Quantitative Prediction of Glaucomatous Visual Field Loss from Few Measurements

Zenghan Liang*, Ryota Tomioka‡§, Hiroshi Murata†, Ryo Asaoka† and Kenji Yamanishi*§

*Department of Mathematical Informatics and †Department of Ophthalmology, The University of Tokyo, Japan

‡Toyota Technological Institute at Chicago, Chicago, Illinois, USA; §CREST, JST, Tokyo, Japan

zenghan_liang@mist.i.u-tokyo.ac.jp   tomioka@ttic.edu   hmurata-tky@umin.net

rasaoka-tky@umin.ac.jp   yamanishi@mist.i.u-tokyo.ac.jp

Abstract—We propose database-aware regression methods for extrapolation from few measurements in the context of quantitative prognosis. The idea is to leverage a database of patients with similar conditions to increase the effective number of samples when we train a predictive model. Applying the proposed method to a database of glaucoma patients, we were able to predict the disease condition at a future time point significantly more accurately than the conventional patient-wise linear regression approach. In fact, our prediction was 50% more accurate than the conventional approach when three or less measurements were available and with only two measurements at least as accurate as the conventional approach with six measurements. Moreover, the proposed method can provide spatially localized prediction and also the (localized) speed of progression, which are valuable for doctors in making decisions.

Keywords-quantitative prognosis, multi-task learning, clustering, spatio-temporal data, extrapolation

I. INTRODUCTION

Quantitative prognosis is crucial for the treatment of progressive diseases, such as glaucoma. Since the cost for a measurement (e.g., a doctor visit) can be expensive and time-consuming, it would be valuable if we could precisely predict how fast the disease is progressing from few measurements.

To this end, we propose a class of learning algorithms that leverages the similarity between patients so that we can train a shared predictive model to make a quantitatively accurate prediction for the patient in question with few measurements using data from other patients. We call our approach a *database-aware* approach in this sense.

We apply the proposed modeling framework to the quantitative prediction of visual field loss of glaucoma patients. Glaucoma is a chronic progressive disease in which the visual field is damaged along the retinal nerve fiber bundle [1]. Since it is clinically challenging to restore defects caused by glaucoma, making accurate prediction with limited measurements at an early stage is crucially important. A conventional approach for predicting visual field loss is the *patient-wise* linear regression. Accurate prediction from few measurements is especially challenging for glaucoma, because the reliability of the commonly used visual field test method [2] could be compromised by the so called learning

effect [3], which is a false improvement in the patients' visual field due to adaptation.

The proposed *database-aware* approach goes beyond the conventional *patient-wise* approach in two ways. First, we combine the medical records in the database with the current patient to make accurate prediction at the very early stage of the disease; quantitatively, with only two measurement points, our method is at least as accurate as the conventional patient-wise approach with seven measurements; since the average interval between measurements is 6–12 moths, this could significantly affect the quality of life of a patient. Second, our method predicts not only the global loss of the visual field (known as the mean deviation; MD), but also the local counterpart (known as the total deviation; TD), and the speed of progression; thus it would tell in which area the disease is progressing most rapidly, a more detailed insight that could guide doctors in making decisions. Figure 1 shows two typical outcomes of the prediction of the complete visual field achieved with the proposed method.

A. Related Work

Chan *et al.* [4] compared various traditional machine learning classifiers for supervised classification for determining the presence of glaucomatous optic neuropathy (GON). Unsupervised classification frameworks based on variational mixture of factor analysis (Ghahramani and Beal [5]) and variational independent component analysis (Chan *et al.* [6]) are also proposed, whose discrimination performance were comparable to the result of glaucoma hemifield test (GHT), a commonly used empirical criteria. However the above studies did not deal with quantitative prediction of future disease condition.

Patient-wise linear regression analysis with various criteria has been conducted [7]–[10]. One of the most commonly used progression analysis method in clinical is point-wise linear regression with sensitivity threshold measurements obtained by Humphrey Visual Field Analyzer (HFA); see Section III.

B. Organization

In the next section, we describe the problem and the proposed method. In Section III, we empirically show that the proposed method clearly outperforms the conventional
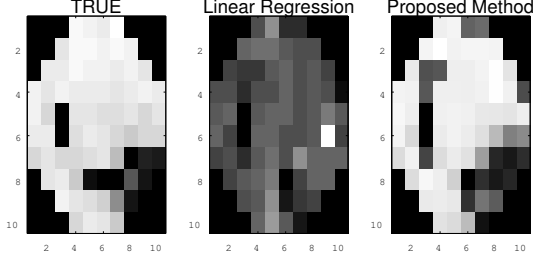
(a) Prediction is based on the three measurements up to the 399th day. The predicted time point is 2462th day. The obtained RMSEs are 19.1 for LR and 7.5 for the proposed method.
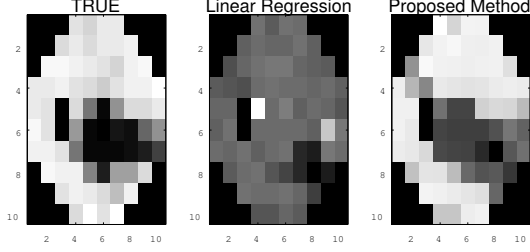
(b) Prediction is based on the three measurements up to the 487th day. The predicted day is th 3477th day. The obtained RMSEs are 18.2 for LR and 6.76 for the proposed method.

Figure 1. Complete visual field prediction we achieved using the proposed spatio-temporal clustering approach. The left panel shows the true result of the visual field test for one of the eyes for a glaucoma patient. The middle panel shows the prediction made only using the test records up to a certain time point. The right panel shows the prediction result obtained by combining the data from the target patient with a database of roughly 1000 follow-up studies. The SVD-EM method for spatial clustering was combined with SC method for temporal extrapolation (see Section II). Darker color indicates lower sensitivity (visual field loss); the two constantly dark pixels in the center left correspond to the blind spot.

patient-wise approach and it can predict the complete visual field accurately. We conclude the paper in Section IV

II. MODELING AND METHOD

A. Problem Setting

Assume that we are given a collection of measurements of D features for N patients. The measurements for the ith patient is a time series consisting of time stamps $\boldsymbol{t}_i = (t_{i1}, t_{i2}, \ldots, t_{iL_i})$ and the observed feature values $X_i \in \mathbb{R}^{D \times L_i}$, where L_i is the number of measurements for the ith patient. We denote the whole data set as $\mathcal{D} = \{\{X_1, \boldsymbol{t}_1\}, \{X_2, \boldsymbol{t}_2\}, \ldots, \{X_N, \boldsymbol{t}_N\}\}$. Note that both the time stamps $\boldsymbol{t}_i$ and the number of measurements L_i are individual to each patient, and the time stamps are not uniformly spaced. We call the rows of X spatial dimension, because the D features in visual field test for glaucoma correspond to spatial features known as the total deviation (TD), though our model can be applied to arbitrary feature vectors.

Our goal is to estimate the measurement value of a new patient, which we call *target patient*, with few measurements at an arbitrary time point in the future, with his or her own data $\{X^*, \boldsymbol{t}^*\}$ combined with the above mentioned database.
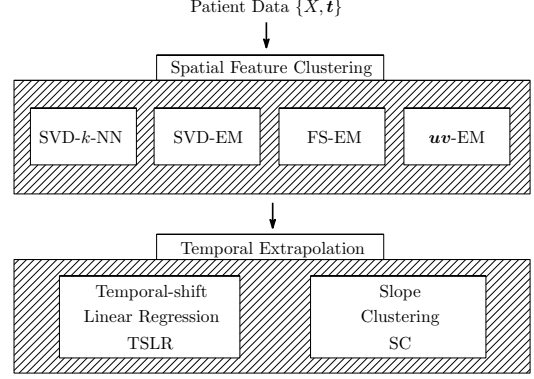

Figure 2. Overall flow of the proposed method.

In this paper, we propose a two-step analysis procedure. The first step is called the spatial feature clustering step, in which we gather patients similar to the target patient X from the dataset $\mathcal{D}$. We propose four alternative approaches, namely, SVD-k-NN, SVD-EM, $\boldsymbol{uv}$-EM, and FS-EM, for spatial feature clustering. The second step is called the temporal extrapolation step, in which we fit the parameters of predictive models to the subset $\mathcal{D}'(\subseteq \mathcal{D})$ of patients who are similar to the target patient found in the first step, and predict the future disease condition. We propose two alternative approaches, namely, temporal-shift linear regression (TSLR) and slope clustering (SC), for the temporal extrapolation step. The proposed framework is schematically illustrated in Figure 2

B. Spatial Feature Clustering

Our method first gathers patients who have similar spatial features. Since the number of measurements for each patient is in general not equal, the challenge here is to derive a patient descriptor having the same length for each patient to make a patient comparable with other patients in the database $\mathcal{D}$. To this end, we propose four spatial feature clustering methods: SVD-k-NN, SVD-EM, $\boldsymbol{uv}$-EM, FS-EM in this subsection. The first three methods uses SVD-style matrix factorization methods for obtaining patient descriptors and the last FS-EM method uses the Fisher score [11].

1) SVD-k-NN and SVD-EM: SVD-k-NN and SVD-EM start from an application of singular value decomposition (SVD) to the data matrix X_i:

$$X_i \simeq U_{ri}\Sigma_{ri}V_{ri}^\top = \sum_{j=1}^r \sigma_{ji}\, \boldsymbol{u}_{ji}\, \boldsymbol{v}_{ji}^\top,$$

where $\Sigma_i = \mathrm{diag}(\sigma_{1i}, \ldots, \sigma_{ri})$, $(\sigma_{1i} \geq \cdots \geq \sigma_{L_i i} > 0)$ and $U_{ri} = [\boldsymbol{u}_{1i}, \ldots, \boldsymbol{u}_{ri}]$, $\Sigma_{ri} = \mathrm{diag}(\sigma_{1i}, \ldots, \sigma_{ri})$, $V_{ri} = [\boldsymbol{v}_{1i}, \ldots, \boldsymbol{v}_{ri}]$. Here, $\boldsymbol{u}_{1i}, \ldots, \boldsymbol{u}_{ri}$ represent spatial features of patient i.

Next we take $r = 1$ for simplicity and denote the spatial feature vector for the target patient by $\boldsymbol{u}_1^*$, and the collection of spatial feature vectors of all the patients in $\mathcal{D}$ by $\mathbf{U} = [\boldsymbol{u}_{11}, \boldsymbol{u}_{12}, \ldots, \boldsymbol{u}_{1N}] \in \mathbb{R}^{D \times N}$. Taking only the first
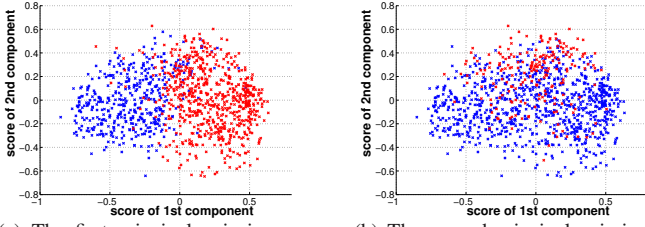
(a) The first principal axis is correlated with the vertical localization of visual field loss. Patients who have more visual field loss on the upper (lower) half of their eyes are plotted as blue (red) crosses, respectively.

(b) The second principal axis is correlated with the horizontal localization of visual field loss. Patients who have more visual field loss on the left (right) half of their eyes are plotted as blue (red) crosses, respectively.

Figure 3. Preliminary analysis of the spatial pattern vectors $\boldsymbol{u}_{1i}$ using PCA. The number of eyes $N = 1086$.

SVD component is not as restrictive as it may sound. In our dataset, the first SVD component contained nearly 85 % of the original signal. Note that taking r greater than one is not straightforward because the ordering of the singular values may not be stable when some singular values are close to each other.

Finally we gather patients in $\mathcal{D}$ that are similar to $\boldsymbol{u}_1^*$ by either performing k-nearest neighbor (k-NN) search in $\mathbf{U}$ or clustering the patients in the dataset by fitting a full-covariance Gaussian mixture model with K components using the EM algorithm (Dempster *et al.* [12]). However, it is known that applying such methods can be challenging when D is large (Beyer *et al.* [13]). To this end, we reduced the dimension by applying principal component analysis (PCA, Pearson [14]) to $\mathbf{U}$ as a preprocessing step. A preliminary analysis of the spatial pattern vectors $\boldsymbol{u}_{1i}$ is shown in Figure 3. The first three PCA components contain roughly 50% of the variance and the first 20 PCA components are enough to capture 90% of the variance. We call the above methods SVD-k-NN and SVD-EM, respectively.

2) $\boldsymbol{uv}$-EM: The next approach for spatial feature clustering called $\boldsymbol{uv}$-EM performs the SVD-like factorization and clustering in a single optimization problem. Let K be the number of clusters and $\{\boldsymbol{u}_k\}_{k=1}^K$ be the cluster centers. Additionally let $\boldsymbol{z}_i \in \{0,1\}^K$ be the latent cluster index variable for the ith subject. Then the complete data likelihood can be written as follows:

$$p(\mathcal{D}, \{\boldsymbol{z}_i\}|\{\pi_k\}, \{\boldsymbol{u}_k\}) = \prod_{i=1}^N \prod_{k=1}^K \{\pi_k p(X_i|\boldsymbol{u}_k)\}^{z_{ki}}, \quad (1)$$

where

$$p(X_i|\boldsymbol{u}_k) = \frac{1}{A_i}\exp\left\{-\frac{1}{2}\|X_i - \boldsymbol{u}_k(\sigma_i\boldsymbol{v}_i)^\top\|_F^2\right\}.$$

Here π_k is the prior class probability for the kth cluster and $A_i = (2\pi)^{DL_i/2}$ is the normalization term. Similarly to the SVD-k-NN and SVD-EM, the data matrix X_i is approximated at $r = 1$, but the approximation and clustering are expressed in a single objective function. Figure 4(a) shows the proposed $\boldsymbol{uv}$-EM algorithm as a graphical model.
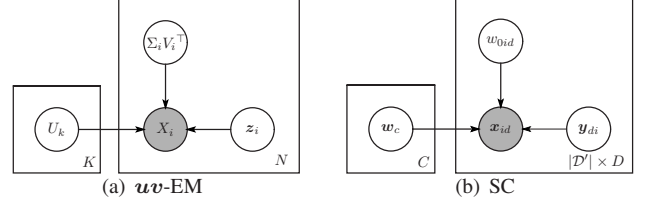

(a) $\boldsymbol{uv}$-EM   (b) SC

Figure 4. Graphical models for $\boldsymbol{uv}$-EM and slope clustering (SC).

Following the standard recipe (see e.g., [15]), an EM algorithm can be derived as follows:

$$\pi_k^{\text{new}} = \sum_i \gamma(z_{ki})/N,$$
$$\boldsymbol{u}_k^{\text{new}} = \frac{\sum_i \gamma(z_{ki})X_i(\sigma_i\boldsymbol{v}_i)}{\sum_i \gamma(z_{ki})\|\sigma_i\boldsymbol{v}_i\|^2},$$
$$(\sigma_i\boldsymbol{v}_i)^{\text{new}} = \frac{\sum_k \gamma(z_{ki})X_i^\top \boldsymbol{u}_k}{\sum_k \gamma(z_{ki})\|\boldsymbol{u}_k\|^2},$$

where $\gamma(z_{ki})$ is the posterior class probability given as $\gamma(z_{ki}) = p(z_{ki} = 1|X_i) = \pi_k p(X_i|\boldsymbol{u}_k)/\sum_k \pi_k p(X_i|\boldsymbol{u}_k)$. We call the above method $\boldsymbol{uv}$-EM in the following.

3) FS-EM: Finally, we employ the Fisher score (Jaakkola and Haussler [11]) for deriving a yet another patient descriptor. Consider the following generative model corresponding to $K = 1$ in the $\boldsymbol{uv}$-EM model (1):

$$p(X_i|\boldsymbol{u}_k) = \frac{1}{A_i}\exp\left\{-\frac{1}{2}\|X_i - \boldsymbol{uv}_i^\top\|_F^2\right\}. \quad (2)$$

The Fisher score for the target patient with the data matrix X_i can be written as follows:

$$s(X|\hat{\boldsymbol{u}}) = \frac{\partial}{\partial\boldsymbol{u}}\log p(X|\boldsymbol{u}, \boldsymbol{v}_i)\Big|_{\boldsymbol{u}=\hat{\boldsymbol{u}}, \boldsymbol{v}_i=\hat{\boldsymbol{v}}_i},$$

where $\hat{\boldsymbol{u}}$ is the maximum likelihood estimator of the generative model (2) with respect to $\mathcal{D}$ and $\hat{\boldsymbol{v}}_i$ is the maximum likelihood estimator for X_i given $\hat{\boldsymbol{u}}$. The maximum likelihood estimators can be obtained by computing the rank one SVD of an auxiliary matrix $\bar{X} = [X_1, \ldots, X_N, X^*]$. Note that the Fisher score is independent of the number of measurements L_i.

Finally we fit a Gaussian mixture model with K components on the Fisher score vectors in a similar manner to the SVD-EM method described above. It is also straightforward to compute the k-NN search based on the Fisher score.

C. Temporal extrapolation

Here we propose two temporal extrapolation models, temporal-shift linear regression (TSLR) and slope clustering (SC).

Let $\mathcal{D}'$ be the subset of the entire dataset consisting of only patients who are found to be similar to the target patient in the spatial feature clustering step including the target patient.

1) Temporal-shift linear regression (TSLR): Let $\phi_m(t)(m = 1, \ldots, M)$ be some fixed basis functions defined over time. The first model assumes the following linear model:

$$X_i = W\Phi_i + \boldsymbol{w}_{i0}\mathbf{1}_{L_i}^\top + E_i \quad (\forall\{X_i, t_i\} \in \mathcal{D}'),$$

where $\Phi_i = \{\phi_m(t_{ij})\}_{mj}$ is an $M \times L_i$ design matrix containing the M predictors, $W \in \mathbb{R}^{D \times M}$ is a coefficient matrix, $\boldsymbol{w}_{i0}$ is a patient specific intercept vector, $\mathbf{1}_{L_i}$ is an L_i dimensional vector containing ones, and E_i is some noise.

Since W does not depend on the patient index i, assuming that the noise E_i is zero mean, we can analytically eliminate $\boldsymbol{w}_{i0}$ as follows:

$$\tilde{X}_i = W\tilde{\Phi}_i + E_i \quad (\forall\{X_i, t_i\} \in \mathcal{D}'),$$

where $\tilde{X}_i := X_i - \frac{1}{L_i}X_i\mathbf{1}_{L_i}$ and $\tilde{\Phi}_i := \Phi_i - \frac{1}{L_i}\Phi\mathbf{1}_{L_i}$ are the centered versions of the measurements and the predictors. The above transformation corresponds to removing the mean from the target variable X_i and the predictor Φ_i.

Then assuming that the number of total measurements $\sum_{i\in\mathcal{D}'} L_i$ is greater than the number of basis functions M, the regression coefficient W can be obtained as follows:

$$W = \bar{X}\bar{\Phi}^\top(\bar{\Phi}\bar{\Phi}^\top)^{-1},$$

where $\bar{X}$ and $\bar{\Phi}$ are the concatenations of $\tilde{X}_i$ and $\tilde{\Phi}_i$ for all patients belonging to the cluster $\mathcal{D}'$ along columns, respectively.

When the predictor is linear $\phi(t) = t$, the estimation of patient specific intercept is equivalent to the estimation of "temporal shift" or time-of-reference; see Figure 6. This is because of the identity $at + b = a(t + b/a)$, in which $t_0 := -b/a$ can be considered as a time-of-reference. For a nonlinear predictor (e.g., sigmoid function), this is not the case anymore.

2) Slope Clustering (SC): The above TSLR model assumes coherence over patients belonging to the same cluster $\mathcal{D}'$ but allows different features D to have different coefficients. The second predictive model assumes that there are C stereotypical progression patterns represented by the regression vectors $\{\boldsymbol{w}_c\}_{c=1}^C$; thus this model allows variation over different patients in the same cluster $\mathcal{D}'$ and at the same time suppresses excess variation over the features; a graphical model for the proposed SC model is shown in Figure 4(b).

More precisely, let $\boldsymbol{y}_{id} \in \{0,1\}^C$ be the latent cluster index variable for the dth feature of the ith patient. Then the complete data likelihood can be written as follows:

$$p(\mathcal{D}', \{\boldsymbol{y}_{id}\}|\{\rho_c\}, \{\boldsymbol{w}_c\}) = \prod_{i\in\mathcal{D}'}\prod_{d=1}^D\prod_{c=1}^C \{\rho_c p(\tilde{\boldsymbol{x}}_{id}|\boldsymbol{w}_c)\}^{y_{cid}},$$

where

$$p(\tilde{\boldsymbol{x}}_{id}|\boldsymbol{w}_c) = \frac{1}{B_i}\exp\left\{-\frac{1}{2}\|\tilde{\boldsymbol{x}}_{id}^\top - \boldsymbol{w}_c^\top\tilde{\Phi}_i\|^2\right\}.$$

Here ρ_c is the prior class probability for class c, $\tilde{\boldsymbol{x}}_{id} = (x_{idj})$ is an L_i dimensional vector containing the centered time evolution of the dth feature for patient i, and $B_i = (2\pi)^{L_i/2}$ is a normalization constant. Note that we have analytically eliminated the patient (and feature) specific intercept $\boldsymbol{w}_{i0}$ by centering $\boldsymbol{x}_{id}$ and Φ_i as above.

Again following the standard derivation, an EM algorithm for the parameters can be obtained as follows:

$$\rho_c^{\text{new}} = \sum_{i\in\mathcal{D}'}\sum_d \gamma(y_{cid})/(D \cdot |\mathcal{D}'|),$$

$$\boldsymbol{w}_c^{\text{new}} = \left(\sum_{i\in\mathcal{D}'}\sum_{d=1}^D \gamma(y_{cid})\tilde{\Phi}_i\tilde{\Phi}_i^\top\right)^{-1}\sum_{i\in\mathcal{D}'}\sum_{d=1}^D \gamma(y_{cid})\tilde{\Phi}_i\tilde{\boldsymbol{x}}_{id},$$

where $\gamma(y_{cid}) = \rho_c p(\tilde{\boldsymbol{x}}_{id}|\boldsymbol{w}_c)/\sum_c \rho_c p(\tilde{\boldsymbol{x}}_{id}|\boldsymbol{w}_c)$ is the posterior class probability.

III. EXPERIMENTS

In this section, we conduct numerical experiments to evaluate our model described in Section II quantitatively.

A. Data

We applied our two-step model to a glaucoma data set.

The data set we analyzed is a sensitivity measurements of 74 locations from 1086 glaucomatous eyes in various stages that had been followed up at Department of Ophthalmology, Graduate School of Medicine, The University of Tokyo. Age at the time of first visual field test was 54.9 ± 13.4 years old and observation follow-up period was 6.6 ± 1.4 years.

The Humphrey Visual Field Analyzer (HFA, Zeiss Humphrey Systems, Dublin, CA), which is the most popular diagnostic system for automated static threshold perimetry, was used to collect the data.

Each patient (or more precisely eye) was measured in total 11.3 ± 3.0 times (measurement interval was 233.3 ± 133.7 days). We evaluated the prediction accuracy of patients at their last visual field observation based on the first d measurements , where $d = 2, \ldots, 10$. For patients who have only $L_i \leq 10$ measurements, we evaluated them based on the first $d = 2, \ldots, L-1$ measurements.

The observed vector contains 74 sensitivity thresholds over all visual field locations, which are acquired at each time of visual field test by HFA central 30-2 full threshold (C30-2) program. This 74-dimensional measurement is called the total deviation (TD) and it quantifies the sensitivity of different locations of the visual field of one eye. Another metric commonly used by clinicians is called the mean deviation (MD) and it is a weighted average of TD with a fixed weight. We present results for predicting both the MD and the TD, in which case $D = 1$ and $D = 74$, respectively.

B. Evaluation Procedure

We split the data into the training set (80%), validation set (10%), and the test set (10%). The training set was used to learn the spatial clustering parameters, namely PCA

projections for SVD-k-NN and SVD-EM, Gaussian mixture model with full covariance for the SVD-EM and FS-EM, cluster centers and mixture weights for uv-EM, maximum likelihood parameter $\hat{u}$ for FS-EM.

We used the validation set for the selection hyper-parameters, namely the number of neighbors k for the SVD-k-NN, and the number of clusters K for the SVD-EM and uv-EM, and FS-EM. Here, the candidate values were $k = 10, 20, 30, 40$ for SVD-k-NN, $K = 10, 30, 50$ for SVD-EM, uv-EM, and FS-EM. The number of PCA dimension was fixed at three, which captures roughly 50% of the variance. We used $C = 3$ for the number of clusters in the SC method in the final evaluation. For simplicity, we use a single basis function $\phi_1(t) = t$ (i.e., $M = 1$) in the following experiments; thus in the case of TSLR, the complexity of the temporal extrapolation step is the same as the conventional patient-wise linear regression approach. The best hyper-parameter values chosen on the validation set were used for the test set, and we repeated the above procedure ten times.

When we evaluated the performance on the test set, we only used the data of the target eye up to the first d measurements, and applied the spatial feature clustering step, which involves SVD, PCA projection, and cluster assignment in the case of SVD-EM method, for example. Note that the clustering parameters were learned on the training set (80%) and fixed. Then we applied the temporal extrapolation methods jointly on the target eye and the eyes belonging to the same cluster (or within the k-nearest neighbor) from the database. The performance is measured by the root mean squared error (RMSE) on the latest visual field measurement of the target eye.

C. Result on the mean deviation (MD)

Predictive errors of all the combinations of the proposed spatial clustering and temporal extrapolation methods, and that of the conventional patient-wise linear regression (LR) are shown in Figure 5(a). We can clearly see that all the proposed methods outperform LR when the number of measurements is smaller than seven. In fact, with only two measurements, all the proposed methods achieve errors that are at least comparable to that of LR with six measurements. This is significant because clinically it is important to provide accurate prediction as early as possible so that the doctors can make necessary actions. The errors of the methods based on SC are slightly lower than those based on the TSLR, especially when the number of measurements is large. This is probably because SC is more flexible than TSLR.

Figure 6 compares the fits obtained by the conventional patient-wise linear regression and the proposed spatial-clustering-based TSLR. We can see that the first couple of MD values for the target patient keep improving due to the learning effect. Thus the patient-wise approach predicts steep
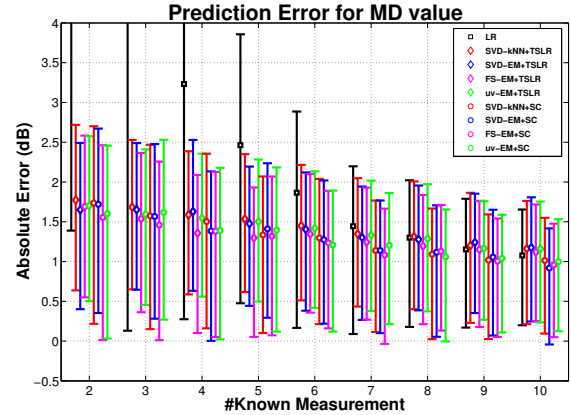
(a) Absolute error for predicting MD value.
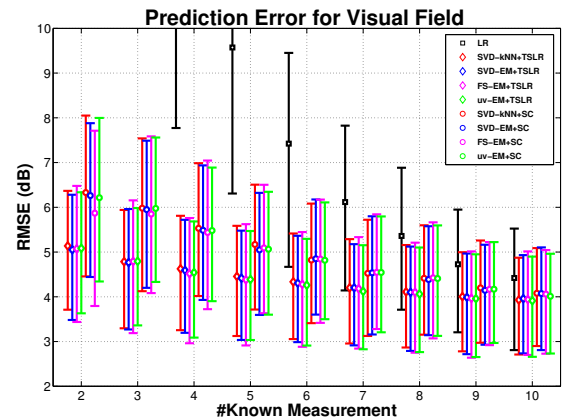
(b) RMSE for predicting TD value.

Figure 5. Prediction error for MD and TD by various methods against the number of known measurements. The center of error-bar is the median of error and the upper end and lower end of the error-bar corresponds to the 25%- and 75%-quartiles.

improvement, though clinically the visual field loss caused by glaucoma is known to be unrecoverable. On the other hand, the proposed method can predict a downward trend because it looks at not only the target patient but also at a collection of similar patients; the predicted line agrees well with the true future measurement. Note also that since we estimate the intercept for each patient individually, we can align the patients along the temporal axis and show in which stage the target patient is in in comparison to other patients in the same cluster.

D. Result on the total deviation (TD)

Predictive errors of all the methods for the prediction of total deviation are shown in Figure 5(b). We see similar trends to the result for MD, but the improvement is larger. In fact, with only two measurements the proposed TSLR method combined with any spatial clustering method achieved a performance that is comparable to the LR method with eight measurements. The error of the SC method is slightly larger than that of TSLR method when the number
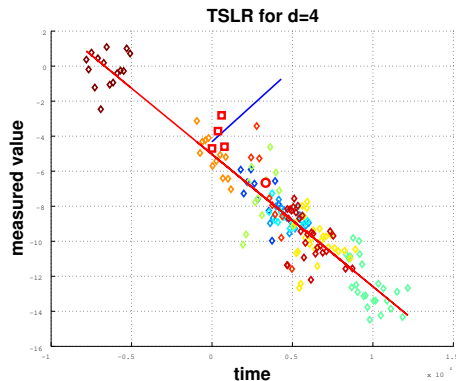
TSLR for d=4

Figure 6. Visualization of the predictions made by the conventional patient-wise linear regression and the proposed TSLR for one patient. The target patient with $d = 4$ measurements is plotted as red squares. The future measurement to be predicted is plotted as a red circle. The fit obtained by the conventional method is shown as a blue line. The fit obtained by the proposed TSLR method is shown as a red line. The data from other patients are plotted as diamonds with individual colors.

of known measurements is small, but they become close for larger number of measurements.

IV. CONCLUSION

In this paper, we have proposed a method to accurately predict the condition of a patient with a progressive disease in the future from few measurements. The main idea is to exploit the similarity between patients and effectively obtain a larger number of samples by applying a spatio-temporal clustering. We have also addressed several issues that typically arise in dealing with medical records, namely, non-uniform sampling, high-dimensionality, and lack of time-of-reference, in a principled manner.

We have applied the proposed method to visual field measurements from glaucoma patients. Our approach was overall 10–70% more accurate than the conventional patient-wise approach depending on the number of measurements. The improvement for TD reached 70% when the number of measurements $d = 2$; this is as accurate as the prediction made by the conventional approach with $d = 8$ (see Figure 5(b)). Considering that the average interval between measurements is half a year to a year, we argue that we have pushed the *prediction of complete visual field* forward to a practical regime for the first time.

Possible future directions include taking the profile of the patients (age, gender, records of other illnesses) into account, making the predictive model more sophisticated, and tackling other types of progressive diseases.

REFERENCES

[1] D. F. Garway-Heath, D. Poinoosawmy, F. W. Fitzke, and R. A. Hitchings, "Mapping the visual field to the optic disc in normal tension glaucoma eyes," *Ophthalmology*, vol. 107, no. 10, pp. 1809–1815, 2000.

[2] J. Katz and A. Sommer, "Reliability indexes of automated perimetric tests," *Arch. ophthalmol.*, vol. 106, no. 9, pp. 1252–1254, 1988.

[3] A. Heijl, G. Lindgren, and J. Olsson, "The effect of perimetric experience in normal subjects." *Archives of ophthalmology*, vol. 107, no. 1, pp. 81–86, 1989.

[4] K. Chan, T.-W. Lee, P. A. Sample, M. H. Goldbaum, R. N. Weinreb, and T. J. Sejnowski, "Comparison of machine learning and traditional classifiers in glaucoma diagnosis," *IEEE Trans. Bio-med. Eng.*, vol. 49, no. 9, pp. 963–974, 2002.

[5] Z. Ghahramani and M. J. Beal, "Variational inference for bayesian mixtures of factor analysers," in *Advances in Neural Informaiton Processing Systems 12*, 2000, pp. 449–455.

[6] K. Chan, T.-W. Lee, and T. J. Sejnowski, "Variational learning of clusters of undercomplete nonsymmetric independent components," *J. Mach. Learn. Res.*, vol. 3, pp. 99–114, 2003.

[7] B. N. Noureddin, D. Poinoosawmy, F. W. Fietzke, and R. A. Hitchings, "Regression analysis of visual field progression in low tension glaucoma," *Brit. J. Ophthalmol.*, vol. 75, pp. 493–495, 1991.

[8] F. W. Fietzke, R. A. Hitchings, D. Poinoosawmy, A. I. Mc-Naught, and D. P. Crabb, "Analysis of visual field progression in glaucoma," *Brit. J. Ophthalmol.*, vol. 80, pp. 40–48, 1996.

[9] C. Mayama, M. Araie, M. Suzuki, K. Ishida, T. Yamamoto, Y. Kitazawa, M. Shirakashi, H. Abe, H. Tsukamoto, H. K. Mishima, K. Yoshimura, and Y. Ohashi, "Statistical evaluation of the diagnostic accuracy of methods used to determine the pregression of visual field defects in glaucoma," *Ophthalmology*, vol. 111, no. 11, pp. 2117–2125, 2004.

[10] R. A. Russell, R. Malik, B. C. Chauhan, D. P. Crabb, and D. F. Garway-Heath, "Improved estimates of visual field progression using bayesian linear regression to integrate structural information in patients with ocular hypertension," *Invest. Ophth. Vis. Sci.*, vol. 53, no. 6, pp. 2760–2769, 2012.

[11] T. S. Jaakkola and D. Haussler, "Exploiting generative models in discriminative classifiers," *Advances in Neural Information Processing Systems*, pp. 487–493, 1999.

[12] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. B. Met.*, vol. 39, no. 1, pp. 1–38, 1977.

[13] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is " nearest neighbor " meaningful?" *International Conference on Database Theory*, pp. 217–235, 1999.

[14] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine Series 6*, vol. 2, no. 11, pp. 559–572, 1901.

[15] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.