

# **A New Marketing Channel Management Strategy Based on Frequent Subtree Mining**

**Daoping Wang  
Peng Gao**

School of Economics and Management  
University of Science and Technology Beijing

## **ABSTRACT**

For most manufacturers, success or failure is determined by how effectively and efficiently their products are sold through their marketing channel members, so the management of marketing channels plays an important role in market competition. Most existing work studies the problem of marketing channel management in a qualitative way. Recently, with the increase of amount of sales data, how to enhance the marketing channel quantitatively is significant. As the marketing channel can be viewed as a tree, in this paper, a new marketing channel management strategy based on frequent subtree mining is proposed. The proposed method is illustrated under the real-world sales data in ERDOS group. Firstly, the tree transaction is formed monthly. For each monthly transaction, only those channel members that pass the basic sales plan will be included. Secondly, we use the TreeMiner algorithm to discover embedded frequent subtrees. Finally, different management strategies are used for different kinds of discovered patterns. We show that our method can correspond to the seven decision areas in traditional marketing channel management.

## **INTRODUCTION**

For most manufacturers, success or failure is determined by how effectively and efficiently their products are sold through their marketing channel members (e.g., agents, wholesalers, distributors, and retailers). Given this situation, considerable marketing channel research has focused on organizational responsibility for managing channel how interrelationships among a firm and its channel members can be managed better (Achrol and Stern 1988; Anderson et al 1997).

Recently, our capabilities of both generating and collecting data have been increasing rapidly. The widespread use of bar codes for most commercial products, and the advances in data collection tools have provided us with huge amounts of data. This explosive growth in data and databases has generated an urgent need for new techniques and tools that can intelligently and automatically transform the processed data into useful information and knowledge. In order to relieve such a *data rich but information poor* plight, during the late 1980s, a new discipline named data mining emerged (Han and Kamber 2006), which devotes itself to extracting knowledge from huge volumes of data, with the help of the ubiquitous modern computing device, i.e., the computer.

Most existing work studies the problem of marketing channel management in a qualitative way (Coughlan et al 2005; Pelton et al 2001). Recently, with the increase of amount of sales data, how to enhance the marketing channel quantitatively is significant.

As the marketing channel can be viewed as a tree, in this paper, a new marketing channel management strategy based on frequent subtree mining is proposed. The proposed method is illustrated under the real-world sales data in ERDOS group. Firstly, the tree transaction is formed monthly. For each monthly transaction, only those channel members that pass the basic sales plan will be included. Secondly, we use the TreeMiner algorithm to discover embedded frequent subtrees. Finally, different management strategies are used for different kinds of discovered patterns. To the best of our knowledge, it is the first time to exploit the problem of marketing channel management by using frequent subtree mining. So our work is explorative, many details should be enhanced step by step in practice.

The remaining of the paper is organized as follows. In Section 2, we briefly revisit the problem of marketing channel management. In Section 3, we discuss some basic concepts of frequent subtree mining. A case study of using the frequent subtree mining in Erdos cashmere group, including representation of database, mining method and the usage of discovered patterns, is reported in Section 4. We conclude this study in Section 5.

## MARKETING CHANNEL MANAGEMENT

Marketing channels can be defined as the set of external organizations that a firm uses to achieve its distribution objectives. Essentially, a channel is the route, path, or conduit through which products or things of value flow, as they move from the manufacturer to the ultimate user of the product (Stern et al 1996). The marketing channel (interorganizational network of institutions comprised of agents, wholesalers, and retailers), by performing a variety of distribution tasks, plays a significant role in the flow of products from producers to consumers and on company profitability. Thus, manufacturers are increasingly concerned about the level of performance their channel institutions provide (Rosenbloom 1987).

Like other areas of business, marketing channels require careful administration, as superior channel management policies and strategies help a firm attain a differential advantage but concomitantly are difficult to duplicate. Marketing channel management refers to the process of analyzing, planning, organizing, and controlling a firm's marketing channels (Stern et al 1996). As discussed in numerous articles and textbooks, it comprises seven decision areas: (1) formulating channel strategy, (2) designing marketing channels, (3) selecting channel members, (4) motivating channel members, (5) coordinating channel strategy with channel members, (6) assessing channel member performance, and (7) managing channel conflict (Rosenbloom 1987; Rosenbloom 1999). All seven areas are critical to superior market performance and long-term customer loyalty (Mehta et al 2000). Most existing work studies the problem of marketing channel management in a qualitative way (Coughlan et al 2005; Pelton et al 2001). Study the marketing channel management via a quantitative method is significant.

### Frequent Subtree Mining

Data mining, whose goal is to discover useful, previously unknown knowledge from massive data, is expanding rapidly both in theory and in applications. A recent trend in data mining research is to consider more complex cases than are representable by (normalized) single-table relational databases such as XML databases, multi-table relational databases, molecular databases, graph databases, and so on (Washio et al 2005).

One of the most general formalisms for modeling complex, structured data is that of the graph. However, graphs in general have undesirable theoretical properties with regard to algorithmic complexity. In terms of complexity theory, currently no efficient algorithms are known to determine if one graph is isomorphic to a subgraph of another.

Fortunately, many practical databases do not consist of graphs that require exponential computations. The root of the complexity of graph algorithms is often the existence of cycles in the graph. In many cases, the number of cycles in graph instances in a database is limited, or the graphs may even be acyclic. The latter case is especially interesting, e.g., when the graphs are trees, because many very efficient algorithms are known for this class of graphs. A study of tree mining algorithms may also reveal insights into approaches that can be taken to deal with databases containing graphs instances with few cycles, not only from a practical point of view, but also yielding formal complexity bounds.

### Tree Concepts

A rooted labeled tree,  $T=(V, E)$  is a directed, acyclic, connected graph with  $V= \{0, 1, \dots, n\}$  as the set of vertices and  $E=\{(x, y)|x, y \in V\}$  as the set of edges. One distinguished vertex  $r \in V$  is designated the root, and for all  $x \in V$ , there is a unique path from  $r$  to  $x$ . Further,  $l: V \rightarrow L$  is a labeling function mapping vertices to a set of labels  $L=\{l_1, l_2, \dots\}$ . If  $x, y \in V$  and there is a path from  $x$  to  $y$ , then  $x$  is called an ancestor of  $y$ . If  $x$  is an immediate ancestor of  $y$ , then  $x$  is called the parent of  $y$ , and  $y$  the child of  $x$ .

Given a tree  $S=(V_s, E_s)$  and tree  $T=(V_t, E_t)$ , we say that  $S$  is an isomorphic subtree of  $T$  iff there exists a one-to-one mapping  $\varphi: V_s \rightarrow V_t$ , such that  $(x, y) \in E_s$  iff  $(\varphi(x), \varphi(y)) \in E_t$ . If  $\varphi$  is onto, then  $S$  and  $T$  are called isomorphic.  $S$  is called an *induced subtree* of  $T=(V_t, E_t)$ , denoted  $S \leq_i T$ , iff  $S$  is an isomorphic subtree of  $T$  and  $\varphi$  preserves labels, i.e.,  $l(x)=l(\varphi(x)); \forall x \in V_s$ . That is, for induced subtrees,  $\varphi$  preserves the parent-child relationships, as well as vertex labels.  $S=(V_s, E_s)$  is called an *embedded subtree* of  $T=(V_t, E_t)$ , denoted as  $S \leq_e T$  iff there exists a 1-to-1 mapping  $\varphi: V_s \rightarrow V_t$  that satisfies: 1)  $(x, y) \in E_s$  iff  $\varphi(x) \leq \varphi(y)$  and 2)  $l(x)=l(\varphi(x))$ . That is, for embedded subtrees,  $\varphi$  preserves ancestor-descendant relationships and labels.

## Frequent Subtrees

Support. If  $S \leq_{(i, e)} T$ , we also say that  $T$  contains  $S$  or  $S$  occurs in  $T$ . Note that each occurrence of  $S$  in  $T$  can be identified by its unique match label, given by the sequence  $\varphi(x_0)\varphi(x_1) \dots \varphi(x_{|S|})$ , where  $x_i \in V_s$ . That is, a match label of  $S$  is given as the set of matching positions in  $T$ . Let  $\delta_T(S)$  denote the number of occurrences of the subtree  $S$  in a tree  $T$ . Let  $d_T$  be an indicator variable, with  $d_T(S)=1$  if  $\delta_T(S)>0$  and  $d_T(S)=0$  if  $\delta_T(S)=0$ . Let  $D$  denote a database (a forest) of trees. The support of a subtree  $S$  in the database is defined as  $\sigma(S)=\sum_{T \in D} d_T(S)$ , i.e., the number of trees in  $D$  that contain at least one occurrence of  $S$ . Typically, support is given as a percentage of the total number of trees in  $D$ . A subtree  $S$  is frequent if its support is more than or equal to a user-specified minimum support (*minsup*) value. We denote by  $F_k$  the set of all frequent subtrees of size  $k$  (also called a  $k$ -subtree).

## Representing Trees as Strings

We represent a tree  $T$  by its string encoding, denoted  $string(T)$ , generated as follows: Add vertex labels to  $string(T)$  in a depth-first preorder traversal of  $T$  and add a unique symbol  $-1 \notin L$  whenever we backtrack from a child to its parent. This format allows us to conveniently represent trees with an arbitrary number of children for each node.

We use the above-mentioned representing method, instead of the standard data structures, such as the adjacency-matrix, the adjacency-list, and the first-child-next-sibling representation, for several reasons. First, string encoding is more compact than standard data structures and hence save space. Second, perhaps most importantly, for some types of labeled trees, there can be multiple ways to represent the same tree using the standard data structures. A string encoding is a unique way to represent a labeled tree.

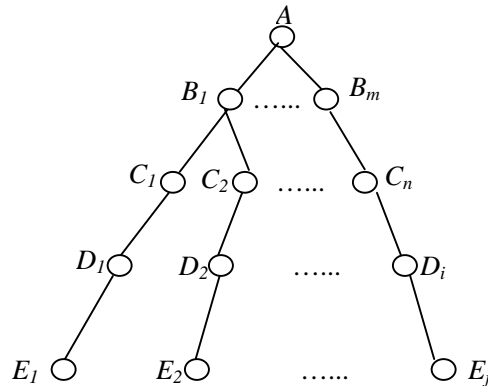
## MARKETING CHANNEL MANAGEMENT IN ERDOS CASHMERE GROUP

ErDOS cashmere group co. LTD. is one of the most famous cashmere products processing firm. Till now, the company has 32 sales corporations, 19 business representative office, 31 merchandise dispatching and distributing center and more than 1,000 sales net around China. Thus, the ERDOS group has established one of the largest marketing channel of the domestic textile industry. Recently, with the rapid computerization of the whole marketing channel, large amount of sales data has been collected. How to make full use of the data to enhance the existing marketing channel is of great importance. To solve this problem quantitatively, rather than only qualitatively, we use frequent subtree mining to discover the interesting tree patterns.

## Representation of Database on Marketing Channel

Generally speaking, the whole marketing channel in ERDOS group can be viewed as a *tree* with five levels (Shown in Fig. 1). In Fig. 1, the nodes  $A; B_1, \dots, B_m; C_1, \dots, C_n; D_1, \dots, D_j; E_1, \dots, E_j$  represent different sales members of the first level, the second level, the third level, the fourth level and the fifth level respectively. Nodes in different levels may represent different channel members, e.g., agents, wholesalers, distributors, and retailers.

Figure 1: Example tree structure of marketing channel in ERDOS.



We have collected sales data of each node from 2002 to 2006, and separated data monthly. That is the database contains  $12 \times 5 = 60$  transactions, and each item in the transaction corresponds to one node in the marketing channel tree in Fig. 1. Note that, we do not need list all the nodes in every transaction. The method we use is that, only those sales elements pass the basic monthly sales plan will be recorded in the transaction corresponds to the same month. For example, suppose the basic monthly sales plan for the nodes in the last level is 50,000 in Oct. 2006, if certain sales element  $E_i$  does not complete the baseline, then  $E_i$  will not be included in the transaction corresponding to Oct. 2006.

Note that the representation of database discussed in this section partially corresponds to the first three areas described in Section 2, i.e., (1) formulating channel strategy, (2) designing marketing channels, (3) selecting channel members.

### Mining Frequent Subtrees

In the way described in Section 4.1, we build a database of marketing channel in ERDOS group with 60 trees, and each tree has  $(A + \sum_{k=1}^m B_k + \sum_{k=1}^n C_k + \sum_{k=1}^i D_k + \sum_{k=1}^j B_k)$  nodes at most. Give minimum support threshold *minsup*, our task is to mine all the *frequent labeled embedded subtrees*. Note that, we mine the embedded, rather than only induced subtrees, because embedded subtrees are a generalization of induced subtrees; they allow not only direct parent-child branches, but also ancestor-descendant branches. As such, embedded subtrees are able to extract patterns “hidden” (or embedded) deep within large trees which might be missed by the traditional definition.

For the specific tree mining algorithm, we use TreeMiner (Zaki 2005). The TreeMiner algorithm follows the combined depth-first/breadth-first traversal idea to discover all frequent embedded subtrees from a database of rooted ordered trees. Other than the general downward closure property (i.e., all subtrees of a frequent tree are frequent), TreeMiner takes advantage of a useful property of the string encodings for rooted ordered trees: removing either one of the last two vertices at the end of the string encoding of a rooted ordered tree  $P$  (with correspondent adjustment to the number of backtrack symbols) will result in the string encoding of a valid embedded subtree of  $P$ . Furthermore, the notions of scope-lists and rightmost extension were introduced in that work. More detail on the specific algorithm can be found in (Zaki 2005).

The TreeMiner algorithm was used in building a structural classifier for XML data (Zaki 2003). Another application of TreeMiner focuses on Bioinformatics applications to mine distinct occurrences of trees: finding common RNA structures and mining common phylogenetic subtrees (Zaki 2005). To the best of our knowledge, there are few papers on studying marketing channel management by using frequent subtree mining.

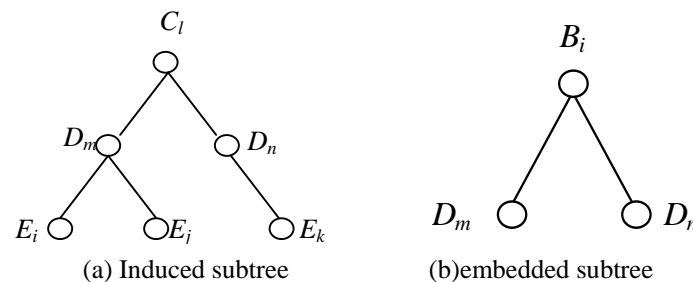
### Usage of Discovered Patterns

We classify the discovered frequent subtrees into two categories. Different strategies are used to deal with different categories respectively.

1) The discovered frequent *induced* subtrees may represent the well-organized marketing channels. Take frequent tree in Fig. 2 (a) for example. The elements in this kind of frequent trees should be retained and popularized.

2) The discovered frequent *embedded* subtrees may represent the hidden interesting alternatives of existing marketing channel. For example, Fig. 2 (b) shows a discovered frequent embedded tree. We can see that the immediate parents of  $D_m$  and  $D_n$ , and the immediate children of  $B_i$  are not included in the frequent pattern. Maybe the existing marketing channel can be adjusted: 1) lower the level of  $B_i$ ; 2) upper the level of  $D_m$  and  $D_n$ ; 3) or delete the channel elements between  $D_m$ ,  $D_n$  and  $B_i$  directly.

**Figure 2: Two classes of example discovered frequent subtrees.**



Note that the frequent subtree mining and usage of discovered patterns in section 4.2 and 4.3 partially correspond to the last four areas described in Section 2, i.e., (4) motivating channel members, (5) coordinating channel strategy with channel members, (6) assessing channel member performance, and (7) managing channel conflict.

### CONCLUSIONS

The management of marketing channels plays an important role for the success of most manufactures. Most existing work studies the problem of marketing channel management in a qualitative way. Recently, with the increase of amount of sales data, how to enhance the marketing channel quantitatively is significant. As the marketing channel can be viewed as a tree, in this paper, a new marketing channel management strategy based on frequent subtree mining is proposed. The proposed method is illustrated under the real-world sales data in ERDOS group. Firstly, the tree transaction is formed monthly. For each monthly transaction, only those channel members that pass the basic sales plan will be included. Secondly, we use the TreeMiner algorithm to discover embedded frequent subtrees. Finally, different management strategies are used for different kinds of discovered patterns.

To the best of our knowledge, it is the first time to exploit the problem of marketing channel management by using frequent subtree mining. So our work is explorative. In our future work, we should enhance the proposed strategy step by step in practice, and popularize the method to other marketing channels similar to that of cashmere.

**REFERENCES**

- Achrol, R. S., Stern, L. W. (1988). Environmental Determinism and Decision Making Uncertainty in Marketing Channels. *Journal of Marketing Research* 25 (1), 36-50.
- Anderson, E., Day, G. S., Rangan V. K. (1997). Strategic Channel Design. MITSloan *Management Review* 38 (4), 59-69.
- Coughlan, A., Anderson, E., Stern, L. W., El-Ansary, A. (2005). *Marketing Channels*. Englewood Cliffs, NJ: Prentice Hall.
- Han, J. W., Kamber, M. (2006). *Data Mining: Concepts and Techniques*. San Francisco, CA: Morgan Kaufmann.
- Mehta, R., Rosenbloom, B., Anderson, R. (2000). The Role of The Sales Manager in Channel Management: Impact of Organizational Variables. *Journal of Personal Selling and Sales Management* 20 (2), 81-88.
- Pelton, L., Lumpkin, J. R., Strutton D. (2001). *Marketing Channels: A Relationship Management Approach*. McGraw-Hill.
- Rosenbloom, B. (1987). *Marketing Functions and the Wholesaler-Distributor: Achieving Excellence in Distribution*. Washington, DC: Distribution Research and Education Foundation.
- Rosenbloom, B. (1999). *Marketing Channels: A Management View*. Hinsdale, IL: Dryden Press.
- Stern, L., El-Ansary, A., Coughlan, A. T. (1996). *Marketing Channels*. Englewood Cliffs, NJ: Prentice-Hall.
- Washio, T., Raedt, L. D., Kok, J. N. (2005). Advances in Mining Graphs, Trees and Sequences. *Fundamenta Informaticae* 66 (1-2), v-viii.
- Zaki, M. J., Aggarwal, C. (2003). XRULES: An Effective Structural Classifier for XML Data. Proceedings of 9th International Conference on Knowledge Discovery and Data Mining, 316-325.
- Zaki, M. J. (2005). Efficiently Mining Frequent Trees in a Forest: Algorithms and Applications. *IEEE Transaction on Knowledge and Data Engineering* 17 (8), 1021-1035.