW-OBS is illustrated to train ANN as well. Some details about the framework and components are also mentioned, especially on how an individual step in ISC works, what's the relationship between them, and how they coordinate to meet every personal demand. ISC, an integrated information processing in the interests of users' individual need, has great advantages over simple searching from network with original keywords.

Keywords: Information Supply Chain (ISC); Supply Chain Management (SCM); Artificial Neural Network (ANN)

1. INTRODUCTION

Now days, it's difficult not to get news any more, but to find out the really needed information from so many data resources. People use searching engines to help themselves, however lots of time is wasted in distinguishing which piece is useful or not. The purpose of our study is to improve the information searching efficiency by some ways of optimization.

Some ideas about Supply Chain Management (SCM) inspire us. In productive industry, marketing rather than producing, is very important to earn profits for factories. The products who could meet the consumers' needs well will occupy an important position in related market. More and more factories put their focus on consumers' satisfaction, according to which, all business segments are recombined as well, including purchasing, producing, inventory, and selling. From 'push' to 'pull', factories assembled their working process through SCM, and this kind of optimization bring them great profits in the end.

We are wondering why not optimize the information searching process on the similar idea of SCM? Accordingly the conception of Information Supply Chain (ISC) is put forward, and more details will be discussed in the next parts.

2. FRAMEWORK OF INFORMATION SUPPLY CHAIN

2.1 How It Works Now?

When use some kind of internet searching engine, as we always do, select some keywords casually, run the engine, then get hundreds even thousands of related results. This process could be summarized in the following chart.

About this chart, some points should be mentioned. The quality of information searching is decided by keywords, which however are chosen by users themselves without any rules. In other words, these keywords usually couldn't represent crucial characteristic attribute of the information needed. The direct impact of the results is that more useless data is got meanwhile it's even harder to find out the real needed ones.

2. 2 "Searching Efficiency"

In order to give a pacific description of "searching efficiency", some indexes should be defined, such as 'precision' and 'repetition'

$$precision = \frac{\left|\{\mathrm{Re}levant\} \cap \{\mathrm{Re}trieved\}\right|}{\left|\{\mathrm{Re}tieved\}\right|} \quad (1)$$

$$repetition = \frac{\left|\{\mathrm{Re}peated\}\right|}{\left|\{\mathrm{Re}tieved\}\right|} \quad (2)$$

Where "Relevant" means the quantity of the information that meets the user's demand; "Retrieved" denotes the amount of actually searching results.
The former index illustrated how much partition is useful

This work was supported by Natural Science Foundation of Shandong Province of China under Grant No.03BS002.

in all results. The later is put forward by us as a complement. It's meaningless for users to get the duplicated information, and more time is also wasted. "Repetition" is set to explain how brief and concise of the searching results.

2.3 The Framework of ISC

Putting more attentions to users' information demand, and assembling the related segments in processing, ISC comes into being. The segments and their interrelationships are depicted through the below chart.

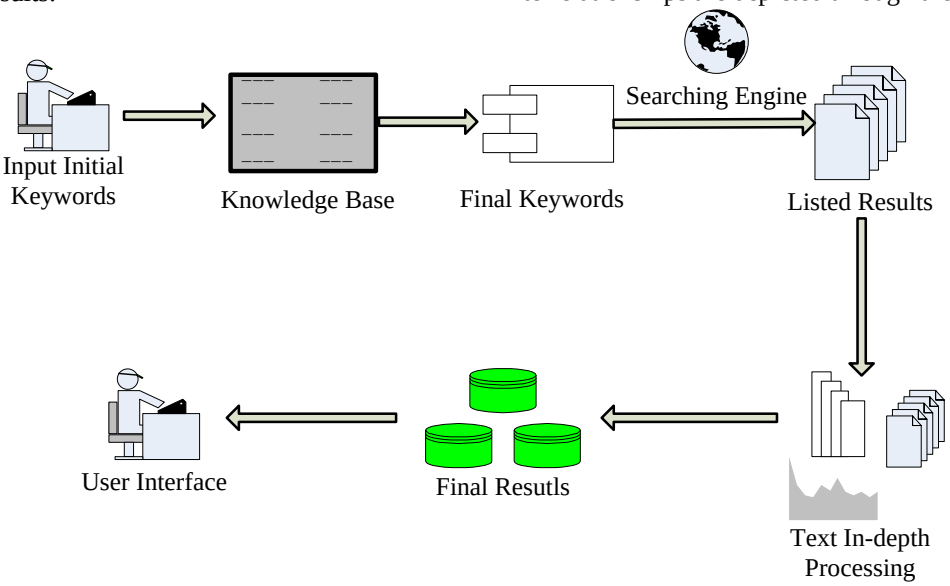

FIGURE 1 Framework of ISC

The transformation of keywords is completed through knowledgebase, in which keywords related to papers of similar categories are stored with table. A fragment is given as follows:

TABLE 1 A Fragment of Knowledgebase

Keywords	Category	Related Ones
term_1	C_1	term_1, ……, term_i
term_2	C_2	term_1, ……, term_j
……	……	……
term_p	C_q	term_1, ……, term_m

Through this kind of transformation, users could expand or delete the keywords, and then strengthen the description about what they really want. For example, If we input "Stock" in any searching engine, thousands of results will come to us. To find out which one is more useful, we have to spend more time and energy. However, in knowledge base, "Stock" is related to "Corporation、Corporation Governance" in the field of "Research"; meanwhile, it is packed with "Trade、Commission、Bonus" together in package of "Finance". Users could fix on the category they are interested in, and with the help of knowledge base, "Stock" and other relative final keywords are input to searching engine, which will bring more specific results comparing to ones above. Some further text processing will be done to the listed results, including text classification, result format standardization, and so on. These are little associated to the technique we discuss in this paper, so no more description about them will be involved here. After the processing of all steps in ISC, final results that are more concise and precise will be offered to users in the end.

3. ARTIFICIAL NEURL NETWORKS (ANN) IN ISC

3.1 About ANN

Artificial neural network is composed of a number of units, connected by weights, and it is well known about its performance on application, such as prediction, pattern recognition, machine learning, and classification. To build a neural network to perform a task, one must first decide how many units are included, how to initialize the weights of network, and which learning algorithm to be applied to a set of training and testing examples. The use of examples also implies that one must decide how to encode them in terms of inputs and outputs of the network.

3.2 Structure of ANN in ISC

As we have mentioned before, one of the crucial parts in ISC is how to build up a proper relationship of keywords in knowledge base. ANN is chosen as the optimization tool, and our purpose is to classify keywords according to category. Naturally, the input/output mapping in ANN is designed as keywords/categories, which determines the number of nodes in input and output layers with the layers being fully connected.

It's noticed that papers belonging to same category usually have different keywords to describe their attributes, which could be illustrated by Figure 2. Where

sample paper A has its unique keywords "U1…Um", and together with sample B, they both have keywords series "S1…Sp", meanwhile all the samples share the conjunct parts "K1…Kn".

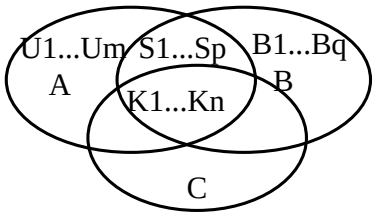

FIGURE 2 Keywords Vs. Papers

Apparently , these keywords have incremental importance in classification. Correspondingly, when a given category is set as the output, these weights connected to keyword inputs should have different value, that is the main purpose of network training.

3.3 Representation

The output data represents the class to which each paper belongs, that is expressed as 1 or 0 indicating the class to which a paper belongs or not. In view of input data format, it is a multi volume vector where 1 or 0 means certain training sample paper has or has not the correlative feature attribute. Take example we explained in the above section, suppose three papers belonging to same category have 15 keywords in sum, the input accordingly is a 15 dimension vector. For sample paper A, on the input volumes of keywords "U1…Um", the value is set as 1, however about "B1…Bq", the value is 0.

3.4 Learning Algorithm

According to the characteristics of this classification task, we choose MW-OBS, a pruning method, as optimization tool, which was put forward by us in a former paper. The basic idea of this method is that it could delete weights from a trained network, whose elimination induce so little increase in error that could be ignored. In this way, it could prune those unnecessary weights even nodes, if all weights connected to which are totally deleted. Reviewing the classification problem we described above, all keywords are set as the attribute input nodes. Through this kind of pruning, the keywords with dissimilar importance to some certain category would have different, even none connected weighs to nodes of next layer, and that is why MW-OBS is chosen as learning algorithm.

3.5 Flowing Chart of MW-OBS in ISC Optimization

In order to illustrate how ANN is trained to satisfy the need of classification, a brief flowing chart is brought forth in next page (Figure 3) considering about the main purpose of optimization in ISC—to find out which keywords are crucial to the category. Based on such information, a knowledge base about keywords could be obtained.

Where E1 is a boundary value, preventing any elimination from inducing great increase in error that can't be ignored, in other words, no more weights could be deleted. Meanwhile M denotes the number of input nodes, and I as well as J is index variable. N(I) is a parameter that represents the number of weights connecting between Ith input node and nodes of next layer. E2 is another boundary value for distinguishing crucial ones from all keywords. When error increase L(IJ) is larger than E2, the Ith input is linked by important weights, whose elimination effects network error too much, and the Ith input is regarded as a pivotal attribute to output category.

4. AN ILLUSTRATIVE INSTANCE

Hundreds even thousands of materials would be included as training samples in order to build up a practical and comprehensive keywords knowledge database in extent or depth, which is a huge and systemic task, and is not to be described in details in this paper. Being part of our work until now, we introduce an illustrative instance listed in table 2 considering its use in ISC.

TABLE 2 An Illustrative Instance

Keywords	Category	Related ones
Basketball	Shopping	Basketball shoes
	Sports	Basketball match, Street, NBA
	Education	Basketball English, history
	Entertainment	Basketball games

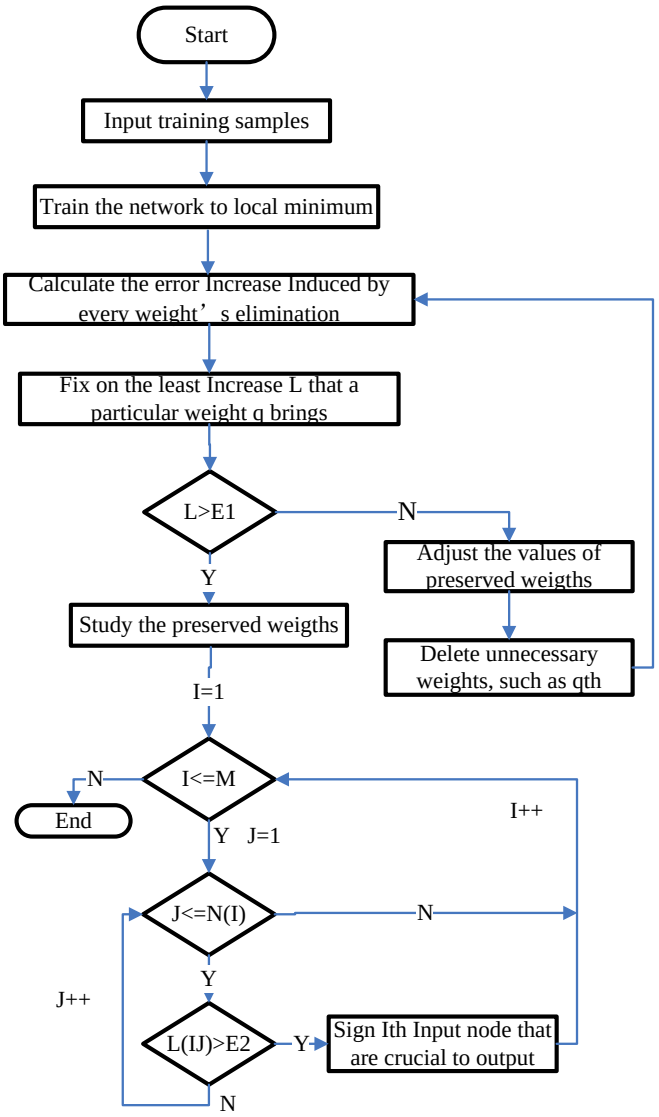

FIGURE 3 Flowing Chart of MW-OBS in ISC

About 200 paper samples are trained before arriving at these results. Reviewing the definition of "precision" offered in the previous section, several experimenters are invited to use this fragment of database to assistant their searching process and give final judgments on results. Some statistical work shows that the average precision increases near 20% through the same searching engine since more accurate keywords are chosen for searching input, and most users are satisfied with the output of ISC.

5. CONCLUSION

The conception of ISC is put forward not only for the achievements in technical innovation, but also to emphasis the importance of coordination in information processing. Through integrating the separated parts into Information Supply Chain, higher efficiency and precision are obtained for users' search.

With the application of knowledge base, users get final keywords, more specific and accurate descriptions than original ones, which increase the searching precision, also help users to extend their comprehension about the related fields. Meanwhile, ISC, as a system, provides other utilities to users, such as defining logical relationship between keywords, text similarity calculation and classification. All these factors are organized and optimized in the purpose of satisfying the needs of users.

MW-OBS, a network pruning algorithm, unlike other optimization methods, deletes or adjusts the network's weight according to their contributions to net error. In the topology of ANN after training, whether the weights connected to a certain node exist or not, as well as whether the values of them are large or little, directly decides this input node is the crucial keyword or not considering the classification task.

ISC, as a new information processing solution, has shown its advantages in primary experiment, and more related research and development of it will be carried out

in the future.

6. REFERENCE

[1] LI Q, WANG Y. "An In-depth Study and Application Analysis of a Fast Network Pruning Approach: MW-OBS", *ICSSSM04*, Vol.2, pp. 620-623, 2004

[2] Thomas J. Westbrook, J.D. "Integrating the Supply Chain", *www.wwwood.net* , 2002

[3] Tyndall, G. R. "The Global Supply Chain Challenge". *Supply Chain Management Review*, Vol. 3, No. 4, pp.13-15.2000

[4]Rich N., "Supply Chain Management and Time-based Competition: The Role of the Supplier Association", *International Journal of Physical Distribution and Logistics Management*, Vol. 27, No. 3, pp.210-225 1997

[5] Wan W, Hirasawa K, Hu J, et al. "A new method to prune the neural network",
IEEE-INNS-ENNS Inter Joint Confer on Neural Networks. Como, ITALY, 2000.

[6]Boozarjomehry R.B.Svrcek W.Y.," Automatic design of neural network structures", *Computers and Chemical Engineering 25* , pp.1075–1088, 2001.