Robust Affinity Propagation using Preference Estimation

Kai-Chao Yang
National Tsing Hua University
30013, Hsinchu, Taiwan
kcyang@vc.cs.nthu.edu.tw

Chang-Hsin Yu, Chen-Jung Hung
National Tsing Hua University
30013, Hsinchu, Taiwan
chyu@vc.cs.nthu.edu.tw

Jia-Shung Wang
National Tsing Hua University
30013, Hsinchu, Taiwan
jswang@cs.nthu.edu.tw

ABSTRACT

Affinity propagation is a novel unsupervised learning algorithm for exemplar-based clustering without the priori knowledge of the number of clusters (*NC*). In this article, the influence of the "preference" on the accuracy of AP output is addressed. We present a robust AP clustering method, which estimates what preference value could possibly yield an optimal clustering result. To demonstrate the performance promotion, we apply the robust AP on picture clustering, using local SIFT, global MPEG-7 CLD, and the proposed preference as the input of AP. The experimental results show that over 40% enhancement of ARI accuracy for several image datasets.

Keywords

Affinity propagation; classification algorithms; clustering method; image classification

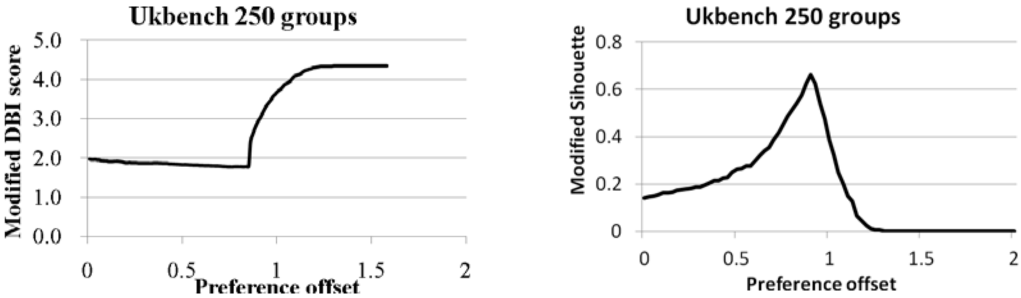
1. INTRODUCTION

Traditionally, clustering problems are solved by learning a set of centers to minimize the sum of squared errors between data points and their nearest centers. These centers might be selected from actual data points called "exemplars" or virtual ones. A representative method is the K-centers clustering algorithm. Different from the traditional clustering algorithms, Frey and Dueck proposed an unsupervised clustering algorithm called Affinity propagation (AP) [Jou01a], which considers every data point as a potential exemplar and iteratively exchanges messages between data points to determine the most suitable exemplars. In [Jou01a], AP has shown its outstanding performance in several areas, such as face detection, gene and exon finding, representative sentence detecting, and air-travel routing. Since the proposal of AP, there have been a lot of relative discussions about it. Some researches focus on variations of AP [Con01a] or combination of K-centers algorithm and AP [Con02a].

Despite the topics extended from AP, this article focuses on improving the accuracy of AP output. The input of AP consists of two parts, the similarity matrix and the preference. Because AP does not need to pre-assign the number of clusters (*NC*), the similarity and the preference are the keys of the clustering results of AP. There are rare discussions about the preference. In [Jou01a], the preference value is suggested to be the median of the similarity matrix. However, according to the observation, the decision of the preference value has a significant impact on clustering results. A median preference value may not lead to the optimal clustering result. In [Jou02a], a method to scan preference for finding the optimal clustering solution was proposed. Preference changes adaptively in the process of AP to find more reliable convergence, but progressive scanning takes a lot of time. Besides, the definition of "optimal solution" was not clearly defined in the paper.

In this article, we use some famous criteria that have been frequently used as the clustering accuracy indices of the AP output, such as the Adjusted Rand Index (ARI) [Jou03a], the Davies-Bouldin Index (DBI) [Jou04a], and the Silhouette index 0[Jou05a] to measure the accuracy of the AP output, and discuss the relationship between the preference and clustering results. We further extend the observations about the preference in [Url01a] to a preference estimation algorithm, which could possibly find out the most suitable preference value. The estimated preference can significantly improve the accuracy of AP. Besides, it can be used in all related researches mentioned above as long as they apply AP for clustering. For demonstration, we choose some image databases as the source material, and then classify images using AP with the proposed preference estimation.

The remaining part of the article is organized as follows. In Section 2, we first address the importance of the preference to AP. Then, according to our observation, the robust AP using preference



(a) (b)

Figure 1. (a) Modified DBI scores and (b) Modified Silhouette Index scores of AP output using different preference values. The selected preference value is reference preference + offset, where the reference preference is set as med(Sim)+min(med(Sim)).

estimation is proposed. In Section 3, we apply the proposed robust AP to the image clustering to demonstrate its performance. Some experimental results are presented here. Finally, the discussions and future works are given in Section 4.

2. ROBUST AFFINITY PROPAGATION

Preference of Affinity Propagation

Affinity Propagation (AP) views each data as a potential exemplar, and recursively passes messages between points until corresponding clusters emerge. In AP, there are two types of messages communicated between data points. The “responsibility” $r(i, k)$, sent from data point i to point k reflects how well point i favors the point k as its exemplar over other candidate exemplars. The “availability” $a(i, k)$ sent from point k to point i reflects how well point k favors itself as an exemplar of point i over other candidate exemplars. For $k = i$, the $r(k, k)$ represents the *preference* of data point k to be chosen as an exemplar.

The preference of AP plays an important role. When the preference of a candidate exemplar k is larger, the responsibility $r(k, k)$ and the availabilities $a(i, k)$ for all i are stronger, so that it is likely that the point k becomes an exemplar eventually. This means that the number of identified clusters is increasing with the preference correspondingly. According to the observation, when the preference is small enough, all data points will be classified into the same cluster. On the contrary, every data point will form a cluster itself when the preference is increasing to near zero. Note that the number of output clusters (*NC*) is a non-decreasing function of preference. In [Jou01a], the preference is suggested to be the median of the similarity matrix. This leads to a moderate number of clusters. However, this selection may not lead to the optimal clustering result because it does not consider dataset content and may produce too many or too few clusters.

Optimality of the Preference and Unimodal

In this sub-section, we discuss how to evaluate the accuracy of clustering results and how the preference affects the clustering results.

The clustering result is usually compared with the ground-truth to evaluate the accuracy. The adjusted rand index (ARI) [Jou03a] is a popular similarity measure of agreement between two partitions. However, there is no ground-truth for comparisons in many cases. In this situation, some other criteria can be used instead of ARI, such as the Davies-Bouldin Index (DBI) [Jou04a] and Silhouette Index [Jou05a].

Let $a(x)$ be the ARI score at preference value x. Then we observe that $a(x)$ tends to be a unimodal function Thus there exists a point p such that $a(x)$ is increasing for $x \le p$ and decreasing for $x \ge p$. Therefore, we can clearly define that , the optimal solution occurs at preference = p.

As for the DBI, we assume $d(x)$ is the DBI score at preference value x. Note that lower DBI scores represent better clustering results. We find that $d(x)$ monotonically decreases with x. This trend results from two facts:

1) *NC* is a non-decreasing function of the preference value x.
2) $d(x)$ is a non-increasing function of *NC*.

DBI takes all data points into consideration, even though some data points only contain itself as singleton or twin. This situation results in the best score of $d(x)$ for these extreme small clusters. As the preference value increases, these trivial clusters contribute much and results in the best DBI score. This is the main reason that ARI and DBI induce different results. In fact, the Silhouette Index that gives the optimal score as long as the preference is larger than a certain value also has the same problem. However, these extreme small clusters caused by large preference values are less informative comparing to those of large size, so it is not

```
top = 0;
bottom = med(Sim) + min(med(Sim));
cut = (top + bottom) / 2;
slope = (DBI_modified(top) – DBI_modified(cut)) / (top – cut);
FOR i = 1 TO round_number
 IF slope < 0
  top = (top-bottom) × 2 + bottom;
  bottom = cut;
  cut = (top-bottom) / 2 + bottom;
  slope = (DBI_modified(top) – DBI_modified(cut)) / (top – cut);
 ELSE
  top = cut;
  cut = (top-bottom) / 2 + bottom;
  slope = (DBI_modified(top) – DBI_modified(cut)) / (top – cut);
 END
END
```

Table 1. Adjusted binary search.

reasonable to give them better scores. Therefore, we modify the evaluating criteria in order to reflect the useful clustering information. The rule is: if the number of data points in a cluster is smaller than a threshold *T*, all points within this cluster are considered isolated and excluded from the evaluation.

Figure 1(a) and (b) show the modified DBI scores and modified Silhouette Index scores, respectively. The dataset is UKbench 250 groups [Url02a], and the preference value is selected as (the referenced preference + offset), where the offset (the x-axis) ranges from 0 to 2. Here the referenced preference is defined as med(Sim) + min(med(Sim)), the sum of the median vector of the similarity matrix and the minimum value of the median vector of the similarity matrix. Here "Sim" denotes the similarity matrix, and "med" represents median. From Figure 1, it can be seen that the curve trend becomes a unimodal function after the removal of trivial clusters.

Preference Estimation

According to the above simulations in Fig. 1, we found that it is highly likely that the above accuracy indices can be transformed to unimodal functions of the preference value. Thus we can estimate the preference that generates the optimal solution. No matter what accuracy index is used, the relationship between the preference and the clustering result of AP can be concluded as follows: Initially, when extremely small preference value is set, every data point is unlikely being the exemplar of other points. When the preference value increases, some data points start being accumulated as different clusters. Then more and more clusters are produced, and the attractive force of every exemplar is getting stronger as well. However, when the preference is larger than a certain value, too many exemplars will be generated. This scatters the attractive force of every exemplar again.

The next step is to search for the preference that generates the optimal clustering result. Table 1 shows the proposed adjusted binary search algorithm that has time complexity of O(log *n*). The proposed algorithm takes advantage of the feature that a unimodal function has exact one extreme value. Since most online image databases have no ground-truth, it is unlikely to evaluate ARI scores for these images. Thus we put emphasis on the DBI, but any other index that can be transformed to a unimodal function can apply the proposed algorithm.

The search starts by calculating the modified DBI scores of two extreme points, preference = med(Sim)+min(med(Sim)) and 0. Then the point with higher scores is iteratively replaced by the median point between the prior two search points. The number of iterations decides the searching time and accuracy. After enough number of iterations, a near optimal preference can be obtained.

Estimated Preference Adjustment

The preference estimation process can generate the best score of the accuracy index. However, for the same dataset we observe that different accuracy indices may imply different optimal preference values. For example, when we use Ukbench 250 image dataset as the input data, the estimated preference offset in the ARI is about 0.88. In Figure 1(b), the same result is observed in the modified Silhouette Index. However, in Figure 1(a), the estimated preference offset becomes 0.76 in the modified DBI. This situation is mainly because different criteria are used by different accuracy indices. Among these accuracy indices, the estimated preference in the ARI should be treated as the optimal preference because ARI scores are evaluated using the ground-truth.

From the simulation using different datasets, we observe that the quotient of the estimated preference offsets between the ARI and modified DBI is roughly fixed. Therefore, at beginning we can train a small number of data points to get the ARI score and DBI score. Then the quotient between them can be calculated. When taking all data points into account, the optimal preference offset can be approximated by the estimated preference offset in the modified DBI multiplying the quotient. Then AP is able to use this preference and the similarity matrix as its input.

Sub-cluster Combination

Although the clustering result of AP using the proposed preference can get nearly optimal ARI scores, we still observe that there are some small clusters containing only one or two members. We call them sub-clusters. To solve this problem, we use AP again to generate new clusters with only exemplars of sub-clusters as input. If two exemplars are classified as the same cluster, all members of these two exemplars are combined as one big cluster. Because the influence of non-sub-clusters has been removed, AP can generate more accurate results than that taking all pictures as input.

Image dataset	ARI scores using median preference	ARI scores using preference estimation and sub-cluster combination		Enhancement (%)	
		9 runs	11 runs	9 runs	11 runs
Ukbench 250	0.633	0.717	0.810	13.21	27.97
Ukbench 350	0.545	0.723	0.743	32.74	36.40
Ukbench 450	0.542	0.746	0.761	37.69	40.59
Ukbench 550	0.536	0.726	0.739	35.42	37.85
Ukbench 650	0.494	0.689	0.713	39.63	44.42
Ukbench 750	0.483	0.570	0.707	18.06	46.50
Average	-	-	-	29.46	39.00

Table 2. ARI score enhancement using preference estimation and sub-cluster combination.

3. CASE STUDY

The proposed algorithm in this article can be used in all areas that are suitable for applying AP or other variations of AP. In this section, we applied AP in the image clustering to demonstrate the performance of the proposed algorithm.

We use SIFT and MPEG-7 to extract features from each image, and then calculate the similarity between images. The input of AP includes the similarity matrix and preference. After the similarity matrix has been obtained, the next step is to determine the preference. Initially, the modified DBI scores are calculated using the preference value equaling med(Sim)+min(med(Sim)) and 0, respectively. Then the adjusted binary searching algorithm is applied to estimate the optimal preference offset in the modified DBI. Next, the estimated preference offset multiplies 1.11 to approximate the ARI result. After that, AP is applied to classify images using this preference. Finally, the proposed sub-cluster combination is applied to combine sub-clusters, and produce the final clustering result.

Six different image datasets from Ukbench database [Url02a] are applied to demonstrate the proposed algorithm. UK benchmark database was proposed by Nister and Stewenius. In this database, we pick out 250, 350, 450, 550, 650, and 750 groups as our experimental datasets. We resize the shorter edge of every picture to 120 pixels, but keep the ratio of height and width unchanged.

The final preference we obtained is used to calculate its ARI score, so that our clustering result can be compared with the ground-truth. In Table 2, we show the ARI scores of AP output using the estimated preference and the ARI scores using preference selection in [Jou01a], respectively. It can be seen that after 11 iterations, the ARI score improves on average 39%.

4. CONCLUSION

In this article, we proposed a robust AP method. Our contribution is to suggest an estimation procedure of what preference value can yield an optimal solution of AP output. First, we transform the accuracy index to a unimodal function. Then, the adjusted binary search algorithm is used to estimate the optimal preference. Furthermore, the sub-cluster combination is proposed to refine the clustering result. The case study of image clustering shows that 39% performance improvement can be achieved.

5. REFERENCES

[Jou01a] Frey, B. J., and Dureck, D., Clustering by passing messages between data points. Science, Vol. 315, No. 5814, pp. 972-976, 2007.

[Con01a] Zhang, X., Gao, J., Lu, P., and Yan, Y., A novel speaker clustering algorithm via supervised affinity propagation, IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 4369-4372, 2008.

[Con02a] Zhang, X., Wang, W., Nørvag, K., and Sebag, M., K-AP: generating specified K clusters by efficient affinity propagation, IEEE International Conference on Data Mining, pp. 1187-1192, 2010.

[Jou02a] Wang, K., Zhang, J., Li, D., Zhang, X., and Guo, T., Adaptive affinity propagation clustering, Acta Automatica Sinica, vol. 33, no. 12, pp. 1242-1246, 2007.

[Jou03a] Hubert, L., and Arabie, P., Comparing partitions, Journal of Classification, vol. 2, no. 1, pp.193–218, 1985.

[Jou04a] Davies, D.L., and Bouldin, D.W., A cluster separation measure, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 1, no. 2, pp. 224-227, 1979.

[Jou05a] Rousseeuw, P., Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, Journal of Computational and Applied Mathematics, vol. 20, no. 1, pp. 53-65, 1987.

[Url01a] Preference and the Number of Clusters. Available: http://www.psi.toronto.edu/affinitypropagation/faq.html#clusters/

[Url02a] UK benchmark database. Available: http://vis.uky.edu/~stewe/ukbench/