Asian Journal of Information Technology 7 (8): 388-393, 2008
ISSN: 1682-3915

Angle Decrement Based Gaussian Kernel Width Generator for Support Vector Clustering

M. Rahmat Widyanto and Herman Hartono
Faculty of Computer Science, University of Indonesia, Kampus UI Depok, 16424, Indonesia

Abstract: A new method to generate Gaussian kernel width parameter (q) for Support Vector Clustering (SVC) is proposed in this study. The proposed method is based on idea of decreasing angle, along with increment of q. This method is a modification of secant method that previously proposed. Experiments are performed using four sets of data, each data set has its own characteristics. Experimental results show that angle decrement based method can generates a valid sequence of q value with simpler computation than secant method. In general, angle decrement based method can improve the performance of SVC so that clustering process can be performed faster.

Key words: Clustering, support vector clustering, angle decrement based, gaussian kernel width

INTRODUCTION

Clustering is a method to divide a set of data into some subset, so that each subset of data shares some common characteristics. Data which are belong to same cluster are more similar than other data that belong to different cluster (Bishop, 2006). Clustering has been used widely in many fields, such as in bioinformatics, pattern recognition and image analysis. There is some categories of clustering algorithm, there are hierarchical (e.g., BIRCH (Zhang *et al*., 1996), CURE (Guha *et al*., 1998), Chameleon (Karypsis *et al*., 1999)), density-based (e.g., DBSCAN (Ester *et al*., 1996), OPTICS (Ankerst *et al*., 1999)), grid-based (e.g., STING (Wang *et al*., 1997), model-based, (e.g., COWEB (Fisher, 1987)) and boundary-detecting (e.g., SVC (Ben-Hur *et al*., 2001)). In this study, we will focus on Support Vector Clustering (SVC), which is proposed in (Ben-Hur *et al*., 2001, 2000).

SVC uses the idea of support vector to cluster the data. In SVC data points are mapped from data space to a high dimensional feature space. The mapping process is done using the Gaussian Kernel (Ben-Hur *et al*., 2000). In feature space, we find a minimal sphere that enclosing the feature space images of data points. One of the most important parameter in SVC, is the gaussian kernel width parameter (q) (Lee and Daniels, 2005). This parameter influences the number of clusters produced. The value of this q parameter is difficult to determine. The value of q is different for each of dataset; it is depend to the characteristic of dataset that will be processed. A secant-like method to generate sequence of q value is proposed in Lee and Daniels (2005) and (2004). This secant-like method is based on calculation of radius for each q value. The calculation of radius is a complex computation, so, the calculation of q value with this secant-like method is also complex. For overall SVC algorithm, secant-like method is increase the complexity of SVC algorithm.

New method for generating sequence of q values is proposed in this study. This method is based on the idea of angle decrement. The proposed method is a modification of secant-like method, so that the calculation of radius for each q value is not needed. This method has simpler calculation than the secant-like method. With this angle decrement based method, generation of q value can be generated faster than using secant-like method. The angle decrement based method is implemented using MATLAB™. We use the four data sets to test the proposed method.

SUPPORT VECTOR CLUSTERING

SVC is a non-parametric clustering algorithm based on support vector approach of Support Vector Machine (SVM). SVM (Vapnik, 1995) is a learning algorithm that has been used widely, especially for data classification. First of all, we use Gaussian Kernel to map data points from data space to a high dimensional feature space. In feature space, we look for the smallest sphere that encloses the images of the data. Then, the enclosing sphere is mapped back to data space. In the data space,

Corresponding Author: M. Rahmat Widyanto, Faculty of Computer Science, University of Indonesia, Kampus UI Depok, 16424, Indonesia

the sphere forms a set of contours, which encloses the data points. These contours are interpreted as cluster boundaries. Points enclosed by each separate contour are associated with the same cluster.

The number of disconnected contours in data space increase, leading to an increasing number of cluster, if the width parameter of the Gaussian Kernel is decreased. SVC can deal with outliers by employing a soft margin constant that allows the sphere in feature space not to enclose all points. For large value of this soft margin constant, we can also deal with overlapping clusters.

Let $\{x_i\} \subseteq X$ be a data set of N points, by using a nonlinear transformation Φ from X to a high dimensional feature space, we look for the smallest enclosing sphere of radius R. this sphere is described by the constraints:

$$\|\Phi(x_j) - a\|^2 \leq R^2, \forall_j \quad (1)$$

where,

$\| . \|$ = The Euclidean norm.

a = The center of the sphere.

Soft constraints are incorporated by adding slack variable, so the constraint is,

$$\|\Phi(x_j) - a\|^2 \leq R^2 + \xi_j, \quad \text{with } \xi j \geq 0 \quad (2)$$

To solve the problem, we introduce the Lagrangian

$$L = R^2 - \sum (R^2 + \xi_j - \|\Phi(x_j) - a\|^2)\beta_j - \sum \xi_j \mu_j + C\sum \xi_j \quad (3)$$

where, $\beta_j \geq 0$ and $\mu_j \geq 0$ are Lagrange multipliers, C is a constant and $C\Sigma\xi_j$ is a penalty term. Setting to zero the derivative of L with respect to R, a and ξ_j, respectively, leads to,

$$\sum_j \beta = 1 \quad (4)$$

$$a = \sum_j \beta \Phi(x_j) \quad (5)$$

$$\beta_j = C - \mu_j \in \quad (6)$$

The definition of Gaussian kernel used in this algorithm is ,

$$K(x_i, x_j) = e^{-q\|x_i - x_j\|^2} \quad (7)$$

with width parameter q. From the derivation of above formulas, we got that:

- Bounded support vector: if $\beta j = C$.
- Support vector: if $0 < \beta j < C$.
- Inner points: if $\beta j = 0$.

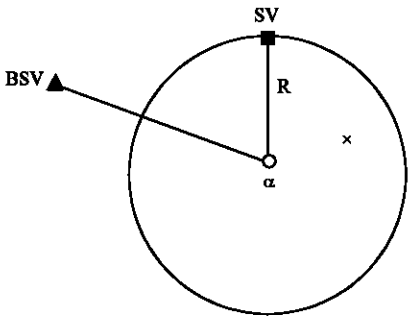

Fig. 1: BSV, SV and minimal sphere

In view of the constraints and the definition of kernels, we have,

$$R^2(x) = K(x,x) - 2\sum_j \beta_j K(x_j, x) + \sum_{i,j} \beta_i \beta_j K(x_i, x_j) \quad (8)$$

The radius of the sphere is $R = \{R(x_i) \mid x_i \text{ is a support vector}\}$. The counters that enclose the points in data space are defined by the data set $\{x \mid R(x) = R\}$. SVs lies on cluster boundaries, BSVs are outsides and all other points lie inside the cluster. Figure 1 illustrates three kinds of point in SVC. SVC can be divided into three major steps, the generation of kernel matrix, solving the quadratic programming to find the Lagrange multiplier and cluster labeling.

In SVC algorithm, the width parameter of Gaussian kernel (q) controls how 'spread out' the data points feature space images are and therefore, determines the size of the minimal sphere. Finding a small set of q values for a given dataset is an important part of SVC algorithm.

SECANT-LIKE METHOD

A secant-like numerical algorithm is proposed in Lee and Daniels (2004, 2005). It is based, on the intuition that significant changes in clustering are less likely to occur in q intervals where R^2 values are fairly stable. This relies on R_2 monotonic. Therefore, we characterizing R^2 as a function of q, this establishes that $R^2 = 0$ for $q = 0$, $R^2 = 1\text{-}1/N$ if $q = \infty$. It is assumed that the value of C is fixed so that the number of outlier is not varied. The number of q values generated by the secant-like algorithm is estimated using known result on secant method. The estimate relies on spatial characteristics of the dataset but not the number of data points or the dimensionality of the dataset. To characterizing R^2 as a function of q, we first observe that as R is the radius of the minimal sphere enclosing data points images, $R^2 \geq 0$ for $0 \leq q \leq \infty$. If $q = 0$, then $R^2 = 0$. These statements described:

$$\frac{1}{N} < C \leq 1 \quad (9)$$

$$q = 0 \text{ if and only if } R^2 = 0 \quad (10)$$

$$\text{If } q = \infty, \text{ then } \beta_i = 1/N, \text{ for all } I \in \{1, \ldots, N\} \quad (11)$$

$$\text{If } q = \infty, \text{ then } R^2(x_i) = 1\text{-}1/N \quad (12)$$

$$R^2 = 1 \text{ if and only if } q = \infty \text{ and } N = \infty \quad (13)$$

The operation of secant-like method is illustrated in Fig. 2. The starting q value for secant-like method is 0, the second q value is from (Ben-Hur *et al.*, 2001), which is:

$$q = \frac{1}{\max_{i,j} \left\| x_i - x_j \right\|^2} \quad (14)$$

this q value is expected to yields a result of one cluster. This secant-like method is based on theorem that R^2 for each value of q ($q<\infty$) is lower than 1-1/N. For each value of q, the associated R^2 value is calculated using the SVC steps of updating a kernel matrix, solving the Lagrangian and computing the radius of the minimal sphere. To generate each subsequent q value, a line through the two previous R^2 curve points is extended until it intersects the line $R^2 = 1\text{-}1/N$. the secant-like algorithm terminates when every data point is an SV or the slope of the line is close to flat. When every data point is an SV, the number of clusters is typically N and no useful clustering information is usually gained for larger q values.

The secant-like method for generating sequence of q values, described in previous section has a major disadvantage. For each of q value, except for q_0 and q_1, the value of R_2 must be calculated. After the value of current R_2 is calculated, find a line that connects previous R_2 with current R_2 and then find the intersection point between this line and 1-1/N. For each of q value, the kernel matrix has to be recalculated and the Lagrangian has to be solved.

The calculation of R^2 is a complex computation, finding a line that connects previous R^2 with current R^2 and finding the intersection point between this line and 1-1/N is also complex computations. It can be seen that the secant-like method increases the complexity of SVC algorithm.

Along with the increase of q, the value of R^2 will also increase. In the high dimensional feature space, it means that the radius of enclosing sphere is increase and the enclosing sphere is become larger. Figure 3 illustrates the expansion of enclosing sphere. As the enclosing sphere expands, the number of BSV is tend to decrease. Some

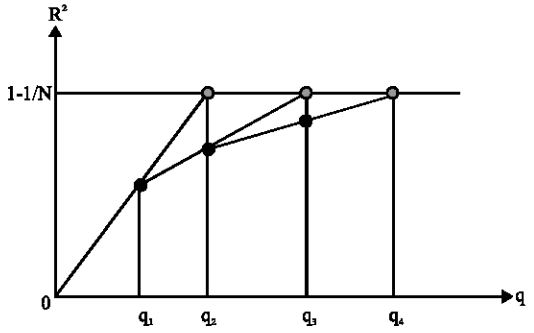

Fig. 2: Secant-like method

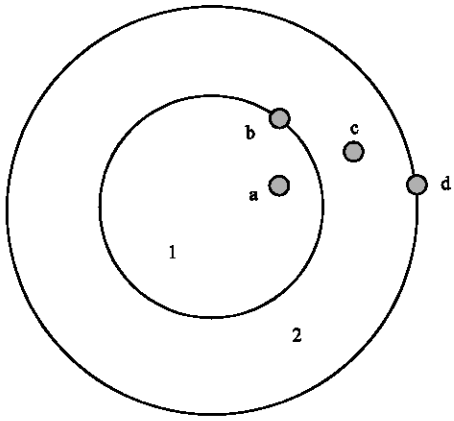

Fig. 3: Increase of enclosing sphere's radius

BSVs will become either inner points or SVs. Figure 3 shows that the enclosing sphere expanded from sphere 1 to sphere 2. Initially, point c and d are BSV, after the sphere expands, point c becomes an inner point and point d becomes a SV.

Proposed Angle Decrement Based Method: Another method to generate sequence of q values is proposed in this study. This method is modified from secant-like method described above. The main idea of this method is based on the fact of the decreasing angle in secant-like method. With this angle-based method, the complexity of computation on SVC can be reduced. Figure 2 shows that the value of R^2 is increasing, along with the increment of q. The lines connecting origin point (0, 0) and sequence of R^2 values will become sloppier, slope of these lines become smaller. We can use this condition to generate sequence of q value. From Fig. 4 shows the decreasing angle with increasing q values. We can see that $\theta_1 > \theta_2 > \theta_3 \ldots$ and $\tan\theta_1 > \tan\theta_2 > \tan\theta_3 > \ldots$. We got that,

$$q_n = \frac{1 - 1/N}{\tan\theta_n} \quad (15)$$

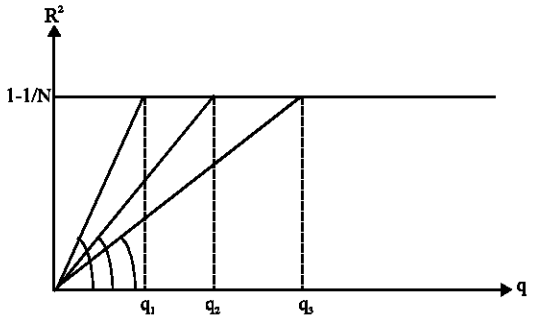

Fig. 4: Basic angle decrement based method

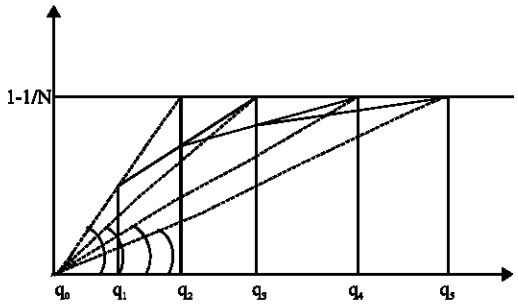

Fig. 5: Angle decrement based method with intersection points

To obtain sequence of q values that approximate the q values produced by secant-like method, we perform angle decrement based method to the intersection points between R^2 max and q. Figure 5 shows the angle decrement based method using the intersection points of secant-like method.

From Fig. 5, we can see that the value of tan θ can be obtained using Eq. (15), for n >2. For n = 1 and n = 2,

$$\tan\theta_1 = \tan\theta_2 = \frac{1-1/N}{q_2} \qquad (16)$$

From implementation of secant-like method to some data sets, we observed that the current value of tan θ is about half of previous tan θ. From this observation we set a hypothesis that $\tan\theta_n \approx 0.5\tan\theta_{n-1}$, for n>2, then we got

$$q_n = \frac{1-1/N}{\tan\theta_n} \approx \frac{1-1/N}{0.5\tan\theta_n}. \qquad (17)$$

EXPERIMENT AND ANALYSIS

To compare the performance of angle decrement based method and secant-like method, we implemented both methods with Matlab. Both methods are tested with 4 datasets, 2 datasets are 2-dimensional data and the others are high-dimensional data. The first dataset is a linearly separable 2-dimensional data, the second dataset is nonlinearly separable 2-dimensional data, third dataset is a 4-dimensional Iris dataset and the last dataset is image-segmentation dataset with 19 dimensions. Both Iris and Image-Segmentation dataset are obtained from UCI machine learning repository (Blake and Merz, 1998) and both 2-dimensional datasets are generated, for experimental purpose.

Table 1: Linearly separable 2-dimensional data with secant-like method

Iterations	q	Nsv
1	0.001156	4
2	0.003488	5
3	0.008149	9
4	0.017008	14
5	0.034198	18
6	0.073686	33

Table 2: Linearly separable 2-dimensional data with angle decrement based method

Iterations	q	Nsv
1	0.001156	4
2	0.003488	5
3	0.006970	9
4	0.013950	12
5	0.027900	17
6	0.055800	28

Table 3: Non-linearly separable 2-dimensional data with secant-like method

Iterations	q	Nsv
1	0.00590	8
2	0.01690	8
3	0.03200	8
4	0.06346	12
5	0.11830	26
6	0.21540	48

Table 4: Non-linearly separable 2-dimensional data with angle decrement based method

Iterations	q	Nsv
1	0.00590	8
2	0.01690	8
3	0.03374	10
4	0.06750	16
5	0.13500	38
6	0.26990	48

Table 1 shows the experiment result from linearly separable 2-dimensional data with secant-like method. Table 2 shows the result of angle decrement based method. Both methods need 6 iterations to produce good clustering result. Both methods produced same clustering result, although the numbers of support vector (Nsv) are not equal. Processing time for angle decrement based method is 18.65 sec, better than processing time of secant-like method (23.26 sec).

Table 3 and 4 show the experiment result from nonlinearly separable 2-dimensional data with secant-like method and angle decrement based method, respectively. Both methods need 6 iterations to produce good clustering result. Processing time for angle decrement based method is 1.65 sec, better than processing time of secant-like method, 1.88 sec.

Table 5: Iris data with secant-like method

Iterations	q	Nsv
1	0.02	4
2	0.06	6
3	0.16	11
4	0.35	16
5	0.70	22
6	1.30	31
7	2.27	47
8	3.87	67

Table 6: Iris data with angle decrement based method

Iterations	q	Nsv
1	0.0200	4
2	0.0600	6
3	0.1248	9
4	0.2497	12
5	0.4994	18
6	0.9990	27
7	1.9975	43
8	3.9950	68

Table 7: Image-segmentation data with secant-like method

Iterations	q	Nsv
1	0.000000431	129
2	0.000001313	194
3	0.000002832	272
4	0.000005745	334
5	0.000010261	438
6	0.000020016	522
7	0.000040614	653
8	0.000083612	751
9	0.000175177	874

Table 8: Image-segmentation data with angle decrement based method

Iterations	q	Nsv
1	0.000000431	129
2	0.000001313	194
3	0.000002626	269
4	0.000005252	327
5	0.000010504	446
6	0.000021008	532
7	0.000042016	668
8	0.000084032	759
9	0.000168064	868

Experiment results from Iris dataset with secant-like method and angle decrement based method are shown in Table 5 and 6. Both methods need 8 iterations to produce good clustering result. Processing time for angle decrement based method is 23.51 sec, better than processing time of secant-like method, 25.85 sec.

For Image-segmentation dataset, the experiment results are shown in Table 7 and 8. Both methods need 9 iterations to produce good clustering result. Processing time for angle decrement based method is 8652.515 sec, better than processing time of secant-like method, 9508.173 sec.

From the experiment results we can see that the angle decrement based method can produce clustering result as good as secant-like method. Angle decrement based method can produce good clustering result with better processing time than secant-like method. This result happened for all datasets that we used in this experiment. We can see that both methods need same number of iterations for producing good clustering result, but angle decrement based method has better processing time because the calculation of q in each iteration of angle decrement based method is faster than calculation of q in secant-like method.

CONCLUSION

A new method to calculate Gaussian kernel width, angle decrement based method, is proposed in this study. The proposed method can produce good clustering result with better processing time than secant-like method. This proposed method improves the overall performance of Support Vector Clustering algorithm. With this angle decrement based method, SVC can produce good clustering result with better performance than before.

The SVC algorithm can cluster a non-linearly separable data, this is one of significant advantage of SVC over other clustering algorithm. Another advantage is the ability to cluster high-dimensional data. SVC can also handle outliers in a dataset. With all of these advantages we can see that SVC is a very good clustering algorithm.

For further research, more intensive experiments with various datasets are needed. Experiment with very large datasets and very high dimension is a good point to do in order to improve this experiment. Experiment with various datasets containing outliers is also needed.

REFERENCES

Ankerst, M., M.M. Breunig, H.P. Kriegel and J. Sander, 1999. Optics: Ordering points to identify the clustering structure. ACM SIGMOD'99 Int. Conf. Management Data.

Ben-Hur, A., D. Horn, H.T. Siegelmann and V. Vapnik, 2000. A Support Vector Clustering Method, International Conference on Pattern Recognition.

Ben-Hur, A., D. Horn, T. Siegelmann and V. Vapnik, 2001. Support vector clustering. J. Machine Learn. Res., 2: 125-137.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning, Springer.

Blake, C.L. and C.J. Merz, 1998. UCI Repository of Machine Learning Databases.

Ester, M., H.P. Kriegel, J. Sander and X. Xu, 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. Proc. 2nd Int. Conf. Knowledge Discovery and Data Mining, pp: 226-231.

Fisher, D.H., 1987. Knowledge acquisition via incremental conceptual clustering. Machine Learn., 2: 139-172.

Guha, S., R. Rastogi and K. Shim, 1998. Cure: An Efficient Data Clustering Method for Very Large Databases, ACM SIGMOD International Conference on Management of Data.

Karypsis, G., E.H. Han and V. Kumar, 1999. Chameleon: A hierarchical clustering algorithm using dynamic modeling. IEEE. Computer: Special Issue on Data Analysis and Mining, 32: 68-75.

Lee, S.H. and K. Daniels, 2004. Gaussian kernel width generator for support vector clustering. ICBA Proceedings.

Lee, S.H. and K. Daniels, 2005. Gaussian kernel width selection and fast cluster labeling for support vector clustering. Technical Report University of Massachusetts Lowell.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory, Springer.

Wang, W., J. Yang and R.R. Muntz, 1997. STING: A statistical information grid approach to spatial data mining. Twenty-Third International Conference on Very Large Data Bases.

Zhang, T., R. Ramakrishnan and M. Livny, 1996. BIRCH: An efficient data clustering method for very large databases. Proceedings of ACM SIGMOD.