1 Generalization of the algorithm for multiple labeling opinion

1.1 background

The problem of conducting similarity studies between different observation has been addressed in many papers. The proposed solutions takes into account the true observation labeling for optimizing the space. In this report we have seen how to implement such algorithms for music based similarity study. However, in many cases we lack the real data labeling. What we have instead; is different labeling provided by more than one user. Those labeling will be based on the user's opinion on the similarity between different observations.

The advancement of query research based on user's behavior and history, raised the need of providing for each user a different label of the observation space. Some problems can be solved by providing for each observation more than one label based on most famous labeling approach. Such problems has been addressed with different solving techniques [2],[3]. This approach however, does not take into account that in many case what is more interesting than finding the true class is to find the most similar response. For example, in our cases we are not interested in finding the true label of the searched query but instead to suggest observations that are the most similar. This notion of similarity is provided by all the previous user's opinion on what two observations(or more) are more similar. We thus have to find a space that respects all the user's opinion and provide the minimum average error between different observations.

In this part of the report, We will consider the problem of having different labeling opinions provided by a ground-truth study. The problem thus will be to minimize the mean error between different labeling using metric learning techniques. We collected 32 different users opinion on similarity between 78 examples of instruments played with different techniques. The metric learning algorithm we used is the same Large Margin Nearest Neighboors [1] and the ranking metrics used for evaluation are the Mean average preicion and the Precision at k.

1.2 Ground Truth Data

The data used for this part was gathered by conducting an online survey. The test included 78 different sounds coming from different instrument and different playing techniques. Subjects were asked to evaluate the similarity between those sounds by hearing each sound and associating it either with a previous sound example or assuming that it is not similar to any previous example. The purpose of such study is to acquire different opinion on the similarity between sounds. Users will thus using their own query will be provided results similar to what they are searching for. This similarity is provided by the acquired opinions of different users alongside the real distance between labels.

1.3 Suggested solutions

To solve this problem different methods where tested with only two providing good results. One of the methods that failed to provide is to use multiple Large Margin Nearest Neighbors in series. This bad results is due to the fact that the last user opinion will have the last saying on how the observations should be presented in the space. We will present in what follows the two solutions, the mathematical motivation and the results.

1.3.1 Sum of LMNN projection matrix

I will start first by taking a simple example of two labeling opinion. Each opinion is given for 5 observations. And for more simplicity I will present the data in a 2D space. The first labeling opinion can be written as $Label_{op1} = [yellow\ yellow\ blue\ blue\ purple]$ and the second one as $Label_{op1} = [blue\ purple\ purple\ purple\ yellow]$. In the top part of figure 1 we can see an example of those two opinion in a 2D space distribution. Note that what interest us in this problem is not finding the true class of each observation rather finding the similarity between different observation. For example, we are interested in the difference in labeling of observation 3 but what is relevant to our problem is the fact that observations 3 and 4 are similar in both labeling.

The first method I will present based on this example, is applying the metric learning algorithm for each opinion alone and take the sum of the projection matrices. The metric learning example that we will base our study on is the same LMNN discussed in a previous chapter. Let us start by looking at figure 1 to further understand how the application of LMNN will affect different labeling opinion. by looking at the first labeling opinion we can clearly see that the effect of applying the LMNN will be to make the observations having the same labels(1,2 and 3,4) closer and the observations having different labels further apart. We will not discuss again this aspect of LMNN but we have to note that for labeling classes containing less than 2 observations the algorithm will not work. This is why we have to train the algorithm with suppressing the classes having less than two observations.

At the end of our algorithm we will be comparing euclidean distances in the projected space. Let us write the matrix of pairwise distances between different observation before

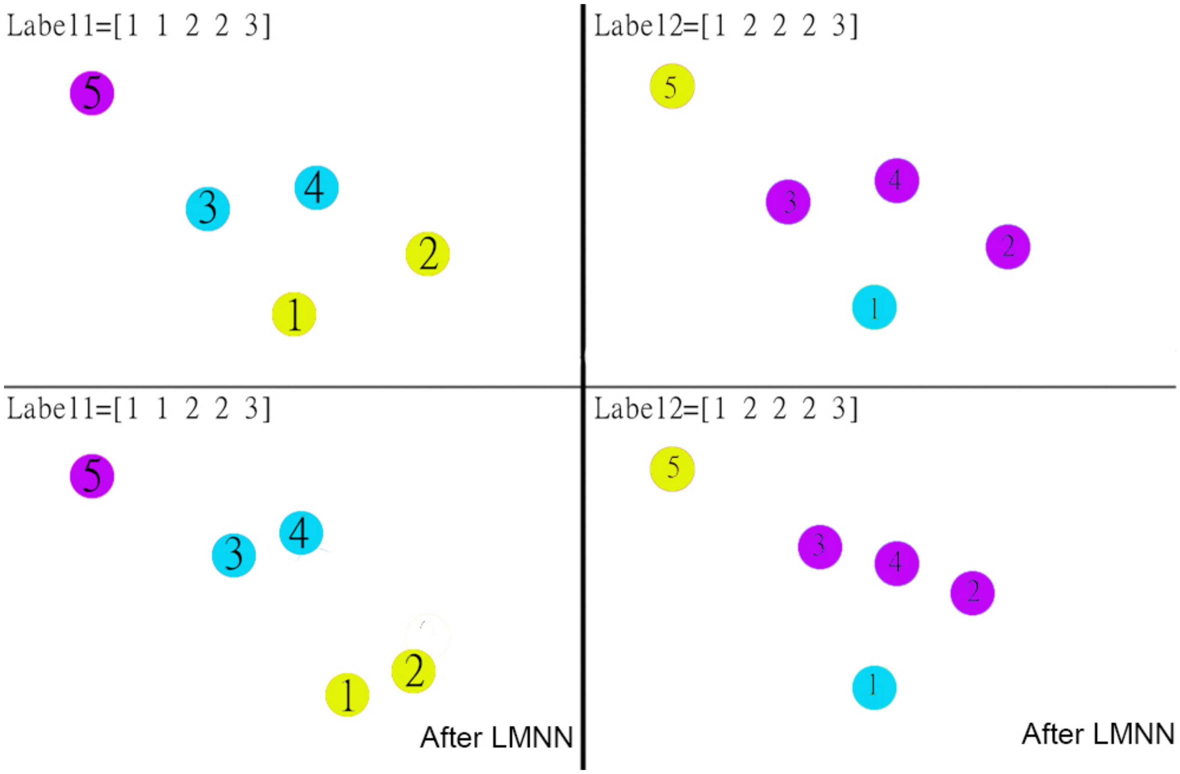

Figure 1: LMNN applied on two different labeling opinion

the application of LMNN as a reference. And then we will compare the distances in the new space and evaluate the performance of the method. The pair wise distance matrix is thus given by :

$$\begin{bmatrix} 1 & d_{12} & d_{13} & d_{14} & d_{15} \\ d_{21} & 1 & d_{23} & d_{24} & d_{25} \\ d_{31} & d_{32} & 1 & d_{34} & d_{35} \\ d_{41} & d_{42} & d_{43} & 1 & d_{45} \\ d_{51} & d_{52} & d_{53} & d_{54} & 1 \end{bmatrix}$$

Now that we have our reference pair wise distance matrix, let us try to find how this matrix will look like after applying LMNN. The sum that we will be computing is based on the matrix L, where $M = L^T L$ is the Mahalanobis matrix. thus we will have the following :

$$(\vec{x_i} - \vec{x_j})^T M_{sum} (\vec{x_i} - \vec{x_j})$$

$$= (\vec{x_i} - \vec{x_j})^T (L_1 + L_2)^T (L_1 + L_2)(\vec{x_i} - \vec{x_j})$$

$$= (\vec{x_i} - \vec{x_j})^T (L_1^T L_1 + L_2^T L_2 + L_1^T L2 + L_2^T L1)(\vec{x_i} - \vec{x_j})$$

$$= (\vec{x_i} - \vec{x_j})^T M_1 (\vec{x_i} - \vec{x_j}) + (\vec{x_i} - \vec{x_j})^T M_2 (\vec{x_i} - \vec{x_j}) + (\vec{x_i} - \vec{x_j})^T L_1^T L_2 (\vec{x_i} - \vec{x_j}) + (\vec{x_i} - \vec{x_j})^T L_2^T L_1 (\vec{x_i} - \vec{x_j})$$

$$= D_{M_1}(\vec{x_i} - \vec{x_j}) + D_{M_2}(\vec{x_i} - \vec{x_j}) + (L_1\vec{x_i} - L_1\vec{x_j})^T (L_2\vec{x_i} - L_2\vec{x_j}) + (L_2\vec{x_i} - L_2\vec{x_j})^T (L_1\vec{x_i} - L_1\vec{x_j}) \; [1]$$

The last two terms of the last equation can not be related to an euclidean distance rather it is a compensation between two difference. Those two subtractions are made in the two different spaces of the two labeling opinions. To understand better this concept, let us look at the first two mahalanobis distance as being an euclidean distance calculated in the new space after LMNN. We know that applying LMNN will result in smaller distances between the k nearest labeling of same class and bigger distance between labeling of different class that violate a certain margin of distances. We will consider in our example that except the observation 5 all other observations will be affected by the application of LMNN. Thus two observation having the same label will have smaller distance in the new space and two observations having different labels will have bigger distances.

We will look at all the different cases that might be encountered and see if the idea of applying the sum on the projection matrix will have good results theoretically .

1. The two labeling opinion agrees on the similarity.
   Since the application of LMNN will conduct to a smaller distance if the labels are similar and the observations are target neighbors we know that in this case both distances will be smaller in the new space. Observation 3 and 4 in figure 1 illustrate that idea. In this case we will have the two distances D_{M_1} and D_{M_2} smaller than the original distance in the space before the application of LMNN. Let us now go back to update our new pair wise distance matrix:

   $$D_{M_1}34 < d_{34} \ and \ D_{M_2}34 < d_{34}$$

   the sum of those two terms is thus :

   $$D_{M_1}34 + D_{M_2}34 < 2d_{34}$$

   . Since the last two parameters of equation 1 can be viewed as euclidean distance effected on two different projection simultaneously we can say that those two terms will also have smaller distance than the distance without the projection. at the and we will have that in the new space

   $$D_{sum}34 < 4d_{34}$$

2. The two labeling opinion agrees on the non similarity.
   The same analogy can be made for the case were both opinion agree that the two observation belongs to different classes. If we take for example the observations 1 and 3 we will have at the end

   $$D_{sum}13 > 4d_{13}$$

3. The two labeling opinion dos not agrees on the similarity.
   If the two labeling opinion disagree on the similarity between two observation, in

one space we will have smaller distance while in the second we will have bigger distance. For example if we look at samples 2 and 4 in figure 1, we can see that in the new spaces distance between those two observations is bigger for the new space with LMNN based on labeling opinion 1, and bigger in the space with LMNN based on labeling opinion 2. we will thus have :

$$D_{M_1}24 > d_{24} \textit{ and } D_{M_2}24 < d_{24}$$

. Let us introduce two factors $\alpha > 1$ and $\beta < 1$ this implies :

$$D_{M_1}24 = \alpha d_{24} \textit{ and } D_{M_2}24 = \beta d_{24}$$

The last two factors in equation 1, will each have a resultant bigger or smaller than the original distance based on the factor by which the distances are represented in the new space. So for example if the distance between two observations in one opinion is twice smaller than the original distance and for the other opinion it is bigger with a factor of 1.2, the resultant distance will be smaller. we will consider this factor as equal to γ with γ having the value smaller or bigger than 1. We will thus have :

$$D_{sum}24 = \alpha d_{24} + \beta d_{24} + 2\gamma d_{24}$$

.

4. An observation is not affected in both spaces by LMNN.
   If the distance between two observations that are not target neighbors for the LMNN is bigger than a certain margin the distance will not be affected directly. This distance will slightly change because the observation themselves have been displaced in their own small margin. We can thus make the assumption that the variation in distance in the new space is negligeable and the for example the distance between observation 1 and 5 will be

$$D_{sum}15 = 4d_{15}$$

Now that we have formulated all the factors we can construct our new matrix based on the sum of the projection matrix L. To simplify the notation we will take on factor to represent the different factors in case 3, for example let $4A = \alpha + \beta + 2\gamma$.

$$\begin{bmatrix} 1 & 4Ad_{12} & > 4d_{13} & > 4d_{14} & 4d_{15} \\ 4Ad_{21} & 1 & 4Cd_{23} & 4Bd_{24} & 4d_{25} \\ > 4d_{31} & 4Cd_{32} & 1 & < 4d_{34} & 4d_{35} \\ > 4d_{41} & 4Bd_{42} & < 4d_{43} & 1 & 4d_{45} \\ 4d_{51} & 4d_{52} & 4d_{53} & 4d_{54} & 1 \end{bmatrix}$$

Let us now examine two cases, one were we have an agreement and one were we have a disagreement :

1. Case were we have agreement between different opinions.
   It should be noted again that we are searching to compare distances and not the exact value of the distances. So let us take two distances were in the first both users agree that two samples belongs to the same class and in the other they agree that they don't. for example we can take d_{13} *and* d_{34}. In the space before the application of LMNN we had : $d_{13} - d_{34}$ in the new space we will have a difference between a value that is 4 time bigger than the first distance and a value that is 4 time smaller than the second distance. So we have been able to have a space that make the difference in distances bigger if we have two exact opinions on labeling. And for both labeling opinions the error will be minimized.

2. Case were we have disagreement between opinions. Let us take again distance $d_1 3$ but this time let us compare it with $d_2 1$. Let us take F as a value that is bigger than 1 to represent the factor by which in the new space the distance between samples 1 and 3 is bigger. we will thus have in the new space $4Fd_{13} - 4Ad_{d21}$. We now that the first distance is going to be bigger than the original distance between samples 1 and 2. On the other hand the distance between 1 and 2 in the new space will depend on the original distance between 1 and 2. So if the distance in the new space is smaller than in the original space, labeling based on first user will have smaller error than labeling based on second user. But since the first factor is well optimized we will have optimization in both case.
   This idea that the "winner" between the two opinion is based on the representation in the original space is very important. And it will be essential in a case were we have more than two opinions, since the probability that two samples will be considered similar is related to the real distance.

Let us now generate the equation to a case of 'nobs' different opinions with 'nlab' different labels :

$$L = \sum_{k=1}^{nobs} L_k$$

$$(\vec{x_i} - \vec{x_j})^T \sum_{k=1}^{nobs} L_k^T \sum_{k=1}^{nobs} L_k (\vec{x_i} - \vec{x_j})$$

$$= (\vec{x_i} - \vec{x_j})^T \sum_{k=1,w=1}^{nobs} L_k^T L_w (\vec{x_i} - \vec{x_j})$$

$$= (\vec{x_i} - \vec{x_j})^T (\sum_{k=1,w=1,k=w}^{nobs} L_k^T L_w + \sum_{k=1,w=1,k\neq w}^{nobs} L_k^T L_w)(\vec{x_i} - \vec{x_j})$$

$$= (\vec{x_i} - \vec{x_j})^T \sum_{k=1,w=1,k=w}^{nobs} L_k^T L_w (\vec{x_i} - \vec{x_j}) + (\vec{x_i} - \vec{x_j})^T \sum_{k=1,w=1,k\neq w}^{nobs} L_k^T L_w (\vec{x_i} - \vec{x_j})$$

$$\sum_{k=1}^{nobs} d_{M_k}(\vec{x_i} - \vec{x_j}) + \sum_{k=1,w=1,k\neq w}^{nobs} (L_k\vec{x_i} - L_k\vec{x_j})^T(L_w\vec{x_i} - L_w\vec{x_j})$$

The same analogy can be made again on the multi-opinion examples, and we will have again a smaller mean average error.

1.3.2 results

	map before	map after
MFCC	50.74	55.02(+4.28)
Scattering	37.73	45.83(+8.1)

Table 1 : Mean average precision comparison before and after the application of the LMNN sum method.

	map before	map after
MFCC	55.66	60.99(+5.33)
Scattering	40.63	53.93(+15.3)

Table 2 : Precision at 5 comparison before and after the application of the LMNN sum method.

References

[1] K. Q. Weinberger, L. K. Saul. *Distance Metric Learning for Large Margin Nearest Neighbor Classification.* Journal of Machine Learning Research (JMLR) 2009

[2] G. Tsoumakas and I. Katakis, “Multi label classification: An overview,” International Journal of Data Warehousing and Mining, vol. 3, no. 3, pp. 1–13, 2007.

[3] Min-Ling Zhang and Zhi-Hua Zhou, “Ml-knn: A lazy learning approach to multi-label learning,” Pattern Recogn., vol. 40, no. 7, pp. 2038–2048, 2007.