Clustering White Matter Fibers: Analysis of Algorithms and Dissimilarity Measures

Yuen Hsi Chang*
CS 318 Winter 2014
*Carleton College: Department of Computer Science

***Abstract*—In this paper, we evaluate the performance of six combinations of dissimilarity measures and clustering methods when classifying nerve fibers in the brain using tractography data generated from diffusion weighted nMR images. We compared the results of both agglomerative and spectral clustering to an expert human clustering of the fibers using both Rand and Adjusted Rand Index taking the human clustering as ground truth.**

***Index Terms*—agglomerative clustering, complete linkage, diffusion weighted MRI, dissimilarity measure, Rand Index, similarity matrix, single-linkage, spectral clustering.**

I. Introduction

Diffusion Weighted MRI can be used to measure the diffusion of water in the brain, which is limited by membrane structures within tissues. In brain regions where axons are similarly orientedand tightly packed, water diffusion is anisotropic, and by measuring water diffusion in various directions, it is possible to determine nerve fiber orientation. Further processing can reconstruct digital fiber tracts by estimating the shape of microstructures.

Since full tractography data sets consist of hundreds of thousands of curves, it is difficult to manually analyze unprocessed tractography data.

Clustering algorithms group fibers into bundles based on their similarity, with the hope that these bundles correspond to known macro-scale anatomical structures within the brain. Such computational approaches allow researchers and doctors to reconstruct white matter structures without having to manually analyze tractography data.

Results from these computer models do not always correspond to actual structures defined by medical professionals, and there remains a gap between artificially-generated clusters and realistic anatomical structures. In this paper, we will evaluate two clustering algorithms, agglomerative and spectral clustering, by comparing our bundles to the clusters presented by Jadrian Miles, assumed to represent the ground truth. The difference between our clustering algorithms and the ground truth will be calculated using variations of the Rand Index.

II. Theory

Clustering a set of curves consists of two computational steps. First, one generates a matrix that indicates the dissimilarity between each pair of curves. Second, one groups the set of curves based on their dissimilarity measures. The resulting classifications generated by these algorithms are quantitatively compared to a ground truth clustering using cluster similarity measures.

A. Dissimilarity Measures

We used three dissimilarity measures: closest point distance, mean minimum distance, and Hausdorff distance.

$$d_c(F_1, F_2) = min_{p\in F_1, q\in F_2}(|p-q|), \tag{1}$$

$$d_m(F_1, F_2) = (\sum_{p\in F_1} min_{q\in F_2}(|p-q|))/l(F_1), \tag{2}$$

$$d_h(F_1, F_2) = max_{p\in F_1}(min_{q\in F_2}(|p-q|)), \tag{3}$$

where F_1 and F_2 are input curves, p and q are points, and $l(F_1)$ refers to the length of curve F_1.

The closest point distance, $d_c(F_1, F_2)$, is the minimum path length that connects a point in F_1 to any point in F_2. The mean minimum distance, $d_m(F_1, F_2)$, is the average distance of all paths connecting a point in F_1 to the closest point in F_2. The Hausdorff distance, $d_h(F_1, F_2)$, is the maximum of the set of minimum paths from F_1 to F_2.

Once the measures are calculated, dissimilarities between pairs of curves are stored in a matrix in which the i, j^{th} entry contains the dissimilarity between F_i and F_j. All distances are symmetric and satisfy the property $d(F_k, F_k) = 0$ meaning that the resulting dissimilarity matrix will be symmetric and hollow. This symmetric and hollow matrix is the input for the agglomerative and spectral clustering algorithms, both of which we implemented.

B. Agglomerative Clusterings

Agglomerative clustering is an iterative clustering method that repeatedly combines the two most similar clusters until all fibers are clustered. When two clusters are combined, the dissimilarity values of the new cluster will be calculated in one of two ways. Single-linkage agglomerative clustering takes the minimum dissimilarity of the two combined clusters. Complete-linkage agglomerative clustering takes the maximum dissimilarity of the two combined clusters.

Single-linkage clustering creates large clusters, because when two clusters are combined, the dissimilarity decreases. Complete-linkage agglomerative clustering creates small clusters, because when two clusters are combined, the dissimilarity increases.

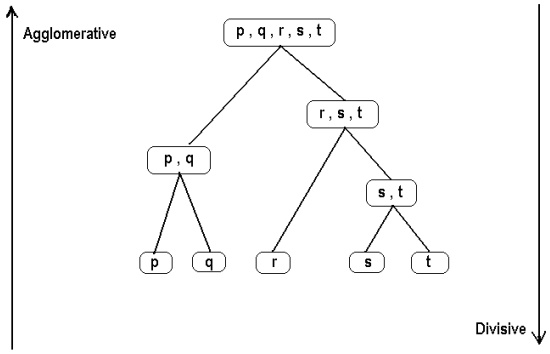

Fig. 1. This figure illustrates the process of agglomerative clustering. When two clusters are combined, a new parent node is created. The data of the new node contains the data of both of its children[3].

Given a similarity matrix for a set of curves, a single-linkage agglomerative clustering set can be obtained by the following steps:

1) Determine the most similar pair of clusters by locating the minimum non-zero entry in the matrix. Let x index the first cluster and y correspond to the second cluster.
2) Replace the ith entry in the xth column with the minimum of the ith entry in the xth and yth columns. Replace each entry in the yth column with 0.
3) Replace the xth entry of each row with the minimum of the xth entry of that row and the yth entry of that row. Replace the yth entry of each row with 0.
4) Repeat Steps 1-3 until the matrix has only 1 non-zero row and column remaining.

Complete-linkage agglomerative clustering follows the same steps, but takes the maximum instead the minimum in steps 2 and 3. The set of clusters at each stage of the algorithm is stored in a binary tree. Initially, the tree has a leaf corresponding to each curve. When two curves or clusters are combined, they become the children of a new node. After the algorithm finishes, we end up with a tree whose root contains an array of the indices of each curve, and the kth level of the tree is a clustering with k clusters.

C. Spectral Clustering

Spectral clustering implements a K-means algorithm to assign curves to groups, but it does so using a $k \times n$ matrix that is derived from the similarity matrix. Spectral clustering makes use of perturbation theory, which refers to the method of approximating the solution of a problem by finding the exact solution to a related problem. Matrix perturbation theory is applied where a transformation of the given similarity matrix is treated as the related problem, whose solution approximates the groupings of the curves within our initial problem.

Specifically, spectral clustering seeks to determine the sensitivity of the eigenvectors within our distance matrix with respect to changes in the system. This is done by finding the k largest eigenvectors of an $n \times n$ matrix derived from the distance between curves, and reducing a $n \times k$ matrix that stores this information to a square $k \times k$ matrix. Each row in the final square matrix corresponds to a group of curves, so this algorithm leaves us with k clusters. The relationship between the $n \times k$ matrix and our initial matrix containing distance between points touches upon math theorems that lie beyond the scope of this paper, but it should be understood that the eigengap (the difference between two successive eigenvalues) of the $n \times k$ matrix reflects the cohesiveness of given individual points. For a more complete understanding on spectral clustering, refer to Ng et al[6].

A description of the clustering algorithm is shown below. For the complete algorithm, refer to Ng et al[6]. Note that the normalized $n \times k$ matrix forms a unit sphere in k-space, and reducing this to a square matrix requires all clusters to lie orthogonal to one another, such that clusters are created based on the differences in orientation rather than the length of respective vectors. Figure 2 shows a visualization of what we would obtain if we graph the points presented by the orthonormal $n \times 3$ matrix Y. The points are color coded to demonstrate which clustering groups they fall under.

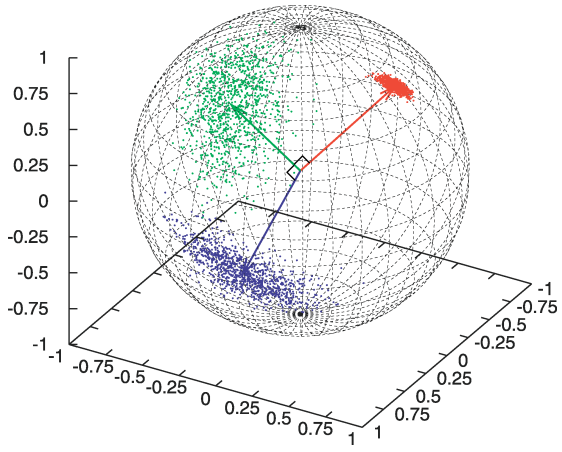

Fig. 2. This figure provides a visualization of the process of grouping normalized curves within a unit sphere, wherein orthonormal groups correspond to the number of clusters that are formed. In this case, we generate three clusters[4].

Given a set of fibers $F = \{F_1, ..., F_n\}$ in $\mathbb{R}^l$ that we want to cluster into k subsets:

1) Form the affinity matrix $A \in \mathbb{R}^{n*n}$ defined by $A_{ij} = e^{(-||s_i - s_j||^2)/2\sigma^2}$ if $i \neq j$, and $A_{ii} = 0$, where $|s_i - s_j|$ is the disimilarity measure between the ith and jth fiber.
2) Find L, the laplacian of A, and form the matrix $X = [x_1 x_2 ... x_k] \in \mathbb{R}^{nxk}$ by stacking the k largest orthogonal eigenvectors of L in columns.
3) From the matrix Y from X by renormalizing each of X's rows to have unit length.
4) Finally, treating each row of Y as a point in $\mathbb{R}^k$, cluster them into k clusters via K-means.
5) Assign the original fiber F_i to cluster j if and only if row i of the matrix Y was assigned to cluster j.

The scaling parameter σ presented in step 1 of the algorithm is a user specified parameter, and should be chosen to restrain

the affinity matrix within a threshold. This way, we end up with tight points on the unit sphere after we normalize the matrix, such that the clusters we generate are clearly divided and independent of one another [6]. A more detailed description regarding how we choose σ is presented in the Methods section.

D. Evaluation of Clusterings

Two clusterings of a set of curves are equivalent to two partitions of the same set. Each pair of entries in the set can be categorized depending on whether that pair is clustered together by the first clustering or second clustering. If a pair of curves is clustered together by both clusterings or neither clustering, then the clusterings agree on that pair of curves. If a pair of curves is clustered together by only one of the clusterings, then the two clusterings disagree. Iterating through each pair of curves, an agreement matrix can constructed as shown in Table 1.

	In C2?	Not In C2?
In C1?	a	b
Not In C1?	c	d

Table 1. Each pair of curves in the data set is put into one of four categories depending on whether they are clustered by C1 and whether they are clustered together in C2.

The Rand Index is calculated by summing the diagonals and then dividing by the total number of curve pairs:

$$R = \frac{a+d}{a+b+c+d}. \tag{4}$$

For identical clusterings, every pair of curves is clustered together by both clusterings or neither clustering and the Rand Index is 1. For dissimilar clusterings, some pairs of curves will be clustered together by one clustering but not the other. As a result, $a+d$ will be less than M and the Rand Index will be less than 1. The Adjusted Rand Index takes chance agreement between pairs of curves into account. The general form of a chance corrected value is given by:

$$S = \frac{S - E(S)}{Max(S) - E(S)}, \tag{5}$$

where $E(S)$ is the expected value of S. For the Rand Index, $Max(R) = 1$ and

$$E(R) = \frac{(a+b)(a+c)+(d+b)(d+c)}{(a+b+c+d)^2} \tag{6}$$

The Adjusted Rand Index for two identical clusterings is 1. For dissimilar clusterings, the Adjusted Rand Index is significantly less than than the Rand Index. The minimum of the Adjusted Rand Index is zero.

These evaluation methods assume that the two input clusterings are partitions of the same set. Adapting these methods to accommodate partitions of different sets is a challenging problem. We did not have time to implement evaluation methods that account for differences in the set of curves.

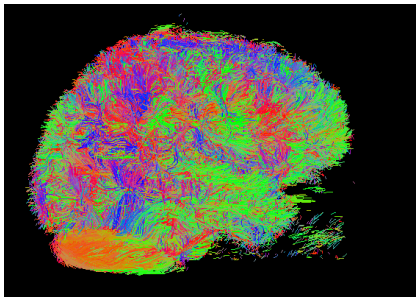

Fig. 3. A sample tractography dataset generated by TrackVis. Color represents the intensity of a measure taken from the centermost point of each fiber, and because of this, coloring at the edges of a fiber does not represent local qualitites. Color indicates fiber orientation at the centermost point. Few fibers have been culled.

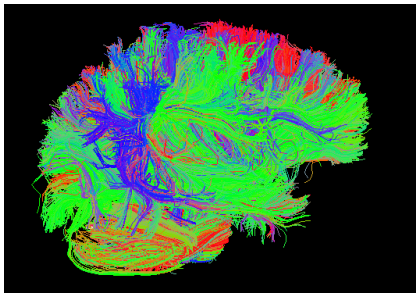

Fig. 4. A TrackVis representation of the complete set of fibers used in this study. This set of fibers is a subset of those shown in Figure 2. Short ($< 50mm$) and long ($> 200mm$) fibers have been culled leaving approximately 28,000 long fibers. The ground truth clustering was performed on this set of fibers. Computational clusterings were performed on random subsets of this set.

III. Methods

We used a dataset containing 64 scans of a healthy human brain, all of the scans are weighted at b-values of 1000 s/mm^2 and are taken from different orientations. The tractography information is acquired by applying the second order Runge-Kutta method on a diffusion tensor imaging model generated by Diffusion Toolkit[2].

Fibers within the brain are interpreted as an extensive set of curves formed by the tractography dataset as shown in

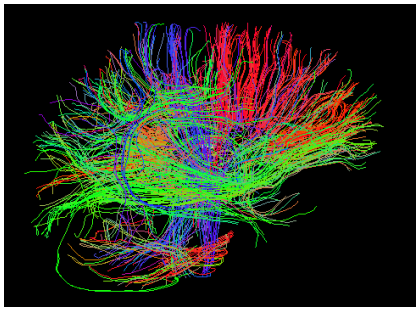

Fig. 5. A TrackVis representation of a random subset of fibers. This set of fibers is a subset of those shown in Figure 3. 500 fibers were randomly selected, which was enough to indicate brain structure. Our clustering algorithms were able to cluster sets of 500 fibers in 5 minutes.

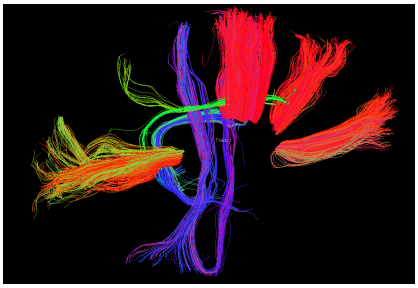

Fig. 6. A TrackVis representation of the fibers in the ground truth clustering. This set of curves is a subset of the curves shown in Figure 3. This set contains about 2000 classified curves.

Figure 3. We culled curves below a certain length threshold and above another length threshold as shown in Figure 4. This way, curves caused by erroneous scans or noise do not get included in the set we cluster. In our case, we set the lower bound of the length of our fibers as x mm, and our upper bound as x mm.

We took in random subsets of approximately 500 fibers each from our set of post processed curves, and generated similarity matrices based on a specified dissimilarity measure. TrackVis visualization of a random subset is shown in Figure 5. These similarity matrices were passed in to our implementation of each clustering algorithms as input, each of which generates an output array that specifies the groups different fibers belong in.

Unlike agglomerative clustering, Spectral clustering on the other hand requires us to specify the number of clusters to be generated. We selected this number by picking it to match the number of clusters the ground truth clustering of each similarity matrix yields. As stated previously, spectral clustering also requires us to iterate through different σ values and select the one that generates the tightest clusters. We did this by calculating the standard deviation of the affinity matrix we create, and selecting the σ value that correspond with the affinity matrix of the highest standard deviation.

We then applied a cluster quality measure to quantify the agreement between the ground truth clustering (shown in Figure 6) and the clusterings generated by our algorithms. We used the Rand Index and the Adjusted Rand Index, which measure the agreement between two partitions of a single set. This process was repeated for each combination of clustering algorithm, dissimilarity measure and evaluation method on 50 subsets, each containing 500 curves.

IV. Results

No significant data has been collected at this point, but we expect to observe the following. The average Rand and Adjusted Rand Indices for each combination of dissimilarity measure and clustering algorithm are shown in Table 2. Of the three algorithms we implemented, spectral clustering had much higher Rand and Adjusted Rand values than both versions of agglomerative clustering. Single linkage agglomerative clustering performed slightly better than complete linkage overall. Additionally, the mean minimum distance returned higher Rand and Adjusted Rand values than the other two similarity measures.

For single linkage agglomerative clustering, using closest point distance resulted in higher Rand and Adjusted Rand values than the Hausdorff distance. The opposite was true for complete linkage clustering.

The similarity of our computational clusterings to the ground truth clustering did not seem to depend on the subset of fibers clustered. However, the variance of the Rand Index was 20% larger than the variance of the Adjusted Rand Index.

V. Analysis

The ultimate goal of fiber clustering of diffusion weighted MRI data is to generate reliable models of brain structures from raw data. However, data collected by MRI machines is noisy and this noise must be accounted for in each step of the process. Unfortunately, our algorithms did not account for noise or uncertainty, and it assumes that the fibers provided by DTK represented ground truth.

Some of the advantages and disadvantages to the different clustering methods we employed are summarized below. Firstly, one disadvantage to single linkage and complete linkage agglomerative clustering is that the similarity values for a cluster depend only on one fiber in that cluster. For single linkage agglomerative clustering, only the closest fiber is considered when determining the similarity values for a new cluster. For complete linkage agglomerative clustering, only the furthest fiber is considered when determining the similarity values for a new cluster.

One of the strengths of spectral clustering is that it does not assume that the density of each cluster is Gaussian, which suits fiber models in the real-world very well. On a related note, studies have shown that even for clusters that do not form convex regions and are not cleanly separated, the algorithm still performs clustering consistent to what a human would have chosen. This is due to effective application of matrix perturbation theory.

There are also several drawbacks to spectral clustering. First, in the study from which our algorithm was derived, the authors did not apply spectral clustering to fibers obtained from scans. Rather, it gives examples of clustering points in k groups without providing the context of the points. In our case, we had to treat fibers as points during the input stage of our similarity matrix, and we are not certain whether this modification of the algorithm would provide unfavorable results. Also, as σ is a human-specified parameter, using spectral clustering in a laboratory setting does require some degree of human intervention and fundamental understanding of the algorithm. Finally, the number of clusters to be generated by the algorithm has to be passed in as input. This is dissimilar to agglomerative clustering, which generates a set of clusterings, each with a different number of clusters. As a result, an agglomerative clustering with k clusters can ideally

TABLE 2

Predicted Results Summary

Clustering Algorithm	**Rand Index**			**Adjusted Rand Index**		
	d_c	d_m	d_h	d_c	d_m	d_h
Agglomerative Clustering (Single Link)	0.54	0.74	0.39	0.32	0.67	0.15
Agglomerative Clustering (Complete Link)	0.35	0.72	0.53	0.21	0.61	0.41
Spectral Clustering	0.63	0.91	0.61	0.50	0.75	0.45

be retrieved in $O(logn)$ time after the clustering algorithm has been run. In general spectral clusering is computationally expensive compared to agglomerative clustering.

VI. Conclusion

Although we only tested our algorithms on the tractography output of a single, healthy human brain, we were able to draw several conclusions regarding the relative merits of our methods. We conjecture that agglomerative clustering performs better when the dissimilarity measure corresponds to the linkage type. However, mean minimum distance yields better clusters than both closest point distance and the Hausdorff distance for both variations of agglomerative clustering. Spectral clustering generates better clusters than agglomerative clustering regardless of the distance metric used. The optimal similarity measure/clustering algorithm combination is spectral clustering with mean minimum distance. While none of these methods are perfect, we have shown that spectral clustering is strictly better than agglomerative clustering in this context.

VII. Future work

Our project left a number of questions unanswered. In particular, we understand that we have only implemented two of many different clustering algorithms. Given more time, we are interested in incorporating and evaluating other clustering algorithms to our project. Different clustering techniques we would like to explore involve using normalized cuts to devise fiber groups, looking at shared-nearest neighbor to calculating similarity between fibers, and further applying spectral theorem with methods involving singular value decomposition.

We would also like to devise a more sophisticated means to assigning similarity to curves. Our current method of basing similarity upon distance between curves is suboptimal. Fibers that are shown to be similar based on Hausdorff distance may not remain similar under the closest point distance or the mean minimum distance, and metrics that work well for a particular structure may not be equally effective for a different set of curves. For instance, the closest point distance measurement is a relatively accurate similarity measure for parallel fibers, but not for kissing, crossing, or branching fibers. It is possible to further our project by creating an algorithm that select a distance metric that is appropriate to the approximate structure of curves.

Also, in order to determine the agreement between the clusters generated by our algorithms and the ground truth partitioning we were provided with, we performed a measure using the Rand Index and the Adjusted Rand Index. Even though Moberts et al. stated that the Adjusted Rand Index is the measure of choice for cluster validation [5], there are a number of other clustering quality measures, some of which we would implement given more time.

Finally, we made the decision to cull curves that fall beneath a certain threshold, such that short curves that are relatively insignificant wont impact our results. However, we did not further investigate the direct impact these short curves can have upon the validity of our results. Again, given more time, we could have explored the consequences this decision has had on our results.

References

[1] Corouge, I., Gouttard, S., and Gerig, G. (2004, April). "Towards a shape model of white matter fiber bundles using diffusion tensor MRI." In Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium

[2] "Diffusion Tookit". trackvis.org/blog/tag/diffusion-toolkit. Web. 13 March 2014.

[3] "Hierarchical Clustering - Intro." http://www.solver.com/hierarchical-clustering-intro. Web. 1 March 2014

[4] Hamelryck, Thomas, John T. Kent, and Anders Krogh. "Sampling Realistic Protein Conformations Using Local Structural Bias." www.ploscompbiol.org/article/slideshow.action?uri=info:doi/10.1371/journal.pcbi.0020131&imageURI=info:doi/10.1371/journal.pcbi.0020131.g004. Web. 14 Mar. 2014.

[5] Moberts, B., Vilanova, A., and van Wijk, J. J. (2005, October). "Evaluation of fiber clustering methods for diffusion tensor imaging." In Visualization, 2005. VIS 05. IEEE (pp. 65-72). IEEE.

[6] Ng, A. Y., Jordan, M. I., and Weiss, Y. (2001). "On Spectral Clustering1 Analysis and an algorithm." Proceedings of Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 14, 849-856.

[7] "TrackVis." trackvis.org. Web. 22 Feb. 2014.

[8] Vilanova, A., Zhang, S., Kindlmann, G., and Laidlaw, D. (2006). "An introduction to visualization of diffusion tensor imaging and its applications. In Visualization and Processing of Tensor Fields" (pp. 121-153). Springer Berlin Heidelberg.

[9] Albatineh, A. N., Niewiadomska-Bugaj, M., and Mihalko, D. (2006). "On Similarity Indices and Correction for Chance Agreement. Journal of Classification", 23(2), 301-313.

[10] Yeung, K. Y., and Ruzzo, W. L. (2001). "Details of the Adjusted Rand Index and Clustering Algorithms, Supplement to the Paper An Empirical Study on Principal Component Analysis for Clustering Gene Expression Data. Bioinformatics, 17(9), 763-774.