

ANEMI: An Adaptive Neighborhood Expectation-Maximization Algorithm with Spatial Augmented Initialization

Tianming Hu^{1,2}, Hui Xiong³, Xueqing Gong¹, and Sam Yuan Sung⁴

¹ East China Normal University

tmhu@ieee.org, xqgong@sei.ecnu.edu.cn

² Dongguan University of Technology

³ Rutgers, the State University of New Jersey

hxiong@andromeda.rutgers.edu

⁴ South Texas College

sysung@southtexascollege.edu

Abstract. The Neighborhood Expectation-Maximization (NEM) algorithm is an iterative EM-style method for clustering spatial data. Unlike the traditional EM algorithm, NEM has the spatial penalty term incorporated in the objective function. The clustering performance of NEM depends mainly on two factors: the choice of the spatial coefficient, which is used to weigh the penalty term; and the initial state of cluster separation, to which the resultant clustering is sensitive. Existing NEM algorithms usually assign an equal spatial coefficient to every site, regardless of whether this site is in the class interior or on the class border. However, when estimating posterior probabilities, sites in the class interior should receive stronger influence from its neighbors than those on the border. In addition, initialization methods deployed for EM-based clustering algorithms generally do not account for the unique properties of spatial data, such as spatial autocorrelation. As a result, they often fail to provide a proper initialization for NEM to find a good solution in practice. To that end, this paper presents a variant of NEM, called ANEMI, which exploits an adaptive spatial coefficient determined by the correlation of explanatory attributes inside the neighborhood. Also, ANEMI runs from the initial state returned by the spatial augmented initialization method. Finally, the experimental results on both synthetic and real-world datasets validated the effectiveness of ANEMI.

1 Introduction

Spatial data have traditional numeric and categorical attributes as well as spatial attributes that describe the spatial information of the objects, such as location and shape. The assumption of independent and identical distribution (IID) is no longer valid for spatial data. In the spatial domain, everything is related to everything else but nearby objects are more related than distant objects [1]. For example, houses in nearby neighborhoods tend to have similar prices which are

affected by one another. In remote sensing images, close pixels usually belong to the same landcover type: soil, forest, etc.

Traditional model based clustering algorithms, such as the Expectation-Maximization (EM) algorithm [2], do not take spatial information into consideration. To this end, Ambroise et al. [3] proposed the Neighborhood Expectation-Maximization (NEM) algorithm, which extends EM by adding a spatial penalty term into the objective function. Such a spatial penalty favors those solutions where neighboring sites are assigned to the same class. The performance of NEM depends mainly on two factors. One is the choice of the spatial coefficient, which is used to weigh the penalty term in the objective function and specifies the degree of spatial smoothness in the clustering solution. Another one is the initial state of cluster separation, from which NEM starts iterative refinement.

For the choice of the spatial coefficient, NEM employs a fixed coefficient that has to be determined a priori and is often set empirically in practice. However, it may not be appropriate to assign a fixed coefficient to every site, regardless of whether it is in the class interior or on the class border. When estimating posterior probabilities, sites in the class interior should receive stronger influence from its neighbors than those on the border. In addition, for initialization, it is usually impossible for NEM to achieve the global optimization, which has been shown to be NP-hard [4]. The clustering performance of NEM is very sensitive to the initial state of cluster separation. As a result, a proper initialization is of great value for the success of finding a better sub-optimal solution in practice. Nevertheless, existing initialization methods for NEM and other EM-style clustering algorithms do not account for such spatial information.

To address the above mentioned challenges, in this paper, we propose a variant of NEM: Adaptive Neighborhood EM with spatial augmented Initialization (ANEMI) for spatial clustering. ANEMI exploits a site-sensitive coefficient, which is determined by the correlation of explanatory attributes inside the neighborhood. In addition, the refinement process of ANEMI starts from the initial state returned by the spatial augmented initialization method. Indeed, by pushing spatial information further into the whole clustering process, our experimental results on both synthetic and real datasets show that ANEMI generally leads to better clustering performance than traditional NEM.

Overview. The remainder of this paper is organized as follows. Section 2 introduces the problem background and related work. In Section 3, after reviewing the basic concepts of NEM, we present the ANEMI algorithm. Experimental results are reported in Section 4, where both the augmented initialization and the adaptive coefficient assignment of ANEMI are evaluated thoroughly. Finally, in Section 5, we draw conclusions and suggest future work.

2 Background

In this section, we first introduce the background by formulating the problem. Then, we briefly review related work.

2.1 Problem Formulation

The goal of spatial clustering is to partition data into groups so that pairwise dissimilarity, in both non-spatial space and spatial space, between those assigned to the same cluster tend to be smaller than those in different clusters. In detail, we are given a spatial framework of n sites $S = \{s_i\}_{i=1}^n$, which are described with an observable set $X = \{\mathbf{x}_i \equiv \mathbf{x}(s_i)\}_{i=1}^n$ of random variables. Note that we overload notation and use X to refer to both the given dataset and their corresponding random variables. Often it is enough to know the neighborhood information, which can be represented by a contiguity matrix W with $W_{ij} = 1$ if s_i and s_j are neighbors and $W_{ij} = 0$ otherwise. We need to infer the unobservable (hidden) set $Y = \{y_i \in \{1, 2, \dots, K\}\}_{i=1}^n$ of random variables, corresponding to the cluster label of \mathbf{x}_i . Due to the spatial constraint, the resulting random field defined over Y is a Markov Random Field (MRF), where $P(y_i | Y - \{y_i\}) = P(y_i | \{y_j : W_{ij} = 1\})$. Hence it is more appropriate to model the posterior distribution of y_i as $P(y_i | \mathbf{x}_i, \{y_j : W_{ij} = 1\})$ instead of $P(y_i | \mathbf{x}_i)$.

2.2 Related Work

There are roughly two categories of work that are related to the main theme of this paper: spatial clustering and the cluster initialization methods for iterative refinement clustering.

Most conventional clustering methods in the literature treat each object as a point in the high dimensional space and do not distinguish spatial attributes from non-spatial attributes. These clustering methods can be divided into the following groups: distance-based [5], density-based [6], hierarchy-based [7], etc.

In the field of spatial clustering, some methods only handle 2-dimensional spatial attributes [8] and deal with problems like obstacles which are unique in clustering geo-spatial data [9]. To incorporate spatial constraints, the simplest method is to directly add spatial information, e.g., spatial coordinates, into datasets [10]. Others achieve this goal by modifying existing algorithms, e.g., allowing an object assigned to a class if this class already contains its neighbor [11]. Another class, where our algorithm falls, selects a model that encompasses spatial information. This can be achieved by modifying a criterion function that includes spatial constraints [12], which mainly comes from image analysis where MRF and EM-style algorithms were intensively used [13,14].

Clustering using mixture models with conventional EM does not account for spatial information. NEM extends EM by adding a weighted spatial penalty term in the objective function. The clustering performance of NEM depends largely on the global fixed coefficient, the weight of the penalty. If further information about structure is available, spatially varying coefficient models can be employed, which has been mainly investigated for regression problems [15].

In practice, the subsequent cluster refinement in NEM is only a sub-task of the whole clustering, which succeeds the execution of a certain initialization method. With the initialization methods returning a set of seed centers, the data are assigned to the closest center and thus an initial clustering is obtained for NEM to refine. Roughly speaking, the cluster initialization methods fall into three

major families: random sampling, distance optimization and density estimation. We will examine three representative methods in more details later.

3 The ANEMI Algorithm

In this section, we first introduce the basics of NEM from the MRF perspective. Then we present the cluster refinement part of the ANEMI algorithm, which exploits an adaptive scheme of coefficient assignment. Finally, we discuss the initialization methods for the ANEMI algorithm.

3.1 The MRF Framework

By the Hammersley-Clifford theorem [16], the prior probability of a cluster label configuration $Y = \{y_i\}_{i=1}^n$ (a realization of the MRF) can be expressed as a Gibbs distribution [13], $P(Y) = \frac{1}{Z_1} \exp(-V(Y))$, where Z_1 is a normalizing constant, and $V(Y)$ is the overall label configuration potential function. In the clustering framework, the conditional probability of X given Y has the form $P(X|Y) = f(X|Y, \Phi)$, a density function parameterized with Φ . The posterior probability becomes $P(Y|X) = \frac{1}{Z_2} P(Y)P(X|Y)$, where $Z_2 = Z_1 P(X)$ is the normalizing constant. Hence finding the maximum a-posteriori (MAP) configuration of the hidden MRF is equivalent to maximizing the logarithm of $P(Y|X)$ (scaled by a constant)

$$U = \ln(f(X|Y, \Phi)) - V(Y) \quad (1)$$

3.2 Neighborhood EM (NEM)

For the potential function $V(Y)$, NEM employs a soft version of the pairwise Potts model, $V(Y) = -\sum_{i,j} W_{ij} I(y_i = y_j)$, where $I(\cdot)$ is the indicator function with $I(\text{true}) = 1$ and $I(\text{false}) = 0$. In detail, let \bar{P} denote a set of distributions $\{\bar{P}_{ik} \equiv \bar{P}(y_i = k)\}$ governing $\{y_i\}$. Termed “spatial penalty”, the potential function used in NEM is $G(\bar{P}) = -\frac{1}{2} \sum_{i,j} W_{ij} \sum_{k=1}^K \bar{P}_{ik} \bar{P}_{jk}$. One can see it becomes the Potts model if we require \bar{P}_{ik} be binary (a hard distribution). Such a model favors spatially regular partitions, which is appropriate in the case of spatial positive autocorrelation.

In NEM, the conditional density $f(\mathbf{x}|\Phi)$ takes the form of a mixture model of K components $f(\mathbf{x}|\Phi) = \sum_{k=1}^K \pi_k f_k(\mathbf{x}|\theta_k)$, where $f_k(\mathbf{x}|\theta_k)$ is k -th component’s density function and $p(\mathbf{x}|y = k) = f_k(\mathbf{x}|\theta_k)$. Following [17], NEM’s soft counterpart of $\ln(f(X|Y, \Phi))$ in Eq. (1) can be written as

$$F(\bar{P}, \Phi) = \sum_{i=1}^n \sum_{k=1}^K \bar{P}_{ik} \ln(\pi_k f_k(\mathbf{x}|\theta_k)) - \sum_{i=1}^n \sum_{k=1}^K \bar{P}_{ik} \ln \bar{P}_{ik} \quad (2)$$

Note that maximizing F is also equivalent to maximizing the log-likelihood criterion function in the conventional mixture model [18]. Then, the new objective function in NEM becomes $U(\bar{P}, \Phi) = F(\bar{P}, \Phi) + \beta G(\bar{P})$, where β is a fixed positive

coefficient to weigh the spatial penalty and controls the desired smoothness of output clustering. U can be maximized via the EM procedure, starting from an initial \bar{P}^0 .

1. M-step: With \bar{P}^t fixed, set $\Phi^t = \operatorname{argmax}_{\Phi} U(\bar{P}^t, \Phi)$, which is exactly the same as the M-step in the standard EM for mixture models, for G dose not depend on Φ .
2. E-step: With Φ^t fixed, set $\bar{P}^{t+1} = \operatorname{argmax}_{\bar{P}} U(\bar{P}, \Phi^t)$.

It can be shown that in the E-step, U will be maximized at \bar{P}^* that satisfies Eq. (3), which can be organized as $\bar{P}^* = O(\bar{P}^*)$. It was proven in [3] that under certain conditions, the sequence produced by $\bar{P}^m = O(\bar{P}^{m-1})$ will converge to that solution to maximize U . Hence \bar{P}_{ik}^* can be regarded as dot product between the estimation from its own \mathbf{x} and the estimation from its neighbors.

$$\bar{P}_{ik}^* = \frac{\pi_k f_k(\mathbf{x}_i | \theta_k) \exp \left(\beta \sum_{j=1}^n W_{ij} \bar{P}_{jk}^* \right)}{\sum_{l=1}^K \pi_l f_l(\mathbf{x}_i | \theta_l) \exp \left(\beta \sum_{j=1}^n W_{ij} \bar{P}_{jl}^* \right)} \quad (3)$$

3.3 NEM with Adaptive Coefficient Assignment

EM for the conventional mixture model is not appropriate for spatial clustering since it does not account for spatial information. In contrast, NEM adds in the criterion a spatial penalty weighted by a fixed coefficient β . However, it may not be appropriate to assign a constant coefficient to every site. For those in the class interior, the whole neighborhood is from the same class and hence the site should receive more influence from its neighbors, especially when their posterior estimates are accurate. For those on the class border, because their neighbors are from different classes, its own class membership should be determined mainly by its own explanatory attributes.

Along this line, ANEMI employs a site-sensitive spatial coefficient for the spatial penalty term. In detail, besides the original β that determines the global smoothness in the solution clustering, every site s_i has another coefficient β_i of its own that determines the local smoothness. Then the new penalty becomes $G(\bar{P}) = \frac{1}{2} \sum_{i=1}^n \beta_i \sum_{j=1}^n W_{ij} \sum_{k=1}^K \bar{P}_{ik} \bar{P}_{jk}$. The original G in NEM can be regarded as a special case with $\beta_i = 1$ for all sites. Let U' denote the Lagrangian of U : $U' = U + \sum_{i=1}^n \lambda_i (\sum_{k=1}^K \bar{P}_{ik} - 1)$, which takes into account the constraints on \bar{P}_{ik} . Based on the necessary optimality Kuhn-Tucker conditions, solving $\partial U' / \partial \bar{P}_{ik} = 0$ for \bar{P}_{ik} yields

$$\bar{P}_{ik}^* = \frac{\pi_k f_k(\mathbf{x}_i | \theta_k) \exp \left(\beta \sum_{j=1}^n W_{ij} \frac{\beta_i + \beta_j}{2} \bar{P}_{jk}^* \right)}{\sum_{l=1}^K \pi_l f_l(\mathbf{x}_i | \theta_l) \exp \left(\beta \sum_{j=1}^n W_{ij} \frac{\beta_i + \beta_l}{2} \bar{P}_{jl}^* \right)} \quad (4)$$

Then the estimation is the same as that in NEM, except that we apply Eq. (4) in the E-step.

What remains is to determine β_i . In our implementation, we employ the local Moran's I measure, which evaluates the local spatial autocorrelation at site s_i based on the explanatory attributes inside the neighborhood [19]. Let z_{ip} denote the normalized p -th attribute of site s_i , i.e., $z_{ip} = x_{ip} - \bar{x}_p$, where \bar{x}_p is the global mean of the p -th attribute. Let σ_p denote the global standard deviation of the p -th attribute. Then, for the p -th attribute at site s_i , the local I is defined as $I_{ip} = \frac{z_{ip}}{\sigma_p^2} \sum_j W_{ij} z_{jp}$, where W is a row-normalized (sum to 1) version of the original binary W . A high I (e.g., $I > 1$) implies a high local spatial autocorrelation at site s_i , which is likely to occur in the class interior. The reverse happens on the border. In ANEMI, β_i is obtained by first averaging I_{ip} over all attributes and then normalizing to $[0, 1]$, i.e., $I_i = \text{mean}_p(I_{ip})$, $\beta_i = \frac{I_i - \min_i\{I_i\}}{\max_i\{I_i\} - \min_i\{I_i\}}$.

3.4 Spatial Augmented Initialization

Like other EM-based algorithms, ANEMI's clustering solution is sensitive to the initial state and hence the study of proper initialization is another focus of this paper. In this paper, we examined three representative methods for clustering initialization: random sampling, K-Means and KKZ. The random sampling method returns K seed centers by uniformly selecting K input instances. For within-cluster scatter minimization, the K-Means algorithm [20] can be regarded as a simplified hard version of EM on Gaussian mixture. While many clustering methods essentially minimize the within-cluster scatter, KKZ [21] is a greedy search method to optimize the complementary between-cluster scatter.

We can see that all initialization methods above only consider normal attributes without accounting for spatial information. If the positive autocorrelation is the major trend within data, then most sites would be surrounded by neighbors from the same class. Based on this observation, we propose to augment feature vector \mathbf{x}_i of site s_i with \mathbf{x}_{Ni} , the average of its neighbors. That is, the augmented vector becomes $\mathbf{x}'_i = [\mathbf{x}_i, \alpha \mathbf{x}_{Ni}]$, where $\alpha > 0$ is a coefficient to weigh the impact of the neighbors, and $\mathbf{x}_{Ni} = \sum_{j=1}^n W_{ij} \mathbf{x}_j / \sum_{j=1}^n W_{ij}$. Then the initialization methods can be run on the augmented $\{\mathbf{x}'_i\}$.

4 Experimental Evaluation

In this section, we first introduce the datasets and the clustering comparison methodology used in our experiments. Then we report comparative results.

4.1 Experimental Datasets

We evaluate ANEMI on five datasets, two synthetic and three real. Some data characteristics are listed in Table 1. The last row gives the spatial smoothness of the target variable y measured with contiguity ratio [19].

The synthetic image datasets are generated in the following way: First, a partition in four classes is simulated from a Potts MRF model with four-neighbor context on a 20×20 rectangular grid. Then, the observations are simulated from this partition based on four Gaussian densities. Fig. 1 shows two sample

Table 1. Some data characteristics

Data	Im1	Im2	Satimage	House	Election
size	400	400	4416	506	3107
#attribute	1	1	4	12	3
#class	4	4	6	n/a	n/a
ratio	0.78	0.84	0.96	0.58	0.61

partitions Im1 and Im2 of different smoothness, together with their observations. The observations for both partitions are drawn from four Gaussian densities: $N(0, 0.5^2)$, $N(1, 0.5^2)$, $N(2, 0.8^2)$, $N(4, 0.8^2)$.

Satimage is a real landcover dataset available at the UCI repository [22]. It consists of the four multi-spectral values of pixels in a satellite image together with the class label from a six soil type set. Because the dataset is given in random order, we synthesize their spatial coordinates and allocate them in a 64×69 grid to yield a high contiguity ratio of 0.96 with four-neighbor context. Fig. 2 illustrates the original partition and a sample obtained partition.

The House dataset records house prices and their environment indices of 506 towns in Boston area [23]. The 12 explanatory variables, such as nitric oxides concentration and crime rate, are used to predict the median value of houses, which is expected to have a small spread in each cluster of a reasonable partition. Fig. 3(a) and (b) show the true house values of 506 towns and their histogram. After normalizing the data to zero mean and unit variance, we fit two Gaussian mixtures, one with two components, the other with four components.

The Election dataset [23] records 1980 US presidential election results of 3107 counties. Originally the three attributes, fraction of population with college degree, fraction of population with homeownership and income, are used to predict voting rate. Here voting rate is used to evaluate clustering performance. Fig. 4(a) and (b) show the voting rates and their histogram. Again, we normalize the data and test two Gaussian mixtures with two and four components respectively.

4.2 Comparison Methodology

We evaluate the clustering quality via two external validation measures. Let C, Y denote the true class label and the derived cluster label, respectively. The conditional entropy $H(C|Y)$ is defined as $H(C|Y) = -\sum_{k=1}^K P_Y(k)H(C|Y=k)$, where probabilities are computed as sample frequencies. Analogously, for the continuous target variable C , we calculate the weighted standard deviation defined as $S(C|Y) = \sum_{k=1}^K P_Y(k)\text{std}(C|Y=k)$, where $\text{std}(\cdot)$ denotes the standard deviation operator and $(C|Y=k)$ denotes the C 's values in cluster $Y=k$. Both measures are minimized to zero in the ideal case.

During experimentation, we concentrate on whether spatial augmented initialization and adaptive coefficient assignment bring any gain in the final clustering quality. For fair comparison, we first compute the augmented version of vectors and then randomly draw K vectors. They are treated as the initial centers returned by the random sampling method on the augmented data. The first half

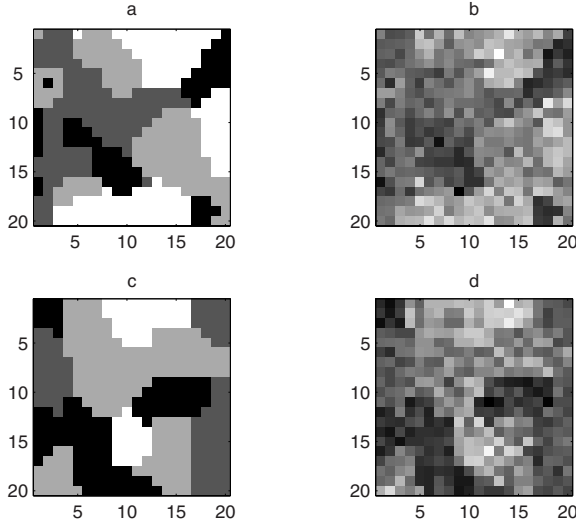


Fig. 1. (a) and (b) show Im1's true partition and observations, respectively. The counterparts of Im2 are shown in (c) and (d).

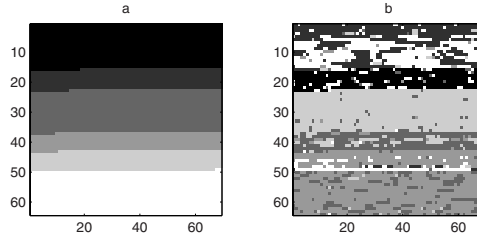


Fig. 2. (a) and (b) show Satimage's true partitions and a sample clustering

of these vectors are treated as those on original data. These vectors also play the role of initial centers for K-Means that runs 10 iterations. Euclidean distance is used in K-Means and KKZ. In the augmented vector $\mathbf{x}'_i = [\mathbf{x}_i, \alpha \mathbf{x}_{Ni}]$, often $\alpha = 1$ led to the best results, so we only report results with $\alpha = 1$.

The cluster refinement part of ANEMI is built upon an implementation of NEM [24]. Specifically, Gaussian mixture is employed as the model. The global coefficient β is tuned empirically in NEM for each dataset and the obtained value is also used in ANEMI. The number of internal iterations of E-step is set to 10. The outer iteration is stopped when $|(U^t - U^{t-1})/U^t| < 0.0001$. Finally, for each dataset, we report average results of 20 runs.

4.3 Results and Discussions

Initialization. First, we run conventional NEM initialized in both original and augmented spaces. The clustering results are given in Table 2, where the best

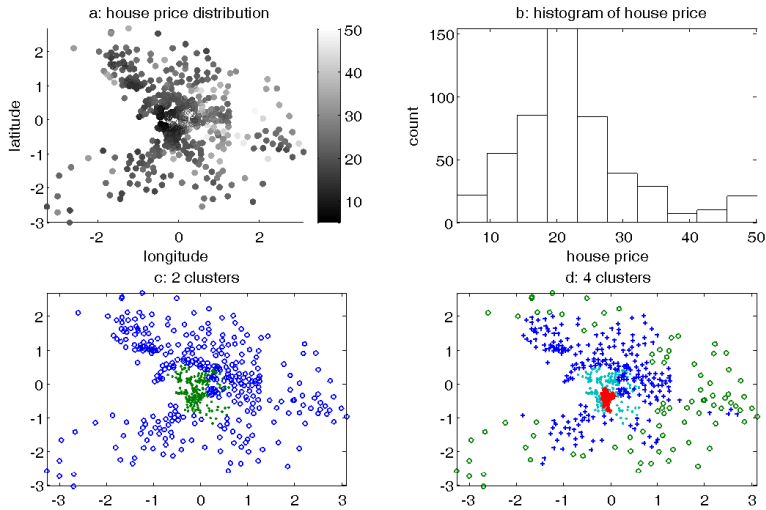


Fig. 3. (a) shows house price distribution in 506 towns in Boston. The corresponding histogram is plotted in (b). Two sample clustering results are shown in (c) and (d).

Table 2. Comparison with different initializations

Data	Sup	Rand	A-Rand	KMeans	A-KMeans	KKZ	A-KKZ
Im1	0.8252	0.9333 \pm 0.0556	0.9027 \pm 0.0264	0.8306 \pm 0.0034	0.8145 \pm 0.0117	0.9578	0.8947
Im2	0.8045	0.9900 \pm 0.0949	0.9558 \pm 0.0874	0.7405 \pm 0.0014	0.7540 \pm 0.0088	0.9362	0.7834
Satimage	0.6278	0.5970 \pm 0.0453	0.6243 \pm 0.0516	0.5170 \pm 0.0344	0.5142 \pm 0.0273	0.8468	0.8630
House:2	n/a	8.0918 \pm 0.0230	8.0011 \pm 0.0282	8.0633 \pm 0.0001	8.0624 \pm 0.0001	8.0632	8.0642
House:4	n/a	7.7830 \pm 0.1427	7.7768 \pm 0.2683	7.8401 \pm 0.0806	7.8170 \pm 0.0496	7.8145	7.8145
Election:2	n/a	0.1088 \pm 0.0023	0.1011 \pm 0.0024	0.0968 \pm 0.0004	0.0965 \pm 0.0001	0.1077	0.1077
Election:4	n/a	0.0992 \pm 0.0018	0.0953 \pm 0.0023	0.0927 \pm 0.0005	0.0919 \pm 0.0004	0.1007	0.0962

results are in boldface. “A-X” means initialization method “X” on the augmented data. For instance, the 3rd and 4th columns show the results with random initialization and augmented random initialization. For the three datasets with discrete target variables, we also list in the column “Sup” the results under supervised mode where each component’s parameters are estimated with all data from a single true class. One can see that initialization using augmented data generally brings improvement. The only exception is Satimage with random sampling and KKZ, possibly because its contiguity ratio is so high that almost every site is surrounded by sites from the same class with very similar observations. Thus using augmented data does not make much a difference to the initialization results. Among the three initialization methods, augmented K-Means always leads to the best or sub-optimal results, though the improvement of augmented versions is often more obvious with random sampling and KKZ.

Coefficient Assignment. Since K-Means generally provides the best initialization, we use it to initialize the mixture model for the subsequent comparison of spatial coefficient assignments. Fig. 5 presents the results corresponding to

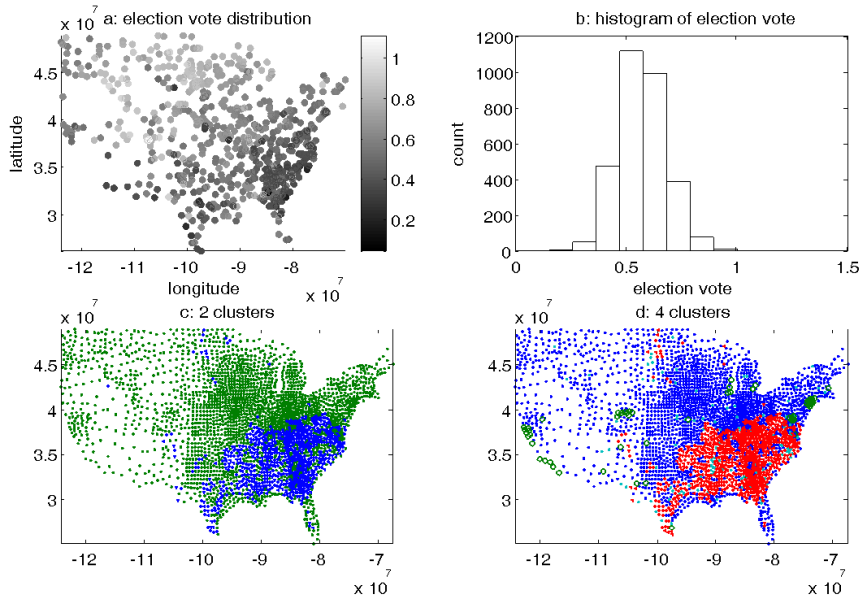


Fig. 4. (a) shows election voting rate distribution in 3107 counties. The corresponding histogram is plotted in (b). Two sample clustering results are shown in (c) and (d).

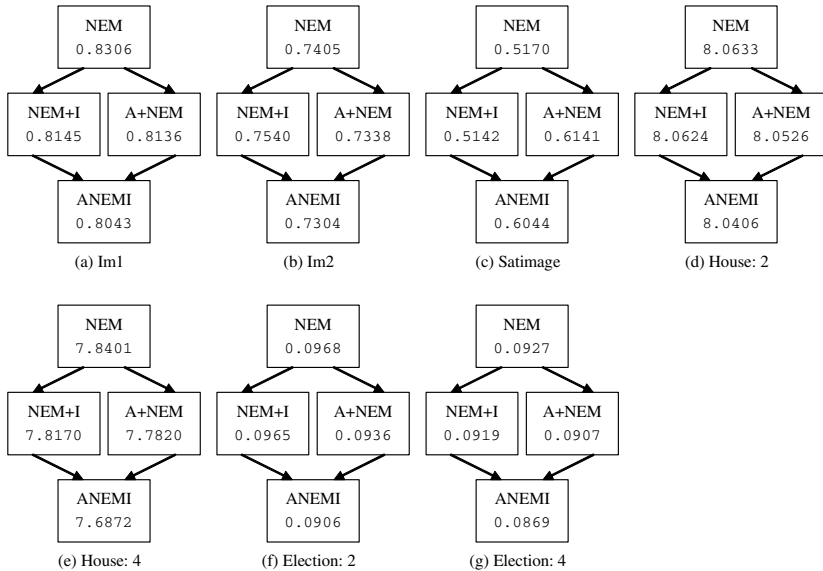


Fig. 5. Impact of spatial augmented initialization and adaptive coefficient assignment

different component combinations: “NEM” denotes conventional NEM with K-Means initialization in the original space; “NEM+I” denotes conventional NEM

with K-Means initialization in the augmented space; “A+NEM” denotes NEM with adaptive coefficient and K-Means initialization in the original space; “ANEMI” denotes NEM with adaptive coefficient and K-Means initialization in the augmented space. One can see that compared to conventional NEM, using site-sensitive coefficients generally yields better results. The only exception is Satimage again. The reasons may be that Satimage’s contiguity ratio is so high that almost every site is surrounded by sites from the same class. Thus it may be always beneficial to assign sites in the neighborhood to the same class. Compared to spatial augmented initialization, the adaptive coefficient assignment has a greater impact on the final clustering quality. The best results are always achieved by combining them two together.

5 Conclusions

In this paper, we introduced an Adaptive Neighborhood Expectation–Maximization with spatial augmented Initialization (ANEMI) algorithm for spatial clustering. ANEMI is an extension of the NEM algorithm, which is built on top of the EM algorithm by incorporating a spatial penalty term into the criterion function. This penalty term is weighed by a spatial coefficient that determines the global smoothness of the solution clustering. Unlike NEM, which assigns an equal weight to every site, ANEMI exploits an adaptive site-sensitive weight assignment scheme, which is determined by the local smoothness inside the neighborhood for each site. In addition, to provide a good initial state for clustering, we proposed to push spatial information early into the initialization methods. Along this line, we also examined three representative initialization methods in the spatial augmented space. Finally, we evaluated the impact of spatial augmented initialization and adaptive coefficient assignment in ANEMI against NEM on both synthetic and real-world datasets. Empirical results showed that with adaptive coefficient assignment, ANEMI using augmented K-Means initialization generally leads to better clustering results than NEM. The gain is most obvious when they are run on datasets with low contiguity ratio.

As for future work, we plan to investigate stochastic versions of NEM to reduce dependence on the algorithm initialization. Also, other optimization techniques, such as genetic algorithms [25], are worth trying to speed up the convergence rate and to improve the final clustering quality further.

Acknowledgments. This work was partially supported by SRF for ROCS, Sci. Tech. Plan Foundation of Guangdong (No. 20070328005), and Sci. Tech. Plan Foundation of Dongguan (No. 2007108101022).

References

1. Tobler, W.R.: Cellular Geography, Philosophy in Geography. In: Gale, W.R., Olson, W.R. (eds.) Reidel, The Netherlands (1979)
2. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Statistical Society B*(39), 1–38 (1977)

3. Ambroise, C., Govaert, G.: Convergence of an EM-type algorithm for spatial clustering. *Pattern Recognition Letters* 19(10), 919–927 (1998)
4. Garey, M.R., Johnson, D.S., Witsenhausen, H.S.: The complexity of the generalized lloyd-max problem. *TOIT* 28(2), 255–256 (1980)
5. Ng, R., Han, J.: CLARANS: A method for clustering objects for spatial data mining. *TKDE* 14(5), 1003–1016 (2002)
6. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: *KDD*, pp. 226–231 (1996)
7. Karypis, G., Han, E.H., Kumar, V.: CHAMELEON: A hierarchical clustering algorithm using dynamic modeling. *Computer* 32(8), 68–75 (1999)
8. Estivill-Castro, V., Lee, I.: Fast spatial clustering with different metrics and in the presence of obstacles. In: *ACM GIS*. (2001) 142 – 147
9. Tung, A.K.H., Hou, J., Han, J.: Spatial clustering in the presence of obstacles. In: *ICDE*, pp. 359–367 (2001)
10. Guo, D., Peuquet, D., Gahegan, M.: Opening the black box: Interactive hierarchical clustering for multivariate spatial patterns. In: *ACM GIS*, pp. 131–136 (2002)
11. Legendre, P.: Constrained clustering. In: Legendre, P., Legendre, L. (eds.) *Developments in Numerical Ecology*. NATO ASI Series G 14, pp. 289–307 (1987)
12. Rasson, J.P., Granville, V.: Multivariate discriminant analysis and maximum penalized likelihood density estimation. *J. Royal Statistical Society B*(57), 501–517 (1995)
13. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *TPAMI* 6, 721–741 (1984)
14. Solberg, A.H., Taxt, T., Jain, A.K.: A markov random field model for classification of multisource satellite imagery. *IEEE Trans. Geoscience and Remote Sensing* 34(1), 100–113 (1996)
15. Congdon, P.: A model for non-parametric spatially varying regression effects. *Computational Statistics & Data Analysis* 50(2), 422–445 (2006)
16. Hammersley, J.M., Clifford, P.: Markov fields on finite graphs and lattices, unpublished manuscript (1971)
17. Neal, R., Hinton, G.: A view of the EM algorithm that justifies incremental, sparse, and other variants. In: Jordan, M. (ed.) *Learning in Graphical Models*, pp. 355–368. Kluwer Academic Publishers, Dordrecht (1998)
18. Hathaway, R.J.: Another interpretation of the EM algorithm for mixture distributions. *Statistics and Probability Letters* 4, 53–56 (1986)
19. Shekhar, S., Chawla, S.: *Spatial Databases: A Tour*. Prentice-Hall, Englewood Cliffs (2002)
20. Tou, J.T., Gonzalez, R.C.: *Pattern Recognition Principles*. Addison-Wesley, Reading (1974)
21. Katsavounidis, I., Kuo, C., Zhang, Z.: A new initialization technique for generalized lloyd iteration. *IEEE Signal Processing Letters* 1(10), 144–146 (1994)
22. Newman, D.J., Hettich, S., Blake, C.L., Merz, C.J.: *UCI repository of machine learning databases* (1998), <http://www.ics.uci.edu/~mllearn/MLRepository.html>
23. LeSage, J.P.: *MATLAB Toolbox for Spatial Econometrics* (1999), <http://www.spatial-econometrics.com>
24. Dang, V.M.: (1998), <http://www.hds.utc.fr/~mdang/Progs/prognem.html>
25. Pernkopf, F., Bouchaffra, D.: Genetic-based EM algorithm for learning gaussian mixture models. *TPAMI* 27(8), 1344–1348 (2005)