Data Cube Based Respiratory Motion Characterization

Jonanthan Klinginsmith, Yuqing Wu, Indiana University, Bloomington
Huanmei Wu, Indiana University – Purdue University, Indianapolis

Purpose: Adequate understanding and precise characteristics of tumor motion is essential for accurate radiation dose delivery in real-time image-guided radiation treatment. We propose a data cube based approach for representing, analyzing and characterizing tumor motion information at various concept levels. More precise patient clustering and characterization can be achieved by combining the tumor motion information and the patient biomedical information.

Method and Materials: Based on a finite state model, a breathing cycle of tumor motion is represented by three line-segments: exhale (EX), end-of-exhale (EOE) and inhale (IN). A *data cube* is constructed to provide the means of multi-dimensional data analysis that computes summary information based on arbitrary combination of dimensions. Data navigation methods – *roll-up, drill-down, slice-and-dice* – are applied to assist in interactively selecting points of interest and navigating among the concept hierarchy. Patient clusters and motion patterns are detected with the assistance of the visualization tools and advanced data mining techniques, such as K-means clustering algorithm. In addition, we incorporate the patient biomedical information in this process to obtain more precise patient motion classification and characterization.

Results: Experiments have been performed on real patient data. Using our approach, summary information is generated automatically and visually displayed. Patient clustering can be easily detected manually or by clustering algorithms. Interesting motion patterns have been discovered from out preliminary results. More precise patient clustering and motion characterization are accomplished when the approach is used on the combination of motion information and patient biomedical information.

Conclusion: The data cube based tumor motion characterization approach not only eliminates all the manual processing on patient motion characterization, but also provides the facility to refine the patient clustering and detect interesting patterns, which will provide valuable input for better understanding of tumor motion and for effective real-time image guided radiation treatment.

1 Introduction

Tumor motion induced by respiratory motion is a big concern for effective image guide radiation treatment. Several respiration strategies have been developed to compensate for tumor motion. However, all of the advanced treatment methods require adequate characterization of tumor motion. We hereby propose a data cube based approach to characterize tumor respiratory motion based on a finite state model. A data cube allows information to be analyzed based on an assortment of attributes which can be summarized at different concept levels. This approach enables effective, efficient, and interactive motion characterization along with motion based patient clustering, which helps obtain more accurate respiratory tumor motion predication. In addition, the data cube based characterization approach can work with heterogeneous biomedical information to further refine patient clustering and to discovery correlations between dynamic tumor respiratory motion and patient biomedical data.

2 Methods

2.1 Introduction of the Finite State Model

Our proposed data cube approach utilizes the piecewise linear representation of modeling results based on a finite state model for tumor respiratory motion as shown in Figure 1. In this model, a single regular breathing cycle has exactly three breathing states, i.e., exhale (EX), end-of-exhale (EOE) and inhale (IN). Regular motion will periodically proceed from state to state in a fixed order: ••• ➔ EX ➔ EOE ➔ IN ➔ EX ➔ •••

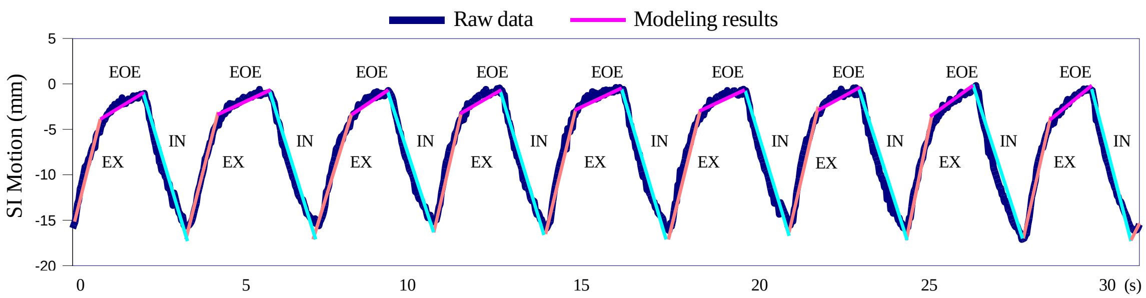

Figure 1: The piecewise linear representation of tumor respiratory motion.

2.2 Multidimensional Analysis

Even though data stored in a relational database is usually represented in the form of two-dimensional tables, a multi-dimensional model is frequently adopted for the purpose of data analysis and knowledge discovery. A fact table is used for data representation, with dimensional tables keeping track of information that describes the dimensions. Each attribute that describes the data (in the fact table) is treated as a dimension that is independent from other dimensions. This model allows the summary information to be computed and analysis to be done on any combination of dimensions. The fact table, as shown in Table 1, is constructed using the data produced by applying the finite state model to the raw tumor motion data collected from patients during radiation treatment.

Patient	Segment	State	Time	SI Motion
1	1	IN	1.60	-23.98
1	1	EX	1.83	22.61
1	1	EOE	2.43	1.37
…	…	…	…	…
1	2	IN	1.63	-16.59
1	2	EX	0.83	16.12
…	…	…	…	…
2	1	IN	1.23	0.00
…	…	…	…	…

Table 1. Example Tumor Motion Data

Multiple aggregate functions are used in computing summary information on arbitrary combination of dimensions. Using just the average SI motion as example along with a fact table based on the data shown in Table 1, we may compute the average SI motion in the following ways: per state per patient per segment, per state per patient among all segments, and per state among all patients. Such summary information can be represented in a table as shown in Table 2.

State	Patient	Segment	Avg. SI motion per State per Patient per Segment	Avg. SI motion per State per Patient	Avg. SI motion per State
EOE	1	1	1.88		
		2	1.62		
		3	2.16		
		…	…		
				1.74	
	2	1	2.92		
		2	3.78		
		3	2.26		
		…	…		
				3.17	
					1.31

Table 2. Summary Info of Patient Tumor Motion on various Concept Levels

2.3 Navigating among the Summary Information

By organizing the computational results in the form of a concept hierarchy, as shown in Table 2, we provide not only the summary information among combination of dimensions, but also the means to navigate among such information. The two major navigation methods available with this data organization approach are *roll-up* and *drill-down*. The *roll-up* operation performs aggregation on the data by climbing up the concept hierarchy for a particular dimension. For example, a roll-up detailed in Table 2 takes the average SI motion per patient per state, to the average SI motion per state. The roll-up operation moves from detailed data to more summarized data. The *drill-down* operation is exactly the reverse of roll-up operation. It navigates from more summarized data to more detailed data, by moving down along the concept hierarchy. For example, from average SI motion per state to average SI motion per state per patient.

These navigation operations enable the user to interactively query and analyze data at hand, and to locate the analytical data on interesting concept levels. However, there are multiple shortcomings of the method as shown in Table 2. First, the data structure is non-relational, with lots of holes (NULL values) in it, even on the primary key columns. Second, it is not convenient; since it fixes the path among the concept hierarchy along which users can roll-up and drill-down. For example, one may not be able to drill-down to average SI motion per state per segment from average SI motion per state, even though it may be very interesting to obtain such information. Finally, the number of columns grows exponentially in relation to the number of aggregated attributes, putting extra burden on the storage system.

2.4 Data Cube

A solution that fixes the problems listed above is a *data cube*. In a data cube, the data in a fact table that has *n* attributes is organized in the form of an *n-dimensional* cube. The cell $(i_1, i_2, \ldots, i_n)$ contains the measure in the fact table where the dimensions takes values $i_1, i_2, \ldots, i_n$ respectively. Figure 2 displays a data cube that contains the information shown in Table 1, focusing on the theme of SI motion. With the data cube, the storage is compact, summary information can be computed on arbitrary combination of dimensions and the navigation (roll-up and drill-down) can be performed along any path in the concept hierarchy.

Another data cube operation we find very useful and plan to adopt in the respiratory motion characterization is *slice and dice*. The *slice and dic*e operation performs a selection on one or more dimensions of the data cube and results in a sub-cube. For example, using slice and dice operations, we can select patients whose SI motion is above a certain threshold, and perform further analysis on the resulting sub-cube.

2.5 Visualization and Motion Clustering

A data cube provides the means to organize, select, and compute summary information based on the data in the cube cells. To assist respiratory motion characterization, we need to detect patterns among individual patients and groups of patients. Some patterns can be detected directly with the assistance of visualization techniques. As shown in the resulting section (Figure 4 and 6), the clustering of patients who breathe with similar

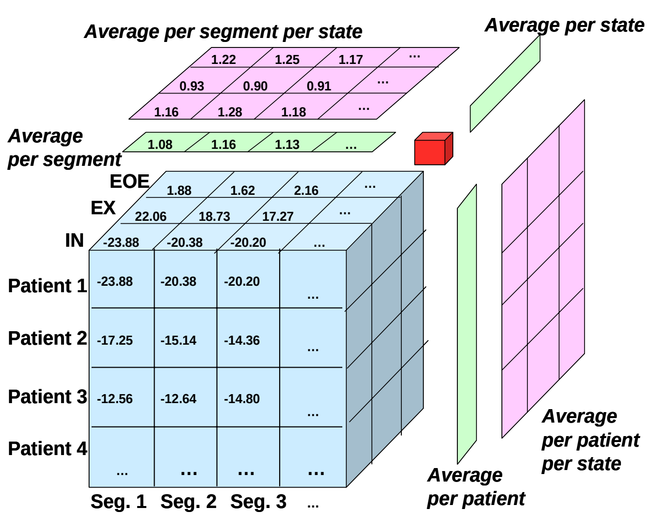

Figure 2. Data Cube on the Theme of SI motion

patterns can be detected from the image generated. However, this naïve approach will cease to work as the number of data points (for example, the number of patients) increase, or when the patterns are much more complicated. Clustering techniques, such as the K-means clustering algorithm, are used to cluster patients who display similar patterns.

2.6 Incorporating Patient Biomedical Information.

Under most circumstances motion data is rich enough to support clustering and characterization; unfortunately, it is not always the case. For example, patients receiving their first treatment will not have motion data available from prior treatments. Additionally, the motion data may be too coarse due to low sampling rate. To further improve the accuracy of motion characterization, we incorporate the patient's biomedical information into the process.

Biomedical information itself can be managed using a data cube and initial patient clustering is computed based only on the patient biomedical information. With the summary information and patient clustering obtained based on motion information alone, multiple approaches can be taken to incorporate the biomedical information with the motion information. One is to construct data warehouse combining the biomedical data and the motion data. The data warehouse can then be used to conduct data mining and knowledge discovery. Another is to take the patient clustering information obtained from the biomedical data into the analysis of the motion data (as discussed in sections above), or vise versa. Yet another is to perform association rule analysis based on the patient clustering obtained from both biomedical data and motion data. We expect that these approaches will yield similar yet slightly different results. We will further investigate the differences and propose the best approach in knowledge discovery in the presence of both patient biomedical data and motion data.

3. Results

We have applied the data cube-based tumor motion characterization approach to a set of tumor motion streams, which contains 17,000 data points, of 12 patient, each patient has 10-20 treatment sessions, and each session contains 100-200 data points. Experiments on another 31 patient data are will be performed soon. Using the finite state model, a treatment session is represented by a sequence of breathing cycles, each with three line segments: EX, EOE and IN. The time and SI motion are measured for each such segment and the result is stored in a table as shown in Table 1.

We construct data cube based on this segmenting representation. Using the navigation methods associated with the data cube, we retrieve and visualize summary information on different concept levels. Figure 3 illustrates a sequence of such images generated by a roll-up operation. With a relatively small data set (12 patient as shown in the figure), the breath pattern for each patient is clearly visible even on the detailed image that portraits the SI motion for each segment of each patient. For larger data set, image based on coarser concept level, such as average SI motion per patient, would turn out to be more appropriate.

We conduct K-means clustering algorithm (with k=3) to cluster the patients based on their average SI motion. The resulting clustering, as shown in Figure 4, indicates that patients that fall in the same cluster (represented by a circle) have a similar breath pattern, based on the measure of average SI motion.

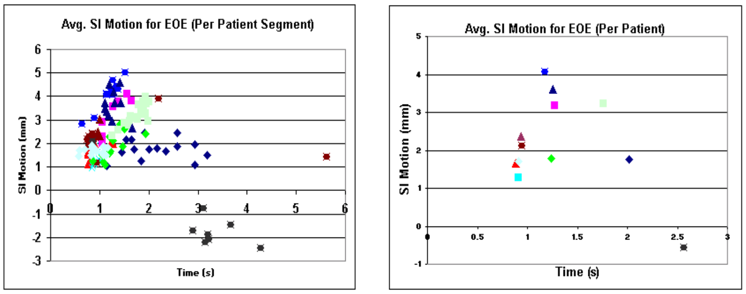

Figure 3. Roll-up from *average SI motion per patient per state* to *average SI motion per patient*

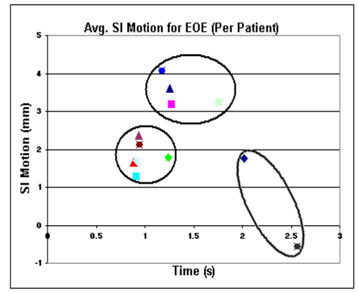

Figure 4. Patient clustering based on average SI motion

We also use the slice and dice operation to select a sub-cube and perform the same roll up operation on the summary information in the sub-cube. For example, Figure 5 shows a set of images generated by the same roll-up operation used in generating Figure 3, but only for patients younger than 70. Comparing Figure 3 and 5, one can identify the substantial differences among the breathing patterns between younger and older patients. Performing K-means clustering on the later result, new patient clustering is generated for younger patients, as shown in Figure 6.

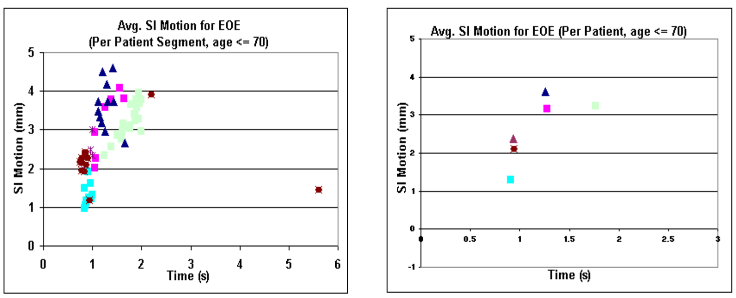

Figure 5. Roll-up from *average SI motion per patient per state* to *average SI motion per patient*, for patients younger than 70.

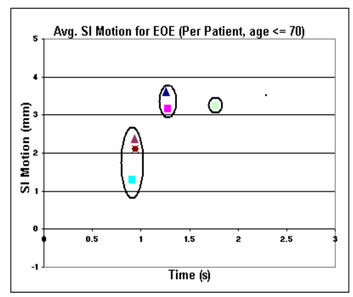

Figure 6. Patient (younger than 70) clustering based on average SI motion