Detection of Keratoconus by Semi-Supervised Learning

Keywords: Semi Supervised Learning, Manifold Learning, Multi Dimensional Scaling, IsoMap

Abstract

Keratoconus, is a non-inflammatory disorder of the eye in which structural changes within the cornea cause it to thin and change to a more conical shape which leads to substantial distortion of vision. Current methods for the detection of the disease mainly are Supervised learning approaches. Our goal is to label suspect patients who have not been clinically diagnosed(unlabeled data). Hence, we propose a new model, which integrates Manifold learning techniques with Semi-Supervised learning, to maximize the utility of the unlabeled data.

1. Introduction

Keratoconus is a degenerative noninflammatory corneal thinning disorder characterized by anterior protrusion of the cornea and its irregular shape. This causes substantial distortion of vision with multiple images, irregular astigmatism and sensitivity to light. Because of this irregular shape, glasses do not adequately correct the vision in the patients since they cannot conform to the shape of the eye. Detection of Keratoconus is important to avoid unpredictable results in refractive surgery. This surgery on a keratoconic eye might lead to even more thinning of the cornea.

Current methods for the detection of Keratoconus mainly consist of building Artificial Neural Networks(Carvalho, 1996; Maeda, 1995) and Decision trees(Twa, 2005) (supervised learning), using the asymmetry of the corneal topography of the labeled data alone. We propose a new model, which integrates Manifold learning techniques with Semi-Supervised learning, to maximize the utility of unlabeled data.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

1.1. Objective

Our aim is to build a model which recognizes patients prone to Keratoconus and classify them based on the advancement of the disease. This involves analyzing differential asymmetry of the corneal elevation, thickness, aberrations and classifying the patients into the 4 major stages(categories), which are:

- Normal
- Mild
- Moderate
- Advanced

Although, Advanced keratoconus is easily diagnosed by slit lamp findings and keratometry readings, it is difficult to detect early keratoconus with these instruments. There exists another class, keratoconic suspects (or just Suspects), which are the corneas with a videokeratographic pattern of localized steepening but none of the traditional clinical signs of keratoconus nor any other circumstances that might explain the topography pattern.

Keratoconus suspects may also be considered to be preclinical patterns of keratoconus when the disease is clinically diagnosed in the fellow eye(other eye). Nevertheless, fellow eye observations are not reliable for ruling out keratoconus in the suspect eye, because the suspect eye may show the first indication of the disease in the patient.

Hence this category can be divided as:

- Initial Suspects : These are clinically normal eyes, but were detected or suspected to be keratoconic by a Neural Network based classifier which takes the corneal topography parameters as the input.
- Fellow normal eye : It is the clinically and topographically normal eye of a patient who has got evident keratoconus in the other eye.
- Abnormal : These are also clinically normal, Keratoconus is not detected, but the pattern is similar. This classification is done by the clinical observer.

Thus the Suspects class data consists of the unlabeled data, that is the data which is not given any of the four major class labels. We have to appropriately classify this data into the Major classes, mainly as Mild, Normal or a third class called Fuzzy. The Fuzzy class consists of the suspect eyes which may not have Keratoconus currently, but might develop later. Thus the data we have consists of 5 categories: Normal, Mild, Moderate, Advanced, Suspects. So we have a set of labeled points which belong to the four major groups and set of unlabeled points (Suspects).

Thus we are looking to:

- Use the suspects data to improve the performance of our classifier.
- Generate severity labels for the suspects that will be verified later on follow up.
- Find the correlation between various measures suggested in the literature (refer section 2) and the class labels.

2. Corneal Topography Analysis

Keratoconus is a disease starting in the inferotemporal part of cornea (bottom part of cornea away from the nose) mostly. So our aim is to bring out the difference between that sector and the other sectors such that asymmetry is established and it can be used for the classification. The differential assymetry of the corneal topography is considered for analysis. The different kinds of topographical data which can be used are : Elevation, Curvature, Corneal Thickness and Aberration. We characterize corneal topography with quantitative corneal indexes(Maeda, 1994; Dingeldein, 1989; Wilson, 1991) such as:

- Opposite sector Index (OSI) : Maximum difference between the opposite sectors.
- Differential sector Index (DSI) : Maximum difference between any two sectors
- Center Surround Index (CSI) : Difference between the central annulus and the annulus surrounding it.
- Successive Sector Index (SSI) : Maximum difference between the successive sectors.
- Irregular Astigmatism Index (IAI) : Average summation of inter-ring power variations along every semi meridian for entire corneal surface.

The corneal elevation and curvature can also be represented as a set polynomials called, Zernike polynomials(Carvalho, 1996) . These are a set of orthogonal polynomials that arise in the expansion of a wavefront function for optical systems with circular pupils. They are used to describe aberrations of the cornea or lens from an ideal spherical shape, which result in refraction errors.

2.1. Related Work

Most of the earlier work in the classification of Keratoconus consisted of building neural network based(Carvalho, 1996; Maeda, 1995) systems. Artificial Neural network models consist of hardware or software that simulates some common aspects of the biological nervous system. Different topography indices, such as the ones escribed above, are used as the input data for the Artificial Neural network models.

Another previous approach for detection of Keratoconus uses decision trees(Twa, 2005). The main idea behind this approach is that ANN models make diagnostic classifications of corneal disease using summary statistics as network inputs, neglect global shape features and produce results that are difficult to interpret clinically. In this approach the use of Zernike polynomials to model the global shape of the cornea and use the polynomial coefficients as features for decision tree classifier is proposed. This model is used to classify a sample of normal patients and patients with corneal distortion caused by keratoconus. The resulting model is easy to interpret using visual cues.

3. Semi Supervised Learning

Many machine-learning researchers have found that unlabeled data, when used in conjunction with a small amount of labeled data, can produce considerable improvement in learning accuracy. One could say that the knowledge on $p(x)$ that one gains through the unlabeled data has to carry information that is useful in the inference of $p(y|x)$.

Thus for Semi-Supervised Learning (SSL) to work, certain assumptions(Chapelle, 2005) will have to hold. One of the assumption states that : The (high-dimensional) data lie (roughly) on a low-dimensional manifold. We are interested in finding this low dimensional manifold which is embedded in the high dimensional input data. This is called Manifold Learning, which is a popular recent approach to nonlinear dimensionality reduction.

The general idea of SSL is to utilize the geometry of the underlying space of possible patterns to con-

struct representations. The simplest example is two disjoint classes on the real line. In this case, given sufficiently many unlabeled points, the structure can be recovered completely with just one labeled point. In general unlabeled data provides us with information about the probability distribution $p(x)$, while the labeled points tell us about the conditional distributions. We see that $p(x)$ puts all its measure on a compact (low dimensional) manifold. Therefore, the unlabeled points are used to estimate the manifold and labeled examples then specify a classifier defined on the manifold.(Belkin, 2004)

Algorithm 1 classical Multidimensional Scaling (cMDS)

Input: $D \epsilon R^{n\times n}$ $(D_{ii} = 0, D_{ij} \geq 0)$, $d \epsilon \{1, \ldots, n\}$
1. Set $B = -\frac{1}{2}HDH$, where $H = I - \frac{1}{n}11^T$ is the centering matrix.
2. Compute the spectral decomposition of $B : B = U\Lambda U^T$.
3. Form Λ_+ by setting $[\Lambda_+]_{ij} = max\{\Lambda_{ij}, 0\}$
4. Set $X = U\Lambda_+^{1/2}$
5. Return $[X]_{n\times d}$

4. Our Model

We are interested in using the entire data(labeled and unlabeled) to find an appropriate low dimensional embedding of the data. Manifold Learning techniques such as Classical Multi Dimensional Scaling (Cox, 2001) and Isometric Feature Mapping(Cayton, 2005) are used to find the low dimensional representations of the data(Embeddings).

Multidimensional scaling : Suppose there are a set of n objects is under consideration and between each pair of objects (r, s) there is a measurement δ_{rs} of the dissimilarity. Multi Dimensional Scaling can be defined as a search for a low-dimensional space, usually Euclidean, in which points in the space represent objects, such that the distances between the points in the space (d_{rs}), match as well as possible with the original dissimilarities(δ_{rs}).

Isometric Feature mapping (IsoMap) : Isomap is perhaps the best known and most applied among the multitude of procedures now available for the problem at hand. It may be viewed as an extension to Multidimensional Scaling (MDS). Isomap consists of two main steps:

- Estimate the geodesic distances(distances along the manifold) between points in the input using shortest-path distances on the data sets k-nearest neighbor graph.
- Use cMDS to find points in low-dimensional Euclidean space whose interpoint distances match the distances found in step 1.

Geodesics are defined to be (locally) the shortest path between points on the space. Thus, the Euclidean distance between nearby points in the highdimensional data space is assumed to be a good approximation to the geodesic distances between these points. For points that are distant in the high-dimensional space, the Euclidean distance between them is not a good estimate of the geodesic distance. The problem is that though the manifold is locally linear (at least approximately), this approximation breaks down as the distance between points increases. Estimating distances between distant points, thus, requires a different technique. To perform this estimation, the Isomap algorithm first constructs a k-nearest neighbor graph that is weighted by the Euclidean distances. Then, the algorithm runs a shortest-path algorithm (Dijkstras or Floyds) and uses its output as the estimates for the remainder of the geodesic distances.

Once these geodesic distances are calculated, the Isomap algorithm finds points whose Euclidean distances equal these geodesic distances. Since the manifold is isometrically embedded, such points exist, and in fact, they are unique up to translation and rotation. As discussed earlier, Multidimensional Scaling is a classical technique that may be used to find such points. The basic problem it solves is this: given a matrix $D \epsilon R^{n\times n}$ of dissimilarities, construct a set of points whose inter-point Euclidean distances match those in D closely. The cMDS algorithm shown above(**Algorithm 1**) returns the low dimensional embedding $[X]_{n\times d}$, where n is the dataset size and d is the new dimension.

4.1. Tests and Results

Initially, the cornea is divided into different number of sectors and the quantitative descriptors are calculated in each case. To check the best number of sectors and the nature of data set, we performed $k-$Means clustering using different number of centroids. We also performed correlation analysis between the calculated indices and the class labels. *Figure 1* shows that for a particular value of number of sectors(15), the indices calculated correlate well with the class labels.

Our analysis can be broadly divided into two parts:

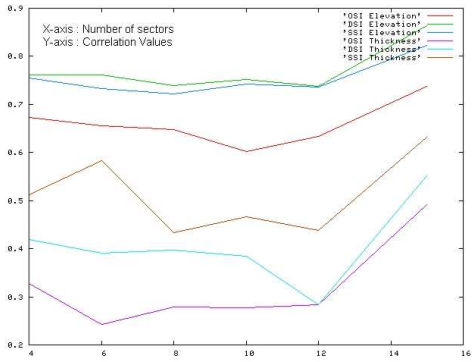

Figure 1. Correlation Analysis

- On the original (high-dimensional) data
- On the embeddings (low-dimensional) obtained using manifold learning

The details about the tests and results are listed below.

An ideal learning machine is sought for this classification problem. We have analyzed the performance of different classifiers and SVMs are chosen because of their nature to maximize margins.

The following tests are performed on both original data and the embeddings obtained.

- Build an SVM classifier and calculate the accuracies.
- Cluster the labeled data and calculate the cluster purity values.

Depending on the accuracies and cluster purity values, choose the appopriate learning machine to classify the Suspects.

The embeddings(low dimensional data) are formed by 3 different methods:

- Isometric Feature Mapping
- Multi Dimensional Scaling
- Principal component analysis

The first two are non-linear dimensionality reduction techniques and PCA uses linear dimensionality reduction. The first two dimensions of all the embeddings are projected and it is observed that for the embeddings obtained by non-linear dimensionality reduction, there is a distinct separation between the classes(*Figures 2 and 3*). As expected, linear dimensionality reduction (PCA) did not give good results as it could not extract the embedding from the high dimensional space.(*Figure 4*)

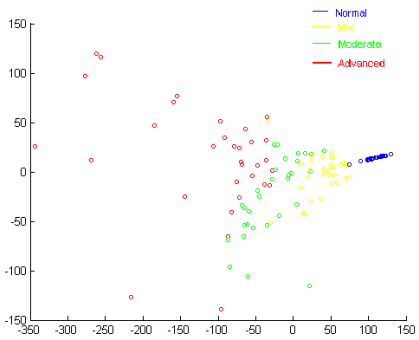

Figure 2. Embedding found by IsoMap : Projection of its first two dimensions

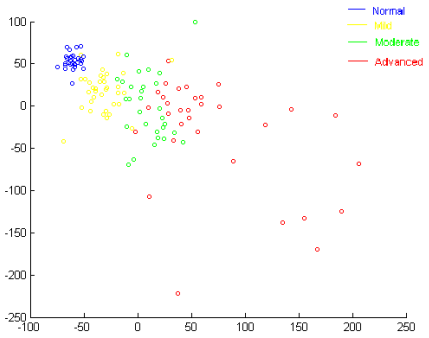

Figure 3. Embedding found by MDS : Projection of its first two dimensions

The manifolds are generated by using : 1. Only labeled data, 2. Both labeled and unlabeled data. The cluster purities of both original data and the embeddings can be seen in *Tables 1 and 2*. *Table 2* also gives a comparison between the embeddings formed by the labeled data and the complete data. It can be seen that manifolds formed by using both labeled and unlabeled data give good results. Also, in general, the

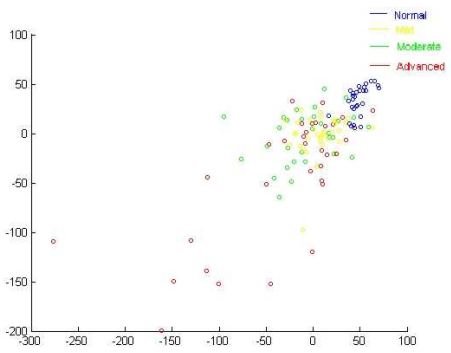

Figure 4. Projection of first two prinicipal components

embeddings give better cluster purities than the original (high dimensional) data.

Similarly, the classification results are given in *Tables 3,4* and *5*. As can be seen the accuracies are better for the embeddings which are formed by the complete data. Thus we were successful in extracting a low-dimensional space that is embedded in the original space by using the unlabeled data also, which was our aim.

Table 1. Results of Cluster Analysis : Original data

Num Sectors	$k=3$	$k=6$	$k=9$
4	0.712	0.716	0.85
6	0.702	0.712	0.82
8	0.712	0.705	**0.86**
12	0.702	0.683	0.702
15	0.743	0.81	0.86

Table 2. Results of Cluster Analysis : Manifold data. Comparison of Cluster purities across different methods, labeled and complete data.

Algorithm	Labeled data	Complete data	k
IsoMap	0.83	0.843	
MDS	0.834	0.84	7
IsoMap	0.87	0.878	
MDS	0.842	0.86	8
IsoMap	0.885	**0.91**	
MDS	0.86	0.887	9

Table 6 gives the comparison between the best results found for both original(high-dimensional) data and the manifold embeddings.

Table 3. Confusion matrix for the classification on the original data

Predicted → Expected	Moderate	Mild	Normal
Moderate	20	4	0
Mild	3	23	1
Normal	0	2	48

Table 4. Confusion matrix for the classification on embeddings formed by both Labeled and Unlabeled data

Predicted → Expected	Moderate	Mild	Normal
Moderate	20	4	0
Mild	1	26	0
Normal	0	0	50

Our main objective is to classify the Suspects (unlabeled data) into one of the major classes. We choose the best learning machine among the obtained (by comparing the classification accuracies or the cluster purity measures) and use it to assign classes(labels) to the Suspects data.

These labels of the suspects will be validated by a follow up on the patients. After an year of follow up on the patient, they will consider his/her state and then compare it with our predicted class. So the learning machine we built needs considerable time to be validated.

But we can partially validate our model by checking if the Initial suspects subclass of Suspects gets classified as Normal or Mild. Given that the Initial suspects are already topographically suspected to have keratoconus(validated by a reliable ANN learning machine), the probability that they being normal should not be too high.

As the Initial Suspects(IS) do not totally resemble the Mild class either, we cannot say that all of the data points in IS class should fall into Mild category. We can say that the probability of they being Mild should be similar to the probability of they being normal. Thus we came up with a metric : If the probability of a suspect being Mild is p and it being Normal is q and if,

- $0 < p \leq 0.2$, it belongs to Normal class.
- $0.2 < p \leq 0.5$ and $p \leq q$, it belongs to Fuzzy class.
- $p > q$, it belongs to Mild class.

Thus using this metric, we got the results as shown

Table 5. Confusion matrix for the classification on embeddings formed by only Labeled data

Predicted → Expected	Moderate	Mild	Normal
Moderate	21	3	0
Mild	3	24	0
Normal	0	1	49

Table 6. Classification accuracies and Cluster purities for Original data and Embeddings.

Method	Original	Embedding
Clustering	86.1%	91.2%
Classification	90.1%	95.04%

in *Table 7.* We can see that, the initial suspects have

Table 7. Classification of Suspects

Predicted → Actual	Mild	Fuzzy	Normal
Abnormals	2	3	15
Fellow Eyes	1	15	4
Initial Suspects	2	16	2

mostly fallen into either Mild or Fuzzy class during the classification. Thus our model does not classify the initial suspects as normal, hence it is valid.

5. Conclusions and Future Work

Different Machine Learning techniques have been applied for the keratoconus problem and empirical results for various methods have been shown. As expected, results seem to suggest better performance on the embeddings generated by Manifold learning.

With accuracies of 95.04% and cluster purities of 91.2%, we can say that our model can be a reasonable predictor of keratoconus. As the results suggest the importance of building a manifold in our problem, further work in this area would lead to better results.

The only limitation of the manifold learning is that we need to obtain the right parameters for number of nearest neighbors and the dimensionality at which an intact manifold exists. This process needs good amount of experimentation.

A perfect medical diagnosis tools require false negative and false positive to be close to zero. As such, the methods attempted in this study cannot claim to provide such a high level of confidence in its results. Still, our model is a reliable prognostic tool to aid ophthalmologists to detect the existence of keratoconus.

References

Maeda N, Klyce SD and Smolek MK, A. (1994). Automated Keratoconus screening with corneal topography analysis. *Investigative Opthalmology Visual Science*

Wilson SE and KJyce SD (1991). Quantitative descriptors of corneal topography: A clinical study. *Arch Opthalmology*

Maeda N, Klyce SD and Smolek MK (1995). Neural network classification of corneal topography *Investigative Opthalmology Visual Science*

Michael K. Smolek and Stephen D. Klyce (1994). Current Keratoconus Detection Methods Compared With a Neural Network Approach *Investigative Opthalmology Visual Science*

Luis Alberto Carvalho (1996). Preliminary Results of Neural Networks and Zernike Polynomials for Classification of Videokeratography Maps. *OPTOMETRY AND VISION SCIENCE, Escola Paulista de Medicina Universidade Federal de Sao Paulo(UNIFESP), Brazil.*

Dingeldein SA, Klycc SD and Wilson SE (1989). Quantitative descriptors of corneal shape derived from computer-assisted analysis of photokeratographs. *Investigative Opthalmology Visual Science*

Michael K. Smolek and Stephen D. Klyce (2005). Algorithms for Manifold Learning.

Mikhail Belkin and Partha Niyogi (2004). Semi-Supervised Learning on Riemannian Manifolds. *Machine Learning*

Olivier Chapelle, Bernhard Schlkopf and Alexander Zien (2005). Semi-Supervised Learning.

Trevor F.Cox and Michael A.A.Cox (2001). MultiDimensional Scaling.

Mikhail Belkin and Partha Niyogi (2004). Manifold Regularization : A Geometric Framework for Learning From Examples. Technical Report, TR-2004-06. *Univ. of Chicago, Department of Computer Science*

Twa MD, Parthasarathy S and Roberts C (2005). Automated Classification of Keratoconus. *Machine Learning*