南京信息工程大学

Support vector machine combined with distance correlation learning for Dst forecasting during intense geomagnetic storms

Jianyong Lu

Nanjing University of Information Science & Technology

2016.10.20

Introduction

Many intentions to forecast the Dst index by using the real-time solar wind data：

- Traditional Statistic Methods: Burton et al.(1975). then, Gonzalez et al.(1994), Murayama (1986), Thomsen et al.(1998), Klimas et al. (1998), and O'Brien and McPherron (2000).
- Neural Network (NN) : Wu and Lundstedt (1996), then Lundstedt et al. (2002) (a recurrent Neural Network for real time forecasting of Dst index. Ji et al. (2012) compared the performance of six Dst forecast models during intense geomagnetic storms.

The present study is the first application of the Support Vector Machine (SVM) combine with Distance Correlation learning (DC) .

Introduction

SVM technique was applied in many space weather forecasting:

- Li et al. (2008) in solar flare and solar proton events
- Huang et al. (2009) in daily values of F10.7
- Chen et al. (2010) in local ionospheric foF2
- Ban et al. (2011) in low-latitude storm-time ionospheric foF2
- Liu et al. (2011) in the solar wind velocity
- Choi et al. (2012) in CMEs
- Ji et al. (2013) in Kp index

Intense geomagnetic storm events

Table 1. 80 intense geomagnetic storms that occurred from 1995 to 2014.

Number	Start time (UT)	End time (UT)	Dst_{min}
1	1995/3/26 0500	1995/3/26 2400	-107
2	1995/4/7 1300	1995/4/9 0900	-149
3	1995/9/27 0100	1995/9/28 0400	-108
4	1995/10/18 1300	1995/10/19 1400	-127
5	1996/10/22 2200	1996/10/23 1100	-105
6	1997/4/21 1000	1997/4/22 0900	-107
7	1997/5/15 0300	1997/5/16 0000	-115
8	1997/10/10 1800	1997/10/11 1900	-130
9	1997/11/7 0000	1997/11/7 1800	-110
10	1997/11/22 2100	1997/11/24 0400	-108
	``		
71	2011/10/24 2000	2011/10/25 1400	-132
72	2012/3/8 1200	2012/3/10 1600	-131
73	2012/4/23 1100	2012/4/24 1300	-108
74	2012/7/15 0100	2012/7/16 2300	-127
75	2012/9/30 1300	2012/10/1 1800	-119
76	2012/10/8 0200	2012/10/9 1700	-105
77	2012/11/13 1800	2012/11/14 1800	-108
78	2013/3/17 0700	2013/3/18 1000	-132
79	2013/5/31 1800	2013/6/1 2000	-119
80	2014/2/18 1500	2014/2/19 1600	-112

The 80 intense storm events ($Dst_{min} \leq -100$) that we used were occurred from 1995 to 2014.

The Dst data is obtained from the World Data Center for Geomagnetism, Kyoto (http://wdc.kugi.kyoto-u.ac.jp/).

The solar wind data observed by ACE spacecraft is obtained from the Coordinated Data Analysis Web (CDA Web) (http://cdaweb.gsfc.nasa.gov/cdaweb/).

DC value of solar wind parameters

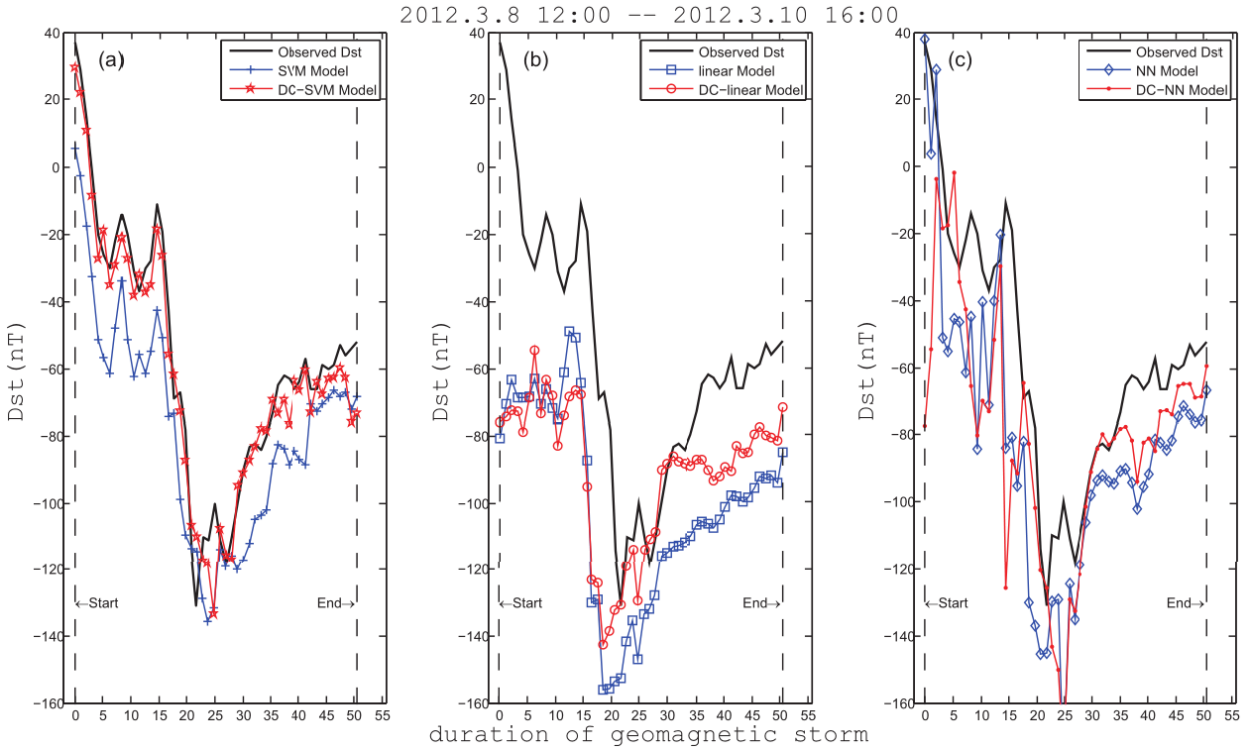
Table 2. DC value of solar wind parameters

Number	Solar wind parameters	DC value
1	E_y	0.3297
2	V_x	0.2596
3	*Plasmabeta*	0.2536
4	*Protondensity*	0.2498
5	B_z	0.2498
6	*Dynamicpressure*	0.2316
7	*Alfvenmach*	0.2215
8	*Coneangle*	0.1941
9	B_y	0.1558
10	V_y	0.1151
11	B_x	0.0911
12	*Temperature*	0.0724
13	V_z	0.0531
14	*Clockangle*	0.0236
15	*Pitchangle*	0.0196

Distance correlation learning (DC) is a method for feature screening which could measure the dependence between random vectors [Székely et al., 2007]. DC satisfies $0 \leq DC \leq 1$, and DC = 0 indicates the random vectors are independent.

In this study, we choose 13 effective parameters by judging the forecasting performance of models.

Comparison of a storm event for six models

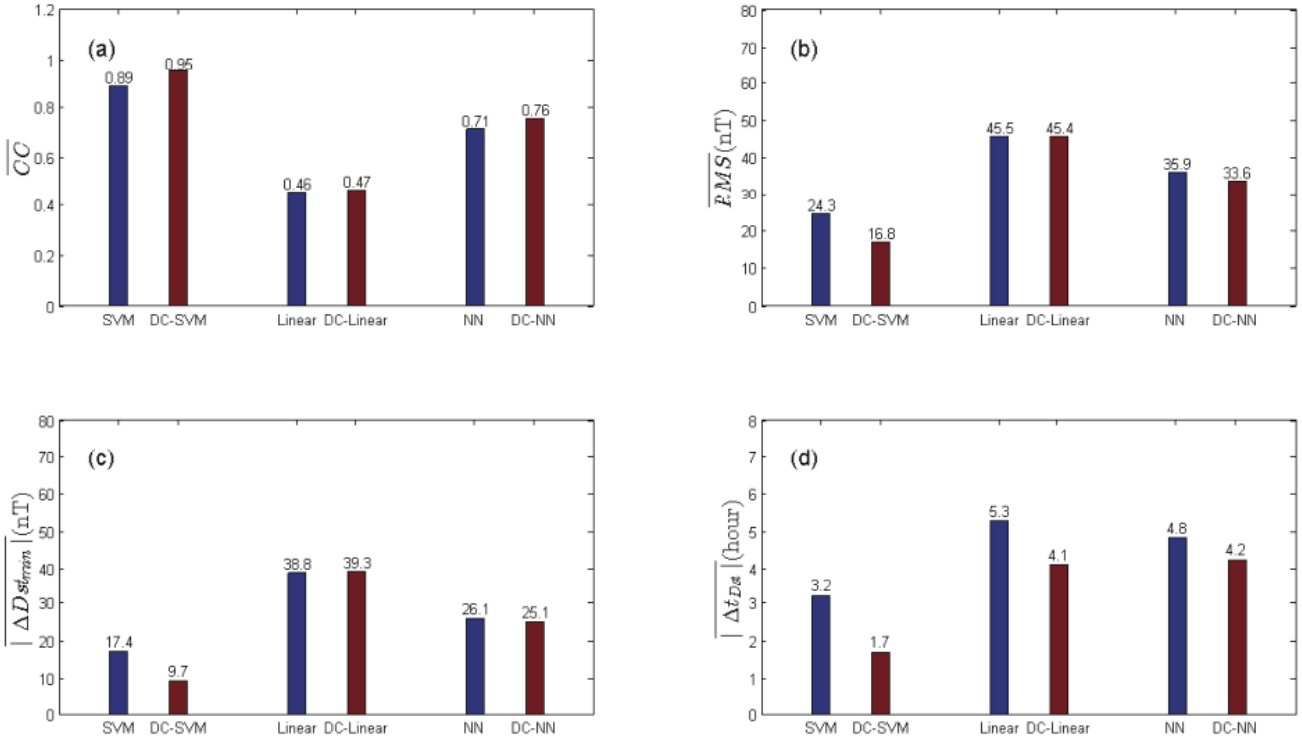


Six models:
1. SVM model.
2. DC-SVM model.
3. Linear (linear machine) model.
4.DC-Linear model.
5.NN (neural network) model.
6.DC-NN model.

This figure shows the comparison between the observed Dst (black) and the predicted from six models for the storm event in March 2012.

We can find from the figures that the DC-SVM model is the best in describing the variation and magnitude of the observed Dst.

Statistical comparisons for all events



CC: correlation coefficients.

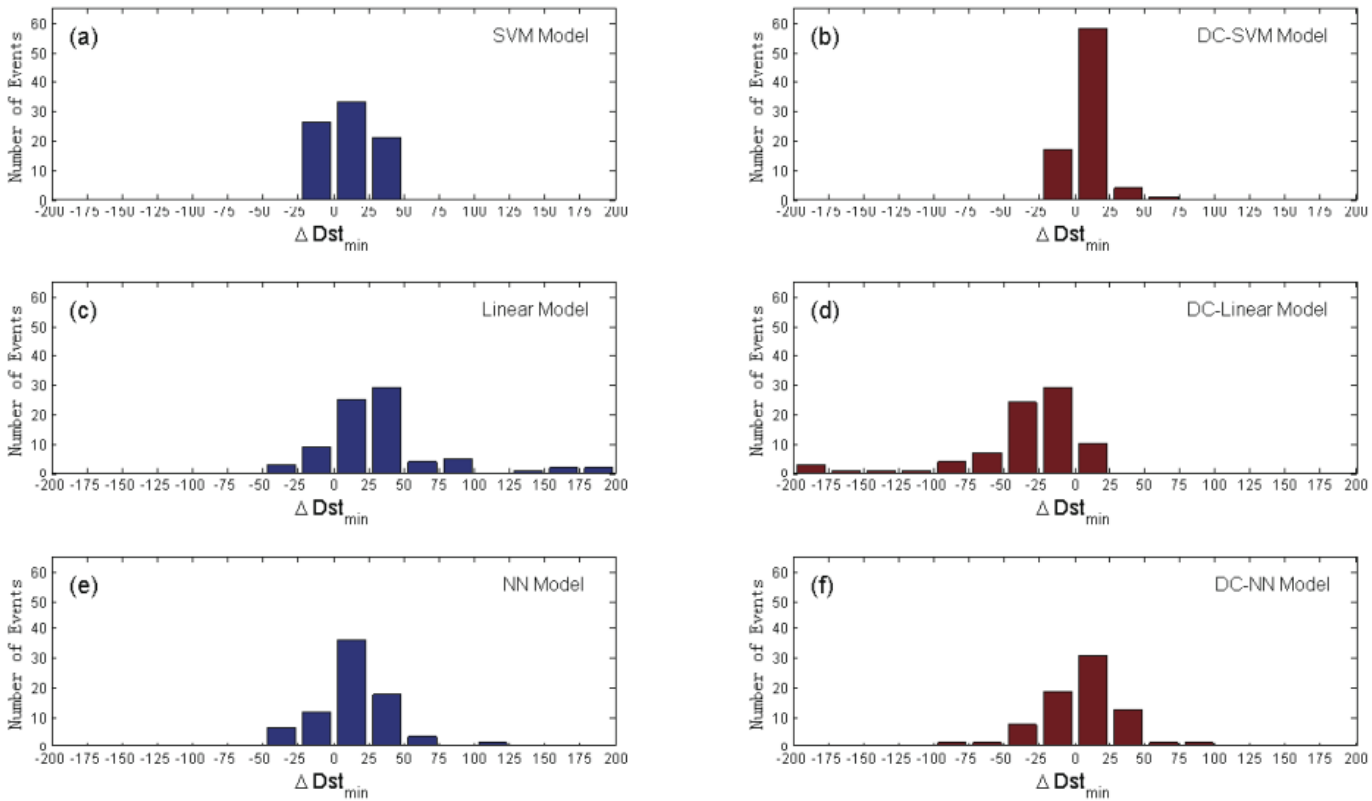
RMS: root mean square error.

$|\triangle Dst_{min}|$: the absolute value of difference in minimum Dst between observed value and predicted one.

$|\triangle t_{Dst}|$: the absolute value of difference in minimum time between observed value and predicted one.

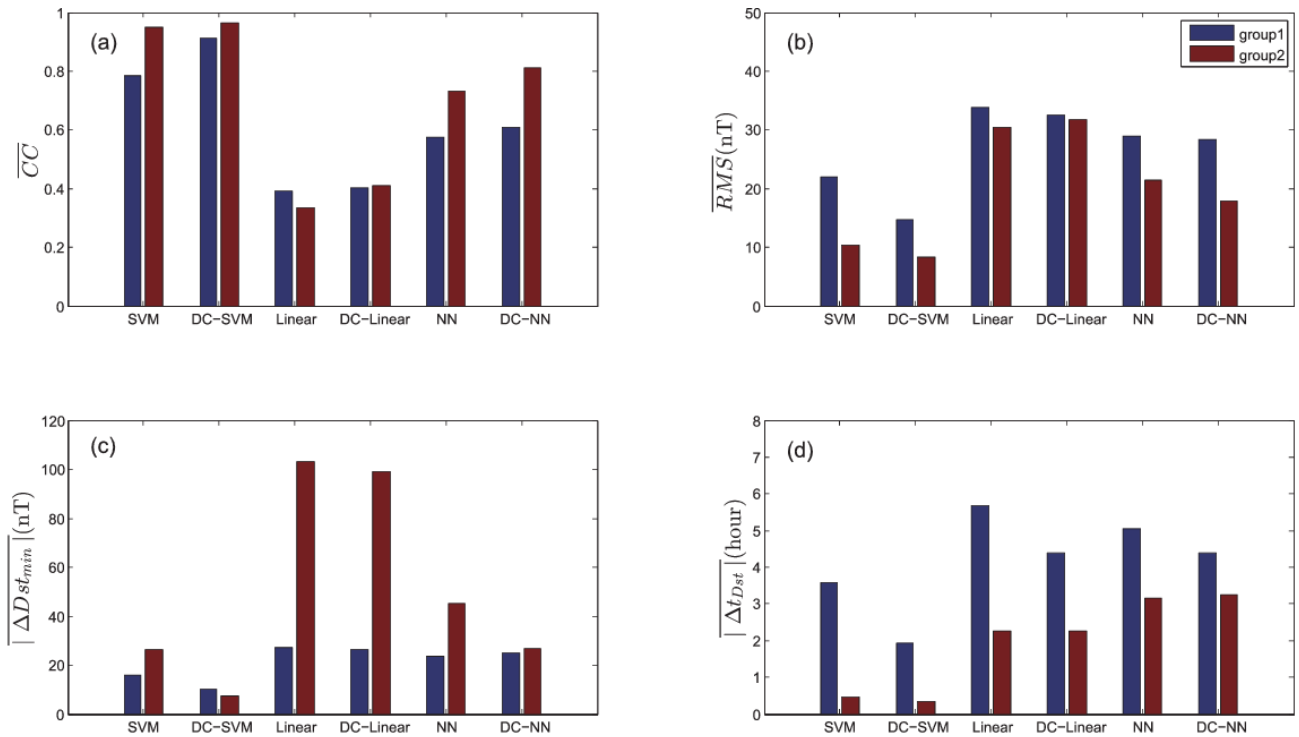
- The $|\triangle Dst_{min}|$ and $|\triangle t_{Dst}|$ of DC-SVM are much smaller than other models, which show the accuracy for DC-SVM to forecast the minimum Dst value
- DC-SVM model shows the best performance for all parameters
- Distance correlation learning significantly improves the performance of all the models.

The events number distribution of $\triangle Dst_{min}$ for six models



$\triangle Dst_{min}$ from the SVM model and DC-SVM model are obviously more concentrated near zero than other models; further more, the DC-SVM model gives the most concentrated $\triangle Dst_{min}$, which is mainly distributed within ± 25 nT.

Comparison dependence on Dst magnitude



For further comparison, we divided storm events into two groups according to the value of Dst_{min}:

Group1:

$-200 < Dst_{min} \leq -100$ nT

Group2:

$Dst_{min} \leq -200$ nT

We can find from the figures:

1. CC for the DC-SVM model is the highest among the six models for both groups.
2. RMS value for the DC-SVM model is the minimum among all the models.
3. Both $|\triangle Dst_{min}|$ and $|\triangle t_{Dst}|$ values of DC-SVM are also the lowest in two groups.
4. DC technology significantly improves the performance of six models to forecast Dst index.

Conclusion

The main purpose of this study is to develop a new effective method to forecast Dst index during intense storms and explore a new technology for space weather real-time forecasting.

We firstly build a DC-SVM model that combines SVM method with distance correlation learning technology in Dst forecast. From the statistical comparison of six Dst forecast models' results, we can find that DC-SVM model is the best in forecasting the amplitude and variation of Dst during both intense and super intense geomagnetic storms.

Thanks

(jylu@nuist.edu.cn)