SMART PLANNER : Divide-and-Conquer with Soft Computing Approach

Amit Saxena, Shivendra Prasad Tiwari, Megha Kothari
Computer Science Department, G.G.University, Bilaspur C.G.- 495009

Abstract

Data mining is an emerging area to discover knowledge from a tremendously large database or data warehouse. Plan mining is an extension activity of data mining. Given a set of plans, the objective is to enable user to know the interesting plans. With the proliferation of database technology, planbase are increasingly common.Efficient patterns of actions discovery, valuating the planbase, and deriving the concise plan is the current challenge to data mining. Soft computing is a consortium of intelligent techniques like Neural Networks, Fuzzy theory, evolutionary computing etc. This paper proposes a soft computing approach to develop a planminer. The attributes of the patterns of the planbase are fuzzified in the first phase. Next, support is associated to each plan. A rule base is generated which is input to a neural network for learning and prediction.

Keywords:- plan mining, Neural Network, Smart Planminer, Learning Algorithm

1. Introduction

With the advent of modern storage devices with capacity in GB, internet, web pages and enormous data is being stored every day. At the moment the major problem lies due to excess of data but lack of meaningful information[1]. Plan Mining is a task of mining significant patterns or knowledge from planbase, a collection of plans. A plan consists of a varying sequence of actions. In sequence mining, a large number of frequently occuring sequences are mined at a very detailed level whereas in the plan mining important or significant generalized patterns are mined or extracted from a plan base.

The application domains of the plan mining includes finding out the significant patterns are used in the maintenance plans for a mechanical device, in the medical domain to uncover the defects in the treatment process and information gathering, Discovering the knowledge and travel patterns of bussiness travelers in an huge online database. Association rule mining, **weblog** mining, Weather forcasting pattern mining, Roadways mining information gathering etc are few emerging issues.

P#	A#	D-Loc	D-Time	A-Loc	A-Time	A-line	Gains ($)
1	1	DLH	800	BMB	1000	IA	90000
1	2	BMB	1000	RPD	1400	HA	225000
1	3	RPD	367	SNG	1967	PA	320000
1	4	SNG	1756	NKN	1956	CA	77000
1	5	NKN	654	KMD	1154	CA	598000
2	1	JBP	1567	CAL	1867	AI	54000
2	2	CAL	1100	DHK	1540	BA	165000
2	3	DHK	345	DSP	1745	BA	123000
2	4	DSP	2000	KTK	2200	IA	87000
..	..	..	..	..	..	..	..

Table 1 : A data base of travel plans **:** An Air Flight Planbase of a bussiness traveler

Zadeh introduced Fuzzy set in 1965[zadeh] to represent data and information processing, non-statistical uncertainty. In this paper, fuzzy logic is used to fuzzify the values of attributes that can be generalized. The higher flexibility is the characteristics feature of neural network produced by learning and hence this suits data driven processing better. Neural networks can blindly generate and refine fuzzy rules from training data [3].

Previous work in plan mining area are divided in several categories. In the area of plan recognition,the objective is to uncover intention and goal of a series of observations and actions. Examples are the works in [2-5,7,8,11,12]. Zaki [9] developed a technique for plan mining which first performs brute-force expansion of all the nodes and then mines sequence patterns using the apriori technique [10] on the data which indicate high incidence of plan failure. They added that knowledge to the planner so that the paths which often result in failure are avoided by the planner. The new approach developed in [11] , is a method for mining significant patters of succesful actions in a large database using Divide –and-Conquer strategy.

In this paper, sequence patterns involving multidimension attributes like gain, cost, time taken, the reachability and many other deciding factors of the nodes are considered. The planbase is generalized by sequencing and using merge operator for the repeating occurences to make highly concised and generalized planbase. The subsequences are generated for all the sequence and their support is calculated to lookup the strengh of the partitioned planbase. On the basis of the highly generalized partioned planbase and their support, the common characteristics properties are derived. To make the system faster and more smart the chacteristics rulebase is sent to the learning system.

2. Problem Outline

Plan mining extracts the important and significant patterns from the plan base which in turn are used by domain experts, planners,or plan valuator. This study is carried out to describe the given database, how to work out on such databases to gather maximum information and queries requested by the user. Consider an air-travel planbase as shown in Table 1, with plan sequences of travel by air. In the planbase each record corresponds to an action ,and the sequence of records having the same plan number are considered as one plan with a sequence of actions. Table 2 stores the detailed information about the airports.Table 1 has an attribute ie Gain or Bussiness amount, which will be further useful to decide the appropriateness of the nodes or actions.

A-Code	City	State	Region	A-Size	…
DLH	Delhi	Delhi	India	1800	…
BMB	Bombey	Maha.	India	1680	…
RPD	Ravalpindi	Raval	Pak	540	…
SNG	Shanghai	Shang	China	900	…
NKN	Nanking	Shang	China	1000	…
KMD	Kathmandu	Kathm	Nepal	556	…
JBP	Jabalpur	MP	India	440	…
CAL	Calcutta	WB	India	530	…
DHK	Dhaka	BD	BD	1975	…
DSP	Dispur	Asam	India	587	…
KTK	Cuttuck	Orissa	India	1189	…
..	..	..	..	..	…

Table 2 : An airport information table

D-Loc	A-Loc	Cost ($)	Dist KM	Time Hrs	...
DLH	BMB	138	2650	2.0	...
BMB	RPD	1823	5845	4.0	...
RPD	SNG	1754	9076	6.0	...
SNG	NKN	580	2345	3.0	...
NKN	KMD	600	6554	5.0	...
JBP	CAL	430	2365	2.0	...
CAL	DHK	1650	7650	4.4	...
DHK	DSP	1160	6943	4.0	...
DSP	KTK	440	2645	2.0	...
..	..	..	..	..	...

Table 3: The assumed database of requirements and features of actions.

The given databases are used to find out miniable patterns. Table 1 has some such attributes which gives information about the sequence of actions, the departure and arrival nodes, the airline to be used, major factor i.e. Gain attribute for each action etc. These attributes of given planbases are to be co-related with the Tables 2 and 3. Table 2 is able to provide some major informations like reachbility, state and the region of the node. For example, if the airport code to any location is not available in the airport database then it is clear that the node is not reachable by plan, if the cost and the time taken is very high than that of the corresponding bussiness amount or gain, then these type of cases of actions should be avoided by the traveler, and the plan should be modified. The planminer has to generate characteristics rules, to be used by the planner, without searching the whole database. But the characteristics rules should be covering the ,maximum areas of queries. Indiscriminate mining of such kind of "rule" may result in a large number of scattered rule only without substantial support, but still missing a clear overall picture. What is to find, is a small number of general (sequential) patterns which may cover a substantial portion if not most of plans, and then search efforts can be divided based on such mined sequences .

A multidimensional database model [3] for the air flight planbase, can be used to facilitate such plan generation. With multi-dimensional model, one may generalize the planbase in different directions and observe when the generalized plans may share some common interesting sequential patterns with.

P#	A#	Cost	Gain	Reach.	A_Size	State	Region	..
1	1	MC	LG	HR	L-L	D-M	I-I	..
1	2	HC	MG	LR	L-S	M-R	I-P	..
1	3	HC	HG	HR	S-M	R-S	P-C	..
1	4	LC	LG	MR	M-M	S-S	C-C	..
1	5	LC	HG	LR	M-S	S-K	C-N	..
2	1	LC	LG	LR	S-S	M-W	I-I	..
2	2	HC	MG	LR	S-L	W-B	I-B	..
..	.	..	..	..	..	..	..	..

Table 4. The database after fuzzifying the Values of generalizable attributes

To find out the values of the attributes in generalized form , the membership function is used as in[16].

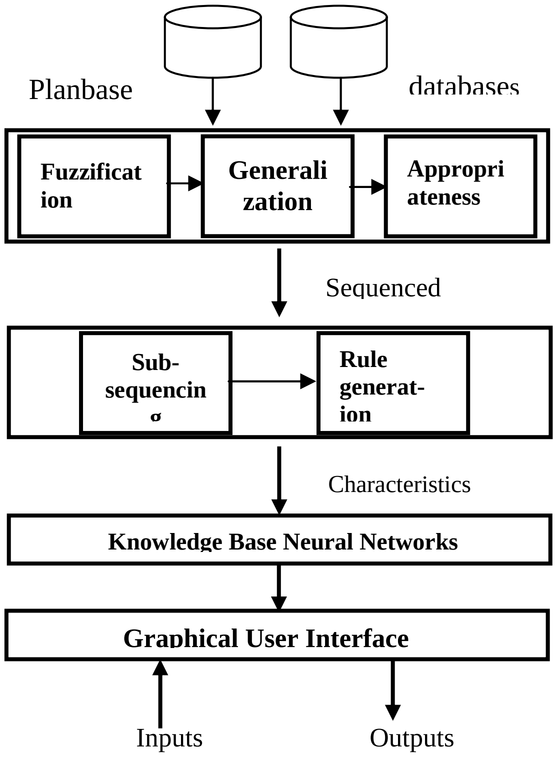

Fig. 1 Schematic Diagram of Smart Planner

Algorithm -

Input:- A planbase and its dimension tables.

Output :- The Fuzzified generalization based sequential patterns of the planbase.

Steps

i. Fuzzification of the attributes:- The deciding attributes like size, cost, reachability, Gain etc. are fuzzified to find out linguistic values using

the trapezoidal fuzzy membership function[16] defined by the eq. 1.

ii. Generate the sequence-planbase.

Generalize the planbase by method [11].

iii. Estimation of generalizable values: - Based on the data in the association dimension table of the planbase, estimate the number of distinct values which may be generalized from the sequence planbase and determine whether it is useful to examine the generalized plan.

iv. Generalization of the planbase at a particular level of abstraction:-Perform multidimensional generalization of attributes, as the AOI developed in [6]. This is similar to the rollup operation in OLAP [3].

v. Apply merge operator "+".

vi. Finding the Most deciding factor :- Based on the various deciding attributes like Gain sequence, Cost sequence, Size sequence etc the Appropriateness-sequence factor is calculated by the method in [14] , which plays a major role to make the decision by the smart planner or plan miner.

vii. Calculate the fuzzified support for each sequence

viii. Use the decision making operators [], () :- The merged sequential patterns derived in the above steps are applied with the decision making oprators [] for option and () for discardable.

Example: On the basis of above algorithm ,the plan base of Tables 1, 2 and 3 are collectively put together on the basis of a releated attributes and corresponding fuzzy values as Table 4. The generalized sequence is:

Delhi-Bombey-Ravalpindi-Shanghai-Nanking-Kathmandu.

The plans can be generalized in different ways e.g. routes may be generalized on the basis of some factors like **Size;** the suitability and the conciseness can be generalized on deciding factors like **Cost, Gain, Reachability, Time taken, Distance** etc. The attributes have fuzzified values e.g. Size factor may have linguistic membership values L (large), M (medium) and S (small),Gain may have HG (high gain), MG (medium gain),and LG (low gain),

P#	Cost	Gain	A_Size	Region	...
1	MC-HC-HC-LC-LC	LG-MG-HG-LG-HG	L-L-S-M-M-S	I-I-P-C-C-N	... ...
2	LC-HC-MC-LC	LG-MG-MG-LG	S-S-L-S-M	I-I-B-I-I	... ...
.	..	..	..	..	

Table 5. Multidim.Gen. Of assumed plan base.

P#	Cost	Gain	A_Size	Reg	...
1	MC-HC-LC^{+}	LG-MG- -LG-HG	L^{+}-S-M^{+}-S	I^{+}-P-C^{+}-N	... ...
2	LC-HC-MC-LC	LG-MG^{+}-LG	S^{+}-L-S-M	I^{+}-B-I^{+}	... ...
..	..	..	..	..	...

Table6. Merging consecutive, identical actions.

P#	App-seq
1	HA-MA-MA-MA-HA
2	MA-LA-HA-MA
..	..

Table7

P#	App-seq	Sup
1	HA-[MA]-HA	79%
2	[MA]-(LA)-HA-MA]	59%
..	..	..

Table 8

After getting these linguistic fuzzy values we can generate a factor which will lead to these all deciding factors. The degree of the appropriateness of the action can be calculated using "Decision Rules" method in [14] . Hence, we can find values HA (highly appro.), MA(medium appro.) and LA (low appro.), the HA valued nodes can be considered as a major bussiness hubs in [11] Rules to get Appropriate attribute may be -

if Flight(x,y) & Appropriateness(x,HA) & Appropriateness(y,LA) then Cost(y)>Cost(x) or Gain(x)>Gain(y) or Reachability(x)>Reachability(y). [conf.] ;
And If Flight(x,y) & AirportSize (x,S) & AirportSize (y,L) then Region(x)=Region(y). [conf.] [11]; if Action(x,y)&Appropriateness(x,HA) & Appropriateness(y,HA) thenCost(x)~Cost(y) is Low or Gain is high or Ratability's high, etc.

The sequence can be replaced by the sequence of their appropiateness values like **HA-MA-MA-HA-MA-MA.**
MA-MA-HA-MA-MA.

In the same way on the basis of size, cost ,gain the generalized sequences may be stored in the Fuzzified generalized planbase. The more generalized form may be highly regular pattern in the form of HA-MA^{+}-HA-MA^{+} ,MA^{+}-HA-MA^{+} for appropriateness factor. Some more oprators like [], ( ) for indicating the status of the action .

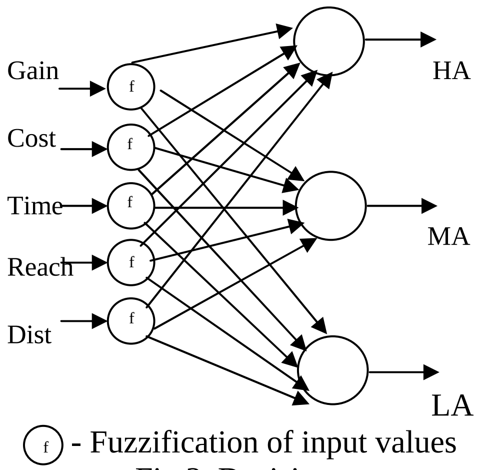

Fig 2. Decision tree

3.2 Subsequencing and rule generation Algorithm:-

Input : The fuzzified generalized planbase.
Output: Commonn characteristics rulebase for each partitioned planbase.

Steps:-

i. Subsequencing of generalized sequence patterns using method in [1].

ii.Constructing the Star/Snowflack data warehouse model:- The resulted subsequences are further stored in the star data warehouse model.There is a fact table and many other dimension tables for each generalized sequenced patterns in which the subsequences with their supports are stired. Only that subsequences are placed which fullfil the defined minimum criteria.

iii. Calculate Support for each subsequence using method in [1].

iv.Selection of subsequences:- The greater support valued subsequences are selected to derive the concise plans.

v.Mine characteristics properties:- Finally, the common characeristics rules are generared for each subsequence pattern on the corresponding partitioned planbase as in [11].

Example: Sub goal identification and its support calculation is performed as in [8,5] and [11]. Now the support is calculated to each subsequence, and store in snowflack data warehouse, and tables 6,7, and 8 are created. Now, on the basis of the subsequences and partitioned patterns, the characteristics rules dy decision table method in [14] and the concise plans are derived. After all a large rule base is created.

3.3 Making the system more intelligent and faster: A learning and updation process

This subsection is one of the most important part of our method.Once the Algorithm 3.2 produces the common characteristics rulebase, the planner or the smart planning system has an overhead to search the rulebase to answer the user's query. If the rule base is forwarded to a Neural Network learning system, then to answer the query, system will not have to search overall rulebase.Mapping of characteristic rule into the neural architecture can be done as in [12].

4.Conclusion

The paper considers an air flight planbase as a problem to mine out the essential data, and to produce rule base by fuzzification and generalization The rule base is learned and refined by an Neural Network, making the system more faster and smart. User can input a planbase receive output as knowledge with GUI. The proposed plan mining technique can be implemented in constructing a smart planner. The work reported earlier constructs a planbase on the basis of bivalent nature of attributes whereas proposed system takes into consideration the multivalued nature of significant attributes for generalization using fuzzy logic. The extension to the current investigation aims to implement actually the soft computing approach on a real database, which will be a scope for researchers.

References

[1] Jiawei Han, Micheline Kamber, Data Mining Cocepts and Techniques, Morgan Kaufmann Harc.India 2001.

[2] J.F.Allen,H.A. Kauth, R.N. Pelavia and J.D.Tenenberg. Reasoning about plans. Morgan Kaufmann publishers .1991.

[3] S.Woods , A.Quilici and Q. Yang. Comnstraints based design recovery for software reengineering : Theory and Experiments.

[4] J.F.Allen ,J. Hendeler and A.Tate. Readings in planning. Margon Kaufmann publishers, 1990.

[5] Q. Yang .Intelligent planning – Decomposition and abstraction based approach . Springer – Verlag 1997.

[6] D.Weld . An Introduction to least – Commitment planning .AI Magagine , Winter 1994:27-61, 1994.

[7] O.Etzioni. Aquiry search –Control knoledge via static analysis .AI 62(2) : 225-302, 1998.

[8] S. Minton, J.G. Carbonell ,C. A. Knoblock ,D. R. Kuokka, O. Etzioni and Y. Gill.Explaination based learnig :A problem solving perspective . AI 40 :63-118, 1985.

[9] M.J. Zaki,N. Lesh , and M.Ogihari.PLAN MINING : Sequence Mining for plan failures .In proc. 4th Int. Conf. Knowledge discovery and data mining (KDD'98) New York, NY ,August 1998.

[10] R. Shrikant and R.Agrawal .Mining Sequential patterns : Genralizations and Performance improvements .In proc. 5th Int. Conf. Extending database technology (EDBT), pp.3-17, Avignon, France ,March 1996.

[11] Jiwei Han, Qiang Yang and Edward Kim .Plan Mining By Divide –And – Conquer . School of Computing Science , Simon Eraser University surnaby, BC Canada ,VSA 1S6. 1999.

[12] Fu,L.M. and Fu,L.C.1990. Mapping rule based systems into neural architecture. Knowledge based systems, 3(1),pp 48-56.

[13] B.Kosko Neural Networks and Fuzzy systems. Englewood Cliffs, NJ:printice Hall 1991.

[14] K.J. Cios, W. Pedryez, R.W.Swiniarski. DATA MINING Methods for KNOWLEDGE DISCOVERY.Kluwer academic publishers.2000.

[15] L.A.Zadeh, "Fuzzy sets", information and control, vol 8,1965.

[16] G.J.Klir and B.Yuan ,Fuzzy logic :Theory and applications in intelligence systems. Boston :Kluwer Academic press,1992.