

A Method for Solving Medical Treatment Problems in Hospitals of China based on Multi-Dimensions Data Analysis

Guan-qi Xu

School of Mechanical Engineering
Shanghai Jiao Tong University
Shanghai, China
wangliya@sjtu.edu.cn

Zhi-bin Jiang

School of Mechanical Engineering
Shanghai Jiao Tong University
Shanghai, China
zbjiang@sjtu.edu.cn

Li-ya Wang

School of Mechanical Engineering
Shanghai Jiao Tong University
Shanghai, China
wangliya@sjtu.edu.cn

Jin-ze Chai

School of Mechanical Engineering
Shanghai Jiao Tong University
Shanghai, China
tju_chai@163.com

Abstract—With the medical technology of China advances rapidly, over medical treatments such as expensive drug preference, over antibiotics and hormone drugs, repeated examinations and so on are also increasing. In order to mitigate the problems, this paper proposed a method of multi-dimensions data analysis to analyze historical medical data and find out the best treatments for recommendation, also the worst to avoid. An algorithm based on k-dominant Skyline Operator is used in the method to sort historical medical data of inpatient with multi-dimensions. The results can provide doctors with reasonable medical treatment advices and patients with suitable prescription respectively.

Keywords—Over Medical Treatment; Medical Data; Multi-Dimensions Analysis; K-dominant Skyline Operator

I. INTRODUCTION

With the development and progress of the society, the progress of medical technology in China is remarkable. However, medical treatment environment is not so good. In China, there are many medical treatment problems such as expensive medical cost and repeated physical examinations. Some researches gave that unreasonable prescriptions and examinations lead to 20% to 30% waste of medical resources [1]. For this issue, many scholars presented different solutions. H.Yan [2] advised government to build up medical security system, improve medical mechanism and so on. S.Xu [3] applied credence good theory to analyze the game between doctors and patients in the medical market. He also explored the settlement mechanism by referring to advanced experience in western countries. R.Han [4] used data mining technique in the clinic analysis of rational drug use by applying result of association rules discovery, to alleviate economic burden of patients. Considering huge amount of valuable historical medical data in Hospital Information System(HIS), we proposes a new approach by sorting and analyzing the data to find best treatment cases with low medicine and examination fee or short time treatment course, which provide patients with

suitable prescription and doctors with reasonable medical treatment advices. On the other hand, those overtreatment cases can also be found to avoid in future. In this paper, a multi-dimensions model of inpatient medical data model is built up at first. Next, an algorithm based on k-dominant Skyline Operator [5] is used to sort inpatient historical medical data. Finally, the correctness and effectiveness of algorithm is proved by experiments.

II. MULTI-DIMENSIONS MODELING OF INPATIENT DATA

Considering multi-dimensional characteristics of medical data such as medicine data, examination data, in-patient data and many other dimensions, inpatient medical data multi-dimensions model should be built up first for analyzing historical medical data. Figure 1 shows star model of inpatient medical multi-dimensions data.

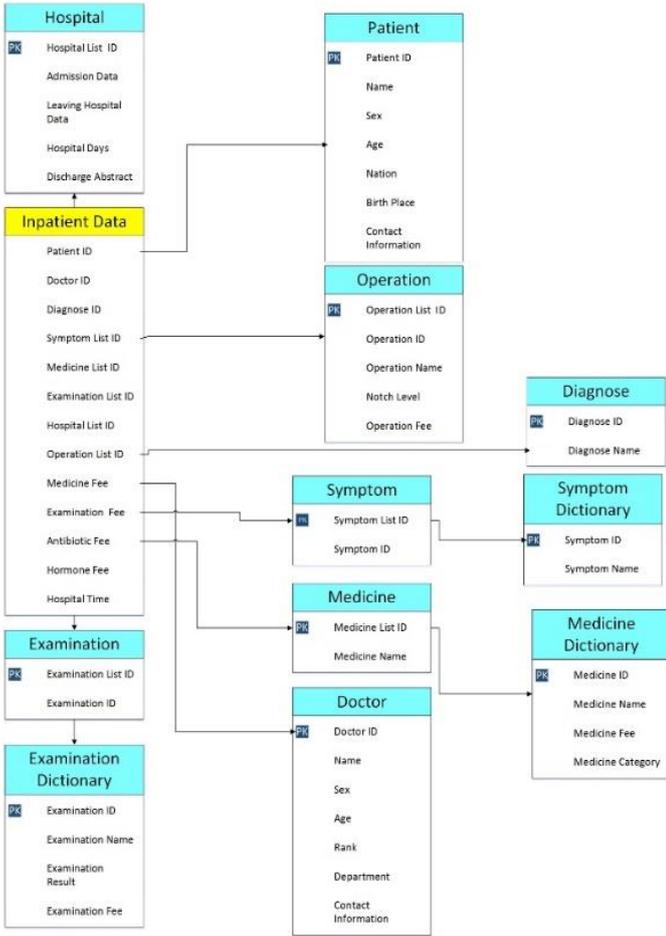


Fig. 1. Inpatient Medical Data Multi-dimensions Star Model

The whole inpatient multi-dimensions model (IMDM) consists of eight dimensions: patient, doctor, diagnoses, symptom, examination, prescription, in-patient and operation. Three data dictionaries: examination, symptom and medicine dictionary. There are five measures: examination fee, medicine fee, antibiotics fee, and hormone fee and hospital days. Antibiotics fee and hormone fee are set as Y.C.Chen [6] presents that use of antibiotics and hormone should be controlled between medical treatments in inpatient. On the other hand, other 3 measures are apparent measures of optimization in inpatient.

III. SORTING METHOD BASED ON IMDM

A. K-dominant Skyline Operator

Sorting method based on IMDM is a multi-dimensions sorting problem, five dimensions including 5 measures set in IMDM should be taken in consideration and meaningful results would be obtained. Regular sorting algorithm such as Bubble Sort and Heap Sort are unsuitable for multi-dimensions sorting problem. Basic Skyline Operator is also unsuitable because Basic Skyline Operator has a shortcoming that as number of dimensions increases, quantity of result is larger[7], which lead to its low effectiveness in high dimensional place. However, k-dominant Skyline Operator overcomes the shortcoming so

that k-dominant Skyline Operator is chosen to process sorting problem.

K-dominant Skyline Operator: For a d-dimensions space $S = \{s_1, s_2, \dots, s_d\}$, data set $D = \{p_1, p_2, \dots, p_n\}$ is a data set on S when and only when a d-dimension is point on S.

$p_i.s_j$ is used to present the value on dimension j of point p_i . For every dimension s_i , assume there is relationship of order which is expressed as \succ_i , according to users' demand, the \succ_i this place can be $<$ or $>$. Without loss of generality, less value of measures are better so that \succ_i is replaced by $<$ infra.

Definition 1: k-dominant. A point p_i dominates point p_j on space S when and only when $\exists S' \subseteq S, |S'| = k$, let $\forall S_r \in S'$, has $p_i.s_r \leq p_j.s_r$ and $\exists s_i \in S'$ let $p_i.s_r < p_j.s_r$.

Definition 2: k-dominant Skyline Set. Point p_i is a k-dominant skyline point when and only when there isn't a point p_j in D, $p_j \neq p_i$ and let p_j k-dominant p_i . $DSP(k, D, S)$ is used to present all k-dominant point in data set D of space S.

B. Inpatient Multi-dimensions Sorting Method One-Scan Algorithm

One-Scan algorithm follows concept of k-dominant Skyline Operator. Here is the thought of One-Scan algorithm. For each point p in data set D, p is first compared with point in T (step 5 to 11). If a point $p' \in T$ is dominated by point p, point p' is not a basic skyline point, then p' is removed from T; Otherwise, if p' is dominated by p or $p' = p$, then p can be ignored. However, if we can make sure p is a basic skyline point, and then p is further compared against the points in R (step 13 to 22). For each $p' \in R$, if p k-dominant p', then p' is deleted from R and moved to T; If p' k-dominant p, then p is not a k-dominant point. Finally, p is inserted to R if p is a k-dominant skyline point; otherwise, p is inserted to T since it's a basic skyline point. Once all points in D is processed, R contains all k-dominant skyline points in D.

And at the start of the algorithm, all points are sorted in advance for a random dimension. The target is to reduce calculation of the algorithm and basic skyline point will be easy and quick to find.

One-Scan Algorithm (D,S,k)

- 1:sort D for a random dimension
- 2:initialize set of k-dominant skyline points $R = \emptyset$
- 3:initialize set of basic skyline points $T = \emptyset$
- 4:for each $p' \in T$ do

```

5: initialize isUniqueSkyline = true (the first point is moved
to T)
6:   for each do
7:     if (p dominates p') then
8:       delete p' from T
9:     else if (p' dominates p) or (p' = p) then
10:      isUniqueSkyline = false
11:      break out of inner for-loop
12:   if (isUniqueSkyline) then
13:     initialize isDominant = true (the first point is moved to R)
14:     for each p' ∈ R do
15:       if (p' k-dominant p) then
16:         isDominant = false
17:       if (p k-dominant p') then
18:         moved p' from R to T
19:   if (isDominant) then
20:     insert p into R
21: else

```

```

22: insert p into T
23: return R

```

C. Result Analysis

One-Scan Algorithm is realized on Visual Studio 2013 and Sql Server 2014. 100 test records of 3 diagnoses are set in database and $k=4$ in our algorithm. Here is the test result.

1) Relationship of k -dominant Skyline Points:

From Figure 2, for diagnose 1, 5 records are recommended. If we use vector medicine fee, antibiotic fee, hormone fee, examination fee, hospital days to present them, these 5 records can be expressed as follows: (779,96,96,185,23), (845,105,94,195,21), (888,95,103,183,19), (946,91,103,213,14) and (989,99,94,201,18). For example, (779,96,96,185,23) and (989,99,94,201,18) are chosen for comparison: it's easy to find that in medicine fee, antibiotics fee and examination fee dimensions, the values of first records are all better than second records and worse than second records in other two dimensions. After comparing relationship of each two records, we find that the relationship is all 3 better-2 worse since k is equal to 4. So that among these 5 records, it doesn't exist 4-dominant relationship which conform to the definition of k -dominant skyline point.

2) Relationship of k -dominant Skyline Points and Other Points:

diagnose id	medicine fee	antibiotic fee	hormone fee	examination fee	hospital days	id
1	779	96	96	185	23	41
1	845	105	94	195	21	49
1	888	95	103	183	19	51
1	946	91	103	213	14	15
1	989	99	94	201	18	71

Fig. 2. Result of One-Scan Algorithm

diagnose id	medicine fee	antibiotic fee	hormone fee	examination fee	hospital days	id
1	766	100	102	202	26	11
1	779	96	96	185	23	41
1	806	102	98	196	21	64
1	845	97	105	199	19	6
1	845	105	94	195	21	49
1	861	96	100	220	24	50
1	888	95	103	183	19	51
1	916	105	101	198	21	62
1	921	101	108	203	23	47
1	931	103	102	185	24	78
1	946	91	103	213	14	15
1	946	99	90	201	23	20
1	962	96	100	205	23	14
1	971	95	97	217	23	74
1	979	107	106	205	21	25
1	980	99	97	196	22	7
1	987	104	103	191	21	46
1	989	99	94	201	18	71

Fig. 3. All Records of A Diagnose

diagnose id	medicine fee	antibiotic fee	hormone fee	examination fee	hospital days	id
1	779	96	96	185	23	41
1	806	102	98	196	21	64
1	845	97	105	199	19	6
1	845	105	94	195	21	49
1	888	95	103	183	19	51
1	946	91	103	213	14	15
1	946	99	90	201	23	20
1	971	95	97	217	23	74
1	980	99	97	196	22	7
1	989	99	94	201	18	71
1	1005	101	94	210	16	67
1	1019	96	104	179	20	57
1	1044	106	99	185	21	40
1	1058	100	96	195	21	18
1	1073	100	101	194	17	87
1	1081	93	105	198	22	3

Fig. 4. Result of Basic Skyline Operator

Figure 3 shows all records of diagnose 1. For example, the third record (806, 102, 98, 196, and 21) which isn't a k-dominant point is chosen for analysis. It isn't difficult to find that the point is 4-dominant by (779, 96, 96, 185, and 23) which is a k-dominant point. In 4 dimensions: medicine fee, antibiotic fee, hormone fee and examination fee, the values of chosen point are all worse than (779, 96, 96, 185, and 23), except in hospital days dimension, the value of chosen point is better. So that according to the definition of k-dominant Skyline Operator, the chosen point won't be recommended as a result point and the result of algorithm proves this. If we compared other points out of result with result points, it's easy to find for each point out of result; there is at least one result point 4-dominant or 5-dominant the point, which proves the correctness of One-Scan algorithm.

3) Comparison of Basic Skyline Operator and K-dominant Skyline Operator:

In order to prove the advances of k-dominant Skyline Operator, text data is used to do comparison between Basic Skyline Operator and k-dominant Skyline Operator. From Figure 4, there are 16 result points for Basic Skyline Operator. Meanwhile, the number of result for k-dominant skyline operator is only 5. It's obvious to find that the result set of basic Skyline Operator contains the result of One-Scan Algorithm, less and better points recommended with thinking of k-dominant Skyline. So that it is suitable to use k-dominant Skyline Operator to process the problem in this paper.

D. Finding best treatment cases and overtreatment cases

When the order relationship is $<$ as smaller value is better, best treatment cases can be found since smaller values presents low fee and less hospital days. An example of best treatment

cases is as Figure 2. In contrast, when the order relationship changes to $>$, overtreatment cases can be found.

E. Effectiveness Analysis

In contrast, when the order relationship changes to $>$, overtreatment cases can be found. To prove effectiveness of One-Scan Algorithm, we set up test data of 3 diagnoses. For each diagnose analysis and comparisons are done. Test result is as follows:

TABLE I. RUNNING TIME

Records	Running Time					
	Diagnose 1		Diagnose 2		Diagnose 3	
	One-Scan	BNL	One-Scan	BNL	One-Scan	BNL
20000	20.0	9.7	8.0	6.0	35.2	19.7
40000	43.4	20.6	17.7	13.5	64.4	41.5
60000	67.0	31.6	29.9	27.0	99.4	68.8

TABLE II. RESULT QUANTITY

Records	Result Quantity					
	Diagnose 1		Diagnose 2		Diagnose 3	
	One-Scan	BNL	One-Scan	BNL	One-Scan	BNL
20000	16	279	19	276	13	348
40000	19	366	21	358	13	421
60000	19	428	21	432	13	474

PS: BNL (Back-Nested-Loop) [6] is an algorithm of Basic Skyline Operator

From Table 1, we find that in the same condition, running time of One-Scan is more than BNL. It's normal since the nature of One-Scan is screening the result of Basic Skyline Operator. However, even there are 60000 records, running time of One-Scan is no more than 100ms, considering that daily outpatient numbers of a hospital is about 10 thousands, the running time is totally acceptable for practical use.

From Table 2, comparing with BNL, One-Scan screens result quantity to a large extent. When the quantity of records increases, screening effect is more apparent. One-Scan algorithm obtains less and better results and supports more effective information to users.

IV. CONCLUSIONS

This paper applies data analysis technique to find best treatment and overtreatment cases from huge historical medical data for mitigation of medical treatment problems in China. Inpatient medical data multi-dimensions models are built up to lay foundation of multi-dimensions data analysis. The k-dominant Skyline Operator is introduced and One-Scan algorithm is used to process the problem. Afterwards, the correctness and effectiveness of One-Scan algorithm is proved by experiments. Authors and Affiliations

ACKNOWLEDGMENT

This research is supported by the National Natural Science Foundation of China. (Grant.no 71432006)

REFERENCES

- [1] D.L. Li, "Causes and Countermeasures of Overtreatment," *Medicine and Philosophy*, vol. 9, 2003, pp. 5-9.
- [2] H. Yan, "Difficulties in Getting Your Questions to See a Doctor," Shanxi: Financial University of Shanxi, 2012.
- [3] S. Xu, "Solving Medical Treat Problem :Based on Analysis and Countermeasures in Medical Market (In Chinese)," Shanghai: Fudan University, 2013.
- [4] R. Han, "Analysis on Rational Outpatient Drug Use Based on the Technology of Data Mining," *China Digital Medicine*, vol. 2, 2014, pp. 18-20.
- [5] Chee-YongChan, H.V. Jagadish, Kian-LeeTan, Anthony K.H. Tung and Z.J Zhang, "Finding k-Dominant Skylines in High Dimensional Space," *International Conference on Management of Data*, Chicago, Illinois, USA, June 27-29, 2006.
- [6] Y.C. Chen, "Study on Monitoring Indicators for Rational Drug Use in Community Health Centers in one District of Shanghai," Shanghai: School of Public Health of Fudan University, 2011.
- [7] Stephan Börzsönyi, Donald Kossmann and Konrad Stocker, "The Skyline Operator," In:Proc.of the 17th Int'l Conf on Data Engineering, Heidelberg, IEEE Computer Society Press, 2001