

APPLICATION OF SUPPORT VECTOR MACHINE TO THE PREDICTION OF GEO-EFFECTIVE HALO CMES

SEONGHWAN CHOI^{1,2}, YONG-JAE MOON², NGO ANH VIEN³, AND YOUNG-DEUK PARK¹

¹ Korea Astronomy and Space Science Institute, Daejeon 305-348, Korea

E-mail : shchoi@kasi.re.kr, moonyj@khu.ac.kr

² Department of Astronomy and Space Science, Kyung Hee University,
Yongin 446-701, Korea

³ Institute for Artificial Intelligence, Ravensburg-Weingarten University of Applied Sciences,
Weingarten 88250, Germany

(Received December 30, 2011; Revised February 11, 2012; Accepted February 26, 2012)

ABSTRACT

In this study we apply Support Vector Machine (SVM) to the prediction of geo-effective halo coronal mass ejections (CMEs). The SVM, which is one of machine learning algorithms, is used for the purpose of classification and regression analysis. We use halo and partial halo CMEs from January 1996 to April 2010 in the SOHO/LASCO CME Catalog for training and prediction. And we also use their associated X-ray flare classes to identify front-side halo CMEs (stronger than B1 class), and the Dst index to determine geo-effective halo CMEs (stronger than -50 nT). The combinations of the speed and the angular width of CMEs, and their associated X-ray classes are used for input features of the SVM. We make an attempt to find the best model by using cross-validation which is processed by changing kernel functions of the SVM and their parameters. As a result we obtain statistical parameters for the best model by using the speed of CME and its associated X-ray flare class as input features of the SVM: Accuracy=0.66, POD_y=0.76, POD_n=0.49, FAR=0.72, Bias=1.06, CSI=0.59, TSS=0.25. The performance of the statistical parameters by applying the SVM is much better than those from the simple classifications based on constant classifiers.

Key words : space weather — CME — flare — machine learning — support vector machine

1. INTRODUCTION

Coronal mass ejections (CMEs), which are magnetized ejecta from the Sun into the heliosphere, are one of the strongest solar activities which are geo-effective. Earthward CMEs interact with the Earth's magnetic field and result in geomagnetic storms. There have been many studies on the relationship between CMEs and geomagnetic storms; front-side halo CME (Webb 2002; Gopalswamy et al. 2007; Zhang et al. 2007), the location of CMEs (Kim et al. 2005; Wang et al. 2002), the speed of CMEs (Srivastava & Venkatakrisnan 2004), and front-side direction parameter of CMEs (Moon et al. 2005; Kim et al. 2008).

Machine learning technology has been widely used in image processing, data classification, and pattern recognition in order to predict solar activities. Machine learning can automatically make a model from data, even in case that they are not clearly understood. Therefore machine learning technology has been employed for space weather applications in the following two aspects: space weather prediction (Al-Omari et al. 2010; Chen et al. 2010; Colak et al. 2009; Gavrischaka et al. 2001; He et al. 2008; Li et al. 2007; Liu

et al. 2011; Olmedo et al. 2005; Qahwaji et al. 2007, 2008; Yuan et al. 2011) and solar feature identification (Henwood et al. 2010; Labrosse et al. 2010; Quaalude et al. 2003, 2005; Martens et al. 2009).

In this study, we apply Support Vector Machine (SVM) to the prediction of geo-effective halo CMEs, which is the first attempt for the forecast of geomagnetic storms based on CME data. The SVM is one of machine learning algorithms, which is to automatically learn to recognize complex patterns and make intelligent decisions based on data. The speed and the angular width of CMEs, and their associated X-ray flare classes are used as input parameters for the SVM classification. In order to verify the SVM performance, we simply classify the CMEs according to each input parameter with simple constant classifiers.

This paper is organized as follows. Section 2 describes the data, the SVM, and statistical verification methods. In Section 3, we present the results of the SVM classifications by using cross-validation. And then we verify the results by comparing them with simple classifications with constant classifiers. Finally, a brief summary and conclusion are given in Section 4.

Corresponding Author: Y.-J. Moon

Table 1.
Data source list

Data	Data Center / URL
SOHO/LASCO CME Catalog	CDAW Data Center http://cdaw.gsfc.nasa.gov/CME_list/
GOES X-ray Flare	NOAA NGDC http://www.ngdc.noaa.gov/nndc/struts/results?t=102827&s=25&d=8,230,9 ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA/SOLAR_FLARES/FLARES_XRAY/
Dst Index	WDC for Geomagnetic in Kyoto http://wdc.kugi.kyoto-u.ac.jp/dstdir/

Table 2.
The number of CMEs from January 1996 to April 2010

Data Sets	Assumption	the Number of Data
CMEs	in the CME Catalog	15,076
Halo and Partial Halo CMEs	with angular width over 120°	1,282
Front-side CMEs	with X-ray flare from 2 hours ago	733
CMEs for training	from 1996 to 2002	627
CMEs for prediction	from 2003 to 2010	106

2. DATA AND METHOD

2.1 Data

We use the SOHO/LASCO CME Catalog, the GOES X-ray flare list, and the Dst index in order to predict geo-effective halo CMEs. The SOHO/LASCO CME Catalog is provided by ‘Coordinated Data Analysis Workshops (CDAW) Data Center’ via HTTP (http://cdaw.gsfc.nasa.gov/CME_list/). There are 15,076 CMEs from January 1996 to April 2010. We select 1,282 halo and partial halo CMEs whose angular widths are larger than 120°. In the CME catalog, we use CME appearance time, its angular width, and its linear speed. And then we extract 733 front-side CMEs by identifying if they have stronger X-ray flares than B1 X-ray class within two hours before the initial appearance times of the halo and partial halo CMEs. The GOES X-ray flare list is provided by ‘NOAA National Geophysical Data Center (NGDC)’ via HTTP and FTP (<http://www.ngdc.noaa.gov/nndc/struts/results?t=102827&s=25&d=8,230,9> and ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA/SOLAR_FLARES/FLARES_XRAY/). From this data set, we use the speeds and the angular widths of CMEs, and their associated X-ray flare classes as input parameters.

To determine their geo-effectiveness, we use Dst index data which are provided by ‘World Data Center (WDC) for Geomagnetic in Kyoto’ via HTTP

(<http://wdc.kugi.kyoto-u.ac.jp/dstdir/>). If Dst index is less than -50 nT within ± 24 hours from the CME arrival time based on the empirical CME propagation model (Gopalswamy et al. 2001), we consider this CME as a geo-effective CME. Fig. 1 summarizes the method of data selection. Table 1 shows the data centers and their URLs that provide the used data, and Table 2 shows the number of CMEs for the present study from January 1996 to April 2010.

In this paper, we do not use the source location of CMEs as an input parameter, even though the CME location is well known as one of important parameters to predict CME geo-effectiveness (Wang et al. 2002). There are 487 CMEs with the location information, which is almost half of the selected CMEs. As a future study, another model using the CME location would be developed.

We have developed several program codes to download the data from the data centers, to extract data records and input parameters, and to apply SVM to the prediction of geo-effective CMEs by using Python and C/C++ languages. Python is a general-purpose and high-level language. It provides open source programs that can be easily integrated with different platforms and languages. There are libraries like NumPy, SciPy and Matplotlib to allow Python to be used effectively in scientific computing. Recently NASA and ESA initiated the development of SunPy to create an open source project for solar physics using the Python

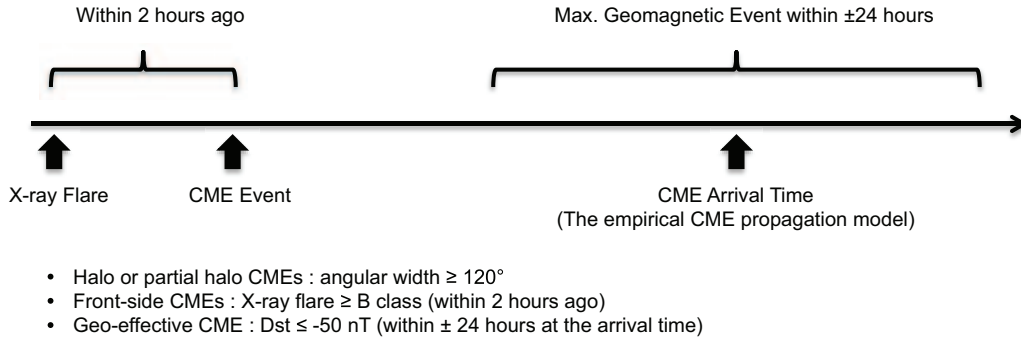


Fig. 1.— The method of data selection.

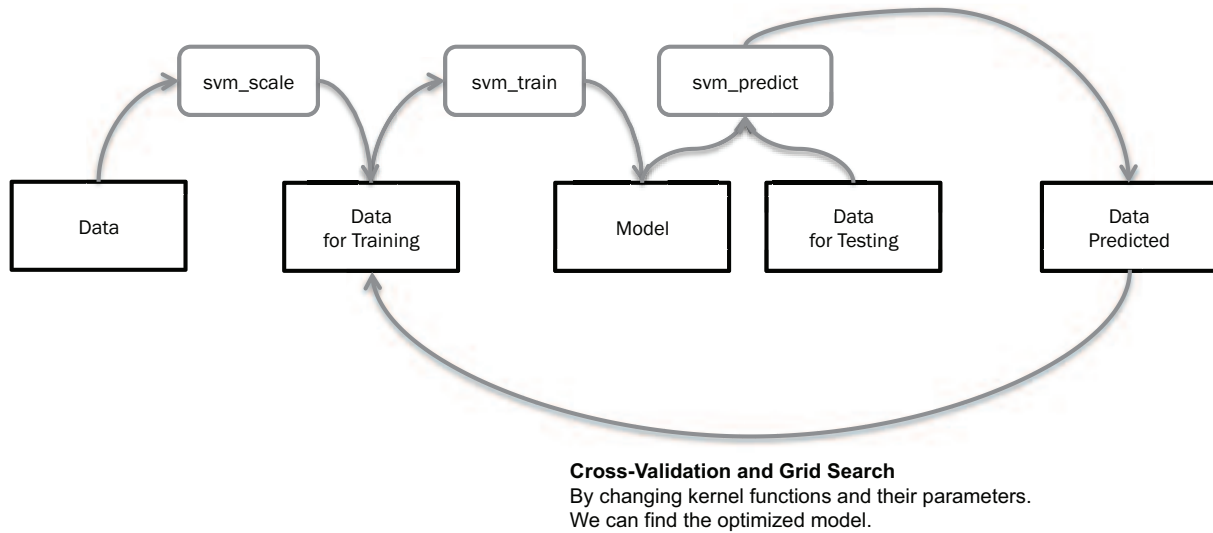


Fig. 2.— Flow chart of how to use LIBSVM.

language (<http://www.sunpy.org>).

2.2 Support Vector Machine

The Support Vector Machine (Boser et al. 1992; Cortes and Vapnik 1995) is one of the supervised learning methods used for classification and regression. It is used in many applications like medical image processing, finger print recognition, voice recognition, data mining and so on. We use LIBSVM (Chang et al. 2001) as a SVM library, which is widely used in the field of machine learning. It was developed via C language and supports many other languages. The main features of LIBSVM include efficient multi-class classification, cross-validation for model selection, probability estimates, various kernels, and weighted SVM for unbalanced data. Its official website is <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.

It provides three main functions; svm-scale, svm-train, and svm-predict. First of all, we have to scale input data sets from -1 to 1 because all data sets have

different ranges. If they have the same data range, we can generally get a higher accuracy performance. Scaling process can be done by using svm-scale. After scaling we train the scaled data sets and make model data by svm-train. By using the model data, we can predict the testing data via svm-predict. The svm-train produces many different models according to kernel functions of the SVM and their parameters. The kernel functions of the SVM are for measuring the similarity of classes. The LIBSVM supports several kernel functions such as linear, polynomial, radial basis function (RBF), and sigmoid.

We use a grid-search on C and γ using cross-validation for each kernel function. $C > 0$ is the penalty parameter of the error term, and γ means how much smoother or steeper in a decision boundary. By changing pairs of (C, γ) , we pick one pair with the best cross-validation accuracy. The practical method to identify good parameters is to try exponentially growing sequences of C ($C = 2^{-5}, 2^{-3}, \dots, 2^{15}$)

and γ ($\gamma = 2^{-15}, 2^{-13}, \dots, 2^3$). Fig. 2 shows a flow chart of how to use LIBSVM.

2.3 Statistical Verification Method

In order to compare the results between the SVM classification and the simple classifications with constant classifiers, we use statistical verification methods as follows. A contingency table is a matrix format that displays the frequency distribution of the variables. It is used to analyze the relation between categorical variables. Basically there are 2×2 grids assigned YES or NO prediction/observation designations. The number of events with YES prediction and YES observation is yy , a YES prediction and a NO observation is yn , and so on. Table 3 shows the contingency table and its grids.

Table 4 summarizes several statistical parameters based on this contingency table. Accuracy is the prediction accuracy that are correctly predicted and observed. PODy and PODn are the proportions of YES and NO observations that are correctly predicted. FAR is the proportion of YES predictions that are incorrect. Bias is the ratio of the number of YES predictions to the number of YES observations, and is a measure of over- or under-prediction. Critical Success Index (CSI) is the proportion of correct YES predictions that are either prediction or observed. It is also known as the Threat Score. True Skill Statistic (TSS) is a measure of the ability of the prediction to discriminate between YES and NO observations.

3. RESULTS

Scaling and digitizing input features are very important for applying the SVM to data. In the case of X-ray flare class, we assigned 1, 2, 3, and 4 for B, C, M, and X class, respectively. After that, we scale all the input features from -1 to $+1$; the speed and the angular width of CMEs, and their associated X-ray flare classes. We separate front-side halo and partial halo CMEs into two data sets. One is for training from 1996 to 2002 and the other is for testing from 2003 to 2010.

As mentioned before, we use the LIBSVM as a SVM library. To optimize the performance of the SVM, we need to determine the kernel functions and their parameters empirically. In this paper we use several kernel functions such as polynomial, radial basis function, and sigmoid in order to select the best kernel. We also change input features that are every combination of the speed and the angular width of the CME, and its associated X-ray flare class. For the application with each combination, we compute the statistical parameters such as accuracy, PODy, PODn, FAR, bias, CSI, and TSS. For each kernel, we run 110 trials by changing kernel parameters (C and γ) and then list the best result based on cross-validation in Table 5.

Among 18 results in Table 5, SX (the combination of speed and X-ray flare class) and kernel 2 (RBF) is

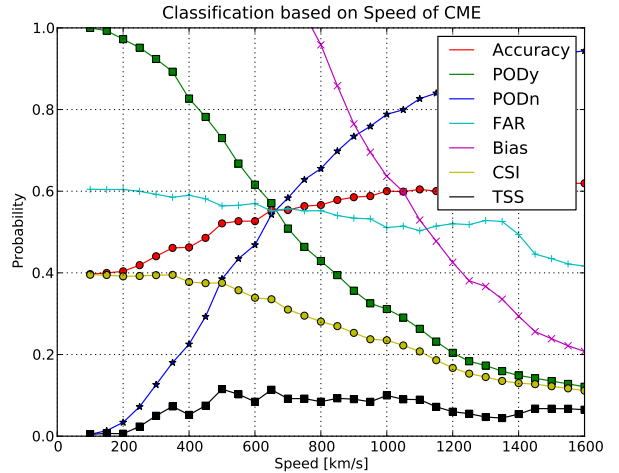


Fig. 3.— Statistical parameters of the classification based on CME speed.

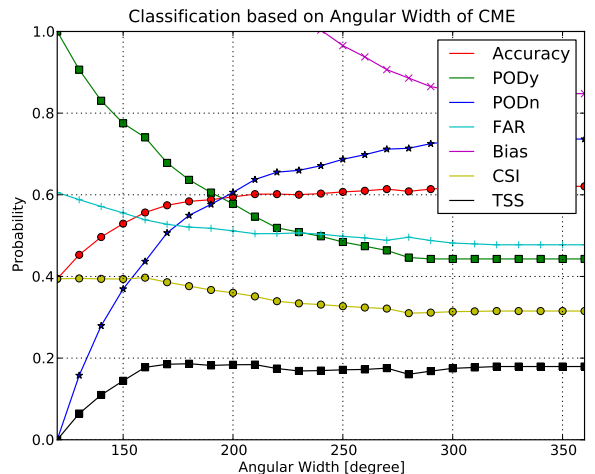


Fig. 4.— Statistical parameters of the classification based on CME angular width.

the best model in terms of the prediction accuracy. It produces the best prediction accuracy of 0.66 as well as the best PODy of 0.76. And the other statistical parameters are PODn=0.49, FAR=0.72, Bias=1.06, CSI=0.59, and TSS=0.25. Generally the RBF is the most popular choice of kernel functions used in the SVM. This is mainly because of their localized and finite responses across the entire range of the real data set. It also produces the best model in this data set.

In order to verify the SVM performance, we simply classify the CMEs according to each input feature with constant classifiers. First, we classify the CMEs based on CME speed from 100 to 1,600 km s^{-1} with

Table 3.
Contingency table for evaluation of prediction

		Observation		
		YES	NO	Total
Prediction	YES	yy	yn	$yy + yn$
	NO	ny	nn	$ny + nn$
	Total	$yy + ny$	$yn + nn$	$yy + yn + ny + nn$

Table 4.
Statistical verification parameters

Statistical Parameter	Description	Definition
Accuracy	Prediction Accuracy	$(yy + nn)/total$
PODy	Probability of Detection of Yes observations	$yy/(yy + ny)$
PODn	Probability of Detection of No observations	$nn/(yn + nn)$
FAR	False Alarm Ratio	$yn/(yy + yn)$
CSI	Critical Success Index	$yy/(yy + ny + yn)$
Bias	Prediction Bias	$(yy + yn)/(yy + ny)$
TSS	True Skill Statistic	$PODy + PODn - 1$

$$total = yy + yn + ny + nn$$

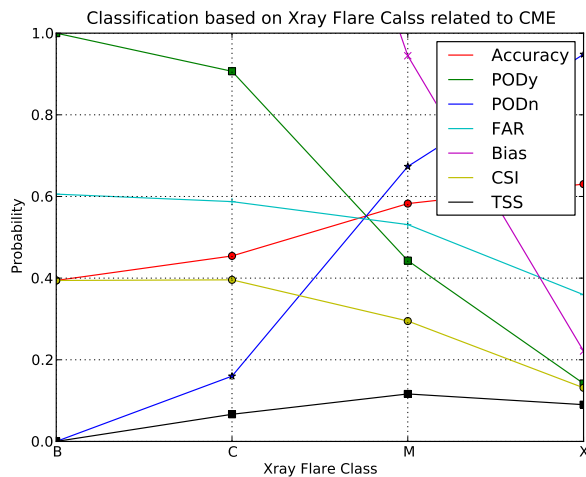


Fig. 5.— Statistical parameters of the classification based on X-ray flare class.

the step of 50 km s^{-1} . Fig. 3 shows the the statistical result of the CME speed classification. If we consider only the prediction accuracy, we can not always take meaningful results. As we can see in Fig. 3, it has a very poor performance in other statistical parameters

although it has a better result in the accuracy. So we consider PODy as another important statistical parameter along with the prediction accuracy. In this work we select the best model which have the best accuracy among the cases with PODy over 0.7. In the case of 500 km s^{-1} classifier, the statistical parameters are Accuracy=0.53, PODy=0.73, PODn=0.39, FAR=0.55, Bias=1.63, CSI=0.39, and TSS=0.12. Second, we classify the CMES based on CME angular width from 120° to 360° with the step of 10° . Fig. 4 shows the the statistical result of the CME angular width classification. In the case of 160° classifier, the statistical parameters are Accuracy=0.56, PODy=0.73, PODn=0.43, FAR=0.53, Bias=1.57, CSI=0.40, and TSS=0.17. Third, we classify CMES based on the X-ray flare class associated with the CME; B, C, M and X. Fig. 5 shows the the statistical result of the X-ray class classification. In the case of C class classifier, the statistical parameters are Accuracy=0.46, PODy=0.90, PODn=0.16, FAR=0.58, Bias=2.14, CSI=0.40, and TSS=0.06.

Table 6 shows the comparison of the results among the SVM classification and the simple classifications. In the case of the SVM classification, most of the important statistical parameters such as Accuracy, CSI, and TSS are all better than those of other classifications. In the case of X-ray class, it has the best PODy but its other statistical parameters are very poor.

In previous study, the classification with CME lo-

Table 5.
CME Classifications of the SVM by using cross-validation

Input	Kernel	c	g	yy	yn	ny	nn	Accuracy	PODy	PODn	FAR	Bias	CSI	TSS
S	1	-5	+3	22	23	45	16	0.36	0.33	0.41	0.49	0.67	0.24	-0.26
	2	+11	-13	29	24	38	15	0.42	0.43	0.38	0.55	0.79	0.32	-0.18
	3	-1	+3	28	10	39	29	0.54	0.42	0.74	0.74	0.57	0.36	0.16
W	1	-3	-7	41	15	26	24	0.61	0.61	0.62	0.73	0.84	0.50	0.23
	2	+7	+3	0	1	67	38	0.36	0.00	0.97	0.00	0.01	0.00	-0.03
	3	-5	-15	0	0	67	39	0.37	0.00	1.00	9.99	0.00	0.00	0.00
SW	1	-3	-7	25	5	42	34	0.56	0.37	0.87	0.83	0.45	0.35	0.24
	2	+15	+1	2	1	65	38	0.38	0.03	0.97	0.67	0.04	0.03	0.00
	3	-1	+3	6	6	61	33	0.37	0.09	0.85	0.50	0.18	0.08	-0.06
SX	1	-3	-7	46	18	21	21	0.63	0.69	0.54	0.72	0.96	0.54	0.23
	2	-5	-15	51	20	16	19	0.66	0.76	0.49	0.72	1.06	0.59	0.25
	3	+9	-1	15	4	52	35	0.47	0.22	0.90	0.79	0.28	0.21	0.12
WX	1	-5	-7	31	10	36	29	0.37	0.46	0.74	0.76	0.61	0.40	0.21
	2	+7	-3	21	11	46	28	0.46	0.31	0.72	0.66	0.48	0.27	0.03
	3	+13	-3	18	9	49	30	0.45	0.27	0.77	0.67	0.40	0.24	0.04
SWX	1	-1	-9	31	10	36	29	0.57	0.46	0.74	0.76	0.61	0.40	0.21
	2	+11	-5	21	11	46	28	0.46	0.31	0.72	0.66	0.48	0.27	0.03
	3	+7	-3	15	6	52	33	0.45	0.22	0.85	0.71	0.31	0.21	0.07

Input (Input Features) : S (Speed), W (Angular Width), X (X-ray Flare Class)

Kernel (Kernel Functions) : 1 (Polynomial), 2 (RBF, Radial Bias Function), 3 (Sigmoid)

Kernel Parameters : $c(C = 2^c)$, $g(\gamma = 2^g)$

Table 6.
Comparison of the results among the SVM classification and the simple classifications

Classifier	Accuracy	PODy	PODn	FAR	Bias	CSI	TSS
Speed($\geq 500 \text{ km s}^{-1}$)	0.53	0.73	0.39	0.55	1.64	0.39	0.12
Width ($\geq 160^\circ$)	0.56	0.73	0.43	0.53	1.57	0.40	0.17
X-ray class ($\geq \text{B}$)	0.46	0.90	0.16	0.58	2.15	0.40	0.06
SVM (RBF)	0.66	0.76	0.49	0.72	1.06	0.59	0.25

cations and speeds showed $POD_y=0.64$, $POD_n=0.68$, $FAR=0.43$, $Bias=1.13$, and $CSI=0.43$ (Kim et al. 2005). When comparing these values, we found that the POD_n and FAR of the SVM result are slightly lower, but POD_y , $Bias$ and CSI are significantly improved.

4. SUMMARY AND CONCLUSION

In this work we have applied the SVM to the prediction of geo-effective halo CMEs. For this we used halo and partial halo CMEs from January 1996 to April 2010 in the SOHO/LASCO CME Catalog. And we also used X-ray classes of their associated flares to identify front-side halo CMEs (stronger than B1 class), and the Dst index to determine geo-effective halo CMEs (stronger than -50 nT). All the combinations of the speed and the angular width of CME, and its associated X-ray class are used for input features for the SVM classification. Since the SVM produces many models according to the kernel functions and their parameters, we have found the best model by using the cross-validation. We consider the POD_y along with the prediction accuracy as an important statistical parameter. To demonstrate the SVM performance, we compare it with the simple classifications with constant classifiers. The performances of the statistical parameters by applying the SVM are much better than those from the simple classifications between geo-effective and non-geo-effective CMEs. By using CME speed and X-ray flare class as input features of the SVM, we have obtained statistical parameters for the best model: $Accuracy=0.66$, $POD_y=0.76$, $POD_n=0.49$, $FAR=0.72$, $Bias=1.06$, $CSI=0.59$, and $TSS=0.25$.

In recent times, the needs of machine learning are growing because it can deal with large amount of data from space missions. For example, automated systems using machine learning algorithms have been developed for the solar dynamics observatory (SDO) mission (Martens et al. 2009; Attrill et al. 2010). From the present study together with the previous ones, the most important advantage of the SVM for space weather forecast is that we can select the best model among many possible models by changing various kernels and their parameters. Our results can be used for the space weather applications. The developed codes can be integrated to an automatic prediction system if CME speed and width are automatically determined (e.g., CACTUS) and GOES X-ray data are provided in near-real time. In addition, such an application can be extended to other space weather forecasts such as solar flare forecast and solar proton event forecast.

ACKNOWLEDGMENTS

This work has been supported by Kyung Hee University (20090714). The CME catalogue we have used is generated and maintained by the CDAW Data Center by NASA and The Catholic University of Amer-

ica in cooperation with the Naval Research Laboratory. SOHO is a project of international cooperation between ESA and NASA.

REFERENCES

- Al-Omari, M., Qahwaji, R., Colak, T., & Ipson, S. 2010, Machine Learning-Based Investigation of the Associations between CMEs and Filaments, *Solar Physics*, 262, 511
- Attrill, G. D. R., & Wills-Davey, M. J. 2010, Automatic Detection and Extraction of Coronal Dimmings from SDO/AIA Data, *Solar Physics*, 262, 461
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. 1992, 5th Annual ACM Workshop on COLT, pages 144152, Pittsburgh, PA, A training algorithm for optimal margin classifiers. In D. Haussler, editor, ACM Press
- Chang, C.-C., & Lin, C.-J. 2001, LIBSVM : A Library for Support Vector Machines, <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- Chen, C., Wu, Z. S., Xu, Z. W., Sun, S. J., Ding, Z. H., & Ban, P. P. 2010, Forecasting the Local Ionospheric f0F2 Parameter 1 Hour ahead during Disturbed Geomagnetic Conditions, *JGR*, 115, A11315
- Colak, T., & Qahwaji, R. 2009, Automated Solar Activity Prediction: A Hybrid Computer Platform Using Machine Learning and Solar Imaging for Automated Prediction of Solar Flares, *Space Weather*, Vol. 7, S06001, 12PP
- Cortes, C., & Vapnik, V. 1995, Support-Vector Networks, *Machine Learning*, 20
- Gavrishchaka, V. V., & Ganguli, S. B. 2001, Support Vector Machine as an Efficient Tool for High-Dimensional Data Processing: Application to Substorm Forecasting, *JGR*, 106, 29911
- Gopalswamy, N., Lara, A., Yashiro, S., Kaiser, M. L., & Howard, R. A. Predicting the 1-AU Arrival Times of Coronal Mass Ejections, *JGR*, 106, A12, 29207
- Gopalswamy, N., Yashiro, S., & Akiyama, S. Geoeffectiveness of Halo Coronal Mass Ejections, *JGR*, 112, A6
- He, H., Wang, H., Du, z., Li, R., Chui, Y., Zhang, L., & He, Y. 2008, Solar Activity Prediction Studies and Services in NAOC, *Advances in Space Research*, 42, 1450
- Henwood, R., Chapman, S. C., & Willis, D. M. 2010, Increasing Lifetime of Recurrent Sunspot Groups Within the Greenwich Photoheliographic Results, *Solar Physics*, 262, 299
- Kim, R.-S., Cho, K.-S., Moon, Y.-J., Kim, Y.-H., Yi, Y., Dryer, M., Bong, S.-C., & Park, Y.-D. 2005, Forecast Evaluation of the Coronal Mass Ejection (CME), Geoeffectiveness Using Halo CMEs from 1997 to 2003, *JGR*, 110, A11104

- Kim, R.-S., Cho, K.-S., Kim, K.-H., Park, Y.-D., Moon, Y.-J., Yi, Y., Lee, J., Wang, H., Song, H., & Dryer, M. 2008, CME Earthward Direction as an Important Geoeffectiveness Indicator, *ApJ*, 677, 1378
- Labrosse, N., Dalla, S., & Marshall, S. 2010, Automated Detection of Limb Prominences in 304 Å EUV Images, *Solar Physics*, 262, 449
- Li, R., Wang, H.-N., He, H., Cui, Y.-M., & Du, Z.-L. 2007, Support Vector Machine Combined with K-Nearest Neighbors for Solar Flare Forecasting, *Chin. J. Astron. Astrophys.*, Vol. 7, No. 3, 441
- Liu, D. D., Huang, C., Lu, J. Y., & Wang, J. S. 2011, The Hourly Average Solar Wind Velocity Prediction Based on Support Vector Regression Method, *MNRAS*, 413, 2877
- Martens, P. C. H., Attrill, G. D. R., Davey, A. R., Engell, A., Farid, S., Grigis, P. C., Kasper, J., Korreck, K., Saar, S. H., Savcheva, A., Su, Y., Testa, P., Wills-Davey, M., Bernasconi, P. N., Raouafi, N.-E., Delouille, V. A., Hochedez, J. F., Cirtain, J. W., Deforest, C. E., Angryk, R. A., de Moortel, I., Wiegelmann, T., Georgoulis, M. K., McAteer, R. T. J., & Timmons, R. P. 2009, Computer Vision for the Solar Dynamics Observatory (SDO), *Solar Physics*, tmp 144
- Moon, Y.-J., Cho, K.-S., Chae, J., Choe, G. S., Kim, Y.-H., Bong, S.-C., & Park, Y.-D. New Extrapolation Method for Coronal Mass Ejection Onset Time Estimation, *JGR*, 110, A7
- Qahwaji, R., & Colak, T. 2007, Automated Short-Solar Flare Prediction Using Machine Learning and Sunspot Associations, *Solar Physics*, 241, 195
- Qahwaji, R., Colak, T., Al-Omari, M., & Ipson, S. 2008, Automated Prediction of CMEs Using Machine Learning of CME-Flare Associations, *Solar Physics*, 248, 471
- Qu, M., Shih, F. Y., Jing, J., & Wang, H. 2003, Automated Solar Flare Detection Using MLP, RBF, and SVM, *Solar Physics*, 217, 157
- Qu, M., Shin, F. Y., Jing, J., & Wang, H. 2005, Automatic Solar Filament Detection Using Image Processing Techniques, *Solar Physics*, 228, 119
- Srivastava, N. K., & Venkatakrishnan, P. 2004, Solar and Interplanetary Sources of Major Geomagnetic Storms during 1996-2002, *JGR*, 109, A10
- Wang, Y. M., Ye, P. Z., Wang, S., Zhou, G. P., & Wang, J. X. 2002, A Statistical Study on the Geoeffectiveness of Earth-Directed Coronal Mass Ejections from March 1997 to December 2000, *JGR*, 107, A11
- Webb, D. F. 2002, CMEs and the Solar Cycle Variation in Their Geoeffectiveness, ISBN 92-9092-818-2, 2002, 409
- Yu, D., Huang, X., Wang, H., & Cui, Y. 2009, Short-Term Solar Flare Prediction Using a Sequential Supervised Learning Method, *Solar Physics*, 255, 91
- Yuan, Y., Shih, F. Y., Jing, J., & Wang, H. 2010, Automated Flare Forecasting using a Statistical Learning Technique, *Res. Astron. Astrophys.*, 10, 785
- Zhang, J., Richardson, I. G., Webb, D. F., Gopalswamy, N., Huttunen, E., Kasper, J. C., Nitta, N. V., Poomvises, W., Thompson, B. J., Wu, C.-C., Yashiro, S., & Zhukov, A. N. Solar and Interplanetary Sources of Major Geomagnetic Storms ($Dst \leq -100$ nT) during 1996-2005, *JGRA*, 112, A10