GENERALIZED NET OF THE PROCESS
OF SUPPORT VECTOR MACHINES CLASSIFIER CONSTRUCTION

Veselina Bureva¹, Mirela Puleva² and Evdokia Sotirova¹

¹ Prof. Asen Zlatarov University
Burgas-8000, Bulgaria
e-mails: vbureva@btu.bg, esotirova@btu.bg

² School of Mathematics and Natural Sciences Acad. Nikola Obreshkov
Burgas-8000, Bulgaria
e-mail: mirrellaa@gmail.com

Abstract: In the current paper is presented a generalized net model of the process of
construction of a support vector machine classifier. It can be helpful in analyzing, managing
and optimizing the process. The implementation of the technique is applied to weather data.

Keywords: Artificial intelligence, Generalized nets, Data mining, Machine learning,
Support vector machines, Weather databases.

1 Introduction

Support vector machines (SVMs) is a statistical method frequently used in the area of
machine learning. In data mining SVMs is applied in order to classifying data and predict
the future trends in different areas – medical diagnosis, bioinformatics, pattern recognition,
image processing and text mining. The support vector classifiers have to be trained and after
that they can be used for predicting future patterns. SVM algorithms try to find an “optimal
hyper-plane” that can maximize the distance two groups of samples. The hyper-plane is
optimal when it stands at maximum distance to the points of two classes. The two parallel
planes to the hyper-plane, which are placed the nearest to the points of the separating sets are
called “support vectors” (Fig. 1). The distance between the support vectors for each class is
called “margin”. In the case, when there are many hyper-planes it searches for the
“maximum hyper-plane”. It separates the classes the best. The elements from the classes that
are closest to the hyperplane are called support vectors. The support vectors must be
maximal near the points of the each class and maximal the farthest from the hyper-plane
[9–12]. When the distance between the support vectors and the hyper-plane is too little, then
the classifier has a “hard margin”. The classes may include the single points, which are mapped on the margin place or they can be mapped on the place for the other class. In this case, these points are perceived as “noise” in the data and the margin technique is called “soft margin” [13, 15]. The optimization is made by applying the other methods (quadratic optimization) and introducing new parameters – “slack variables $\varepsilon$”. When the sets cannot separate linearly it is needed to be mapped the input points on high dimensional space and the technique “kernel trick” to be applied. This is made while the classes become linearly separable. In this high dimensional feature space, linear classification can be performed. Therefore, the SVM is used to successfully classify linearly separable and non-linearly separable datasets [4–8].

![Figure 1. The separating hyperplane and support vectors](image)

In the following paragraph are described the simple steps of algorithm for SVM in non-linear case:

- Preprocess the data – (missing values, presumably incorrect values, noise in the data, Numerisize the data, Normalize the data);
- Develop the model(s) – Select kernel type, Determine kernel parameters based on the selected kernel type (This is a hard problem. One should consider using cross validation and experimentation to determine the appropriate values for these parameters);
- If the results are satisfactory, finalize the model, otherwise change the kernel type and/or kernel parameters to achieve the desired accuracy level;
- Extract and deploy the model.

In Multi-class classification problems can be included algorithms based thoroughly on the two-class classification [19, 20]. On the area of these techniques belong all other techniques that have and own process of training the data. Widely-used methods are One-against-all and One-against-one, Decision tree SVM, DAG SVM (directed acyclic graph SVM), Fuzzy SVM. While they training the data, they group all the classes in pairs.
2 Generalized network model

The concept of Generalized nets (GNs) is introduced in [1, 2]. The Generalized Net model of the process of support vector machines classifier construction is presented in Fig. 2. In [3, 14, 18] are presented other methods for knowledge discovery, which are modelled by generalized nets. The implementation of these techniques is also made by using a weather data.

The GN model of the process of support vector machines contains 11 transitions and 48 places. It contains the following set of transitions:

\[ A = \{Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7, Z_8, Z_9, Z_{10}, Z_{11}\}, \]

where the transitions describe these process:

- \( Z_1 \) – “preprocessing the input data”;
- \( Z_2 \) – “a multi-class classification”;
- \( Z_3 \) – “separating the input data on training, validation and testing sets”;
- \( Z_4 \) – “choosing a method for multi-class classification”;
- \( Z_5 \) – “mapping data on R^2, finding the best hyper-plane, the support vectors, and the margin”;
- \( Z_6 \) – “choosing the another method for multi-class classification or one multi-class classifier needs additional parameters (Decision Tree SVM, directed acyclic graph SVM (DAGSVM))”;
- \( Z_7 \) – “optimization of the classifier”;
- \( Z_8 \) – “finding nonlinear classifier”;
- \( Z_9 \) – “finding a multi-class classifier”;
- \( Z_{10} \) – “validating the classifier”;
- \( Z_{11} \) – “testing the classifier”.

Initially, there is one \( \alpha_1 \)-token that is located in place \( L_6 \) with initial characteristic: “data for preprocessing”.

In the next time-moments this token is split into two or more. The original \( \alpha_1 \)-token will continue to stay in place \( L_6 \), while the other \( \alpha \)-tokens will move to the next transitions.

Initially, there is one \( \beta_1 \)-token that is located in place \( L_5 \) with initial characteristic: “preprocessing techniques”.

In the next time-moments this token is split into two. The original \( \beta_1 \)-token will continue to stay in place \( L_5 \), while the other \( \beta \)-token will move to transition \( Z_2 \) via place \( L_3 \).

From places \( L_1 \), \( L_2 \), and \( L_7 \) the \( \alpha_2 \)-, \( \beta_2 \)- and \( \beta_2 \)-tokens enter the net with initial characteristic respectively: “new data for classifying” in place \( L_1 \), “new preprocessing technique” in place \( L_2 \), and “method for multi-class classification” in place \( L_7 \).
The transition $Z_1$ has the form:

$$Z_1 = \langle \{L_1, L_2, L_4, L_6\}, \{L_1, L_4, L_5, L_6\}, R_1, \lor (L_1, L_2), L_4, L_5, L_6 \rangle,$$

where:
\[ R_1 = \begin{array}{c|ccccc}
L_1 & L_2 & L_4 & L_5 & L_6 \\
\hline
L_1 & \text{false} & \text{false} & \text{false} & \text{true} \\
L_2 & \text{false} & \text{false} & \text{true} & \text{false} \\
L_4 & W_{4,3} & W_{4,4} & \text{false} & \text{false} \\
L_5 & \text{false} & W_{5,4} & W_{5,5} & \text{false} \\
L_6 & \text{false} & W_{6,4} & \text{false} & W_{6,6} \\
\end{array} \]

and:

\begin{itemize}
    \item \( W_{4,3} = \text{"The data is preprocessed"}, \)
    \item \( W_{4,4} = \neg W_{4,3} \).
    \item \( W_{5,4} = \text{"The method for preprocessing is selected"}, \)
    \item \( W_{5,5} = \neg W_{5,4} \).
    \item \( W_{6,4} = \text{"The data to be preprocessed is selected"}, \)
    \item \( W_{6,6} = \neg W_{6,4} \).
\end{itemize}

The \( \alpha \)-tokens that enter in place \( L_5 \) obtain the characteristic: "Data for partitioning".

The transition \( Z_2 \) has the form:
\[ Z_2 = (\{ L_7, L_{10}, L_{15} \}, \{ L_8, L_9, L_{10} \}, R_2, \lor (\land (L_7, L_{11}), L_{10})) , \]
where:
\[ R_2 = \begin{array}{c|ccc}
L_7 & L_9 & L_{10} \\
\hline
L_7 & \text{false} & \text{false} & \text{true} \\
L_{10} & W_{10,8} & W_{10,9} & W_{10,10} \\
L_{15} & \text{false} & \text{false} & \text{true} \\
\end{array} \]

and:

\begin{itemize}
    \item \( W_{10,8} = \text{"The method for the input data, based on two-class classification, is chosen"}, \)
    \item \( W_{10,9} = \text{"The other method for the input data is chosen"}, \)
    \item \( W_{10,10} = \lor (W_{10,8} \land W_{10,9}) \).
\end{itemize}

The \( \alpha \)-tokens that enter in the places \( L_8 \) and \( L_9 \) have the following characteristics: "the method based on two-class classification and input data" in place \( L_8 \), and "other method and input data" in place \( L_9 \).

The token \( \beta_4 \) enters the net via place \( L_{17} \) with initial characteristics: "method for grouping the classes".

The transition \( Z_3 \) has the form:
\[ Z_3 = (L_3, L_{11}, L_{16}), \{ L_{12}, L_{13}, L_{14}, L_{15}, L_{16} \}, R_3, \lor (L_3, L_{11}, L_{16})) , \]
where:
\[
R_3 = \begin{array}{cccccc}
L_2 & L_3 & L_4 & L_5 & L_6 \\
\text{false} & \text{false} & \text{false} & \text{false} & \text{true} \\
L_7 & L_8 & L_9 & L_{10} & L_{11} \\
\text{false} & \text{false} & \text{false} & \text{false} & \text{true} \\
L_{12} & L_{13} & L_{14} & L_{15} & L_{16} \\
W_{16,12} & W_{16,13} & W_{16,14} & W_{16,15} & W_{16,16}
\end{array}
\]

and:
- \(W_{16,12} = W_{16,13} = W_{16,14} = W_{16,15} = \text{"The data is partitioned"},\)
- \(W_{16,16} = \neg (W_{16,12} \land W_{16,13} \land W_{16,14} \land W_{16,15}).\)

The \(\alpha\)-tokens that enter in the places \(L_{12}, L_{13}, L_{14}, L_{15}\) have the following characteristics: “Training set for two-class problem”, “Validation set”, “Testing set”, “Training set for multi-class problem”.

The transition \(Z_4\) has the form:
\[
Z_4 = (\{L_8, L_{17}, L_{19}, L_{20}\}, \{L_{18}, L_{19}, L_{20}\}, R_4, \lor (\land (L_8, L_{17}), L_{19}, L_{20}))
\]

where:
\[
R_4 = \begin{array}{cc}
L_8 & L_9 \\
\text{false} & \text{false} \\
L_{18} & L_{19} \\
\text{false} & \text{false} \\
L_{20} & L_{21} \\
\text{false} & \text{true} \\
\end{array}
\]

and:
- \(W_{19,18} = \text{"There is a two-class combination for all classes"},\)
- \(W_{19,19} = \neg W_{19,18},\)
- \(W_{20,19} = \text{"There are classes for combining"},\)
- \(W_{20,20} = \neg W_{20,19}.\)

The \(\alpha\)-tokens that enter in the places \(L_{18}\) and \(L_{19}\) have the following characteristics: “two classes for classification” in the place \(L_{18}\) and “there is a classes for grouping” in the place \(L_{19}\).

The token \(\beta_4\) enters the net via place \(L_{21}\) with initial characteristics: “labels for the classes: 1 and –1”.

The transition \(Z_5\) has the form:
\[
Z_5 = (\{L_{13}, L_{14}, L_{21}, L_{26}, L_{27}\}, \{L_{22}, L_{23}, L_{24}, L_{25}, L_{26}, L_{27}\}, \)
\[
R_5, \lor (\land (L_{13}, L_{14}, L_{21}), L_{26}, L_{27}))
\]

where:
and:

- $W_{26,22}$ = “The best hyper-plane, the support vectors and the margin are not determined and the model have to be constructed in higher dimension (nonlinear separable case)
- $W_{26,23}$ = “The best hyper-plane, the support vectors and the margin are determined and the model will be optimized (soft margin - noise, outliers)”,
- $W_{26,24}$ = “The best hyper-plane, the support vectors and the margin are determined and the model will be validated”,
- $W_{26,25}$ = “The best hyper-plane, the support vectors and the margin are determined and the constructed model will be for multi-class classification”,
- $W_{26,26}$ = ¬($W_{26,22}$∧$W_{26,23}$∧$W_{26,24}$∧$W_{26,25}$),
- $W_{27,26}$ = “There is the input data for mapping on the $\mathbb{R}^2$”,
- $W_{27,27}$ = ¬$W_{27,26}$.

The $\alpha$-tokens that enter in places $L_{22}, L_{23}, L_{24}, L_{25}, L_{26}$ have the following characteristics: “model for constructing in higher dimension” in place $L_{22}$, “the support vectors and the margin for model optimization” in place $L_{23}$, “the support vectors and the margin for multi-class model” in place $L_{23}$, “linear model for validation” in place $L_{25}$, and “input data for mapping in $\mathbb{R}^2$” in place $L_{26}$.

The tokens $\beta_5$ enters the net via places $L_{23}$ with initial characteristic: “other parameters for multi-class classification”.

The transition $Z_6$ has the form:

$Z_6 = \langle \{L_9, L_{28}, L_{30}, L_{44}\}, \{L_{29}, L_{30}\}, R_6, \lor (\land (L_9, L_{28}, L_{44}), L_{30}) \rangle,$

where:

$R_6 = \begin{array}{c|cc}
L_9 & L_{29} & L_{30} \\
\hline
false & true \\
false & true \\
L_{30} & W_{30,29} & W_{30,30} \\
L_{44} & false & true \\
\end{array}$

and:

- $W_{30,29}$ = “There is a model ready”,
- $W_{30,30}$ = ¬$W_{30,29}$.
The \( \alpha \)-token that enters in the place \( L_{29} \) has the following characteristic: “multi-class model ready”.

The tokens \( \beta_6 \) and \( \beta_7 \) enter the net via places \( L_{31} \) and \( L_{32} \) with initial characteristics: “values for slack variables \( (\epsilon_1, \ldots, \epsilon_n) \)” in place \( L_{31} \), and “values for an additional parameter \( C \)” in place \( L_{32} \).

The transition \( Z_7 \) has the form:

\[
Z_7 = \{ L_{23}, L_{31}, L_{32}, L_{40}, L_{36} \}, \{ L_{33}, L_{34}, L_{35}, L_{36} \}, R_7, \lor ( L_{23}, L_{31}, L_{32}, L_{40} ), ( L_{36} )
\]

where:

\[
R_7 = \begin{bmatrix}
L_{23} & \text{false} & \text{false} & \text{false} & \text{true} \\
L_{31} & \text{false} & \text{false} & \text{false} & \text{true} \\
L_{32} & \text{false} & \text{false} & \text{false} & \text{true} \\
L_{36} & L_{36,33} & L_{36,34} & L_{36,35} & L_{36,36} \\
L_{40} & \text{false} & \text{false} & \text{false} & \text{true}
\end{bmatrix}
\]

and:

- \( W_{36,33} = “\text{The model is optimized}” \)
- \( W_{36,34} = “\text{The model cannot be optimized and it needs to be mapped on higher dimension}” \)
- \( W_{36,35} = “\text{The model is part of the multi-class problem (model)}” \)
- \( W_{36,36} = \neg ( W_{36,33} \land W_{36,34} \land W_{36,35} ) \).

The \( \alpha \)-tokens that enter in the places \( L_{33}, L_{34}, L_{35} \) have the following characteristics: “optimized linear model for validation” in place \( L_{33} \), “model for mapping in higher dimension” in place \( L_{34} \), and “model for multi-class problem” in place \( L_{35} \).

The token \( \beta_8 \) enters the net via place \( L_{37} \) with initial characteristics: “nonlinear \( \phi \)-functions that map the input on a higher dimensional space and kernel trick (choosing the kernel)”.

The transition \( Z_8 \) has the form:

\[
Z_8 = \{ L_{22}, L_{34}, L_{37}, L_{41} \}, \{ L_{36}, L_{39}, L_{40}, L_{41} \}, R_8, \lor ( L_{22}, L_{37}, L_{34} ), ( L_{41} )
\]

where:

\[
R_8 = \begin{bmatrix}
L_{22} & \text{false} & \text{false} & \text{false} & \text{true} \\
L_{34} & \text{false} & \text{false} & \text{false} & \text{true} \\
L_{37} & \text{false} & \text{false} & \text{false} & \text{true} \\
L_{41} & W_{41,38} & W_{41,39} & W_{41,40} & W_{41,41}
\end{bmatrix}
\]

and:

- \( W_{41,38} = “\text{The nonlinear classifier is constructed}” \)
- \( W_{41,39} = “\text{The nonlinear classifier for multi-class problem is constructed}” \).
• $W_{41,40} = \text{"The nonlinear classifier needs optimization"}$,
• $W_{41,41} = \neg (W_{41,38} \land W_{41,39} \land W_{41,40})$.

The $\alpha$-tokens that enter in the places $L_{38}$, $L_{39}$ and $L_{40}$ have the following characteristic:
“nonlinear model for validation” in place $L_{38}$, “nonlinear model for multi-class problem” in place $L_{39}$, and “nonlinear model for optimization” in place $L_{40}$.

The transition $Z_9$ has the form:
$$Z_9 = \langle \{L_{25}, L_{29}, L_{35}, L_{39}, L_{45}\}, \{L_{42}, L_{43}, L_{44}, L_{45}\}, R_9, \land (L_{25}, L_{29}, L_{35}, L_{39}, L_{45}) \rangle,$$
where:

<table>
<thead>
<tr>
<th>$L_{25}$</th>
<th>$L_{29}$</th>
<th>$L_{35}$</th>
<th>$L_{39}$</th>
<th>$L_{45}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>false</td>
<td>false</td>
<td>false</td>
<td>true</td>
<td></td>
</tr>
</tbody>
</table>

and:
• $W_{45,42} = \text{"The classifier have to be validated"}$,
• $W_{45,43} = \text{"The classifier have to be tested"}$,
• $W_{45,44} = \text{"The classifier is needed from additional processing"}$,
• $W_{45,45} = \neg (W_{45,42} \land W_{45,43} \land W_{45,44})$.

The $\alpha$-tokens that enter in places $L_{42}$, $L_{43}$ and $L_{44}$ have the following characteristic:
“multi-class model for validation” in place $L_{42}$, “multi-class model for testing” in place $L_{43}$, and “multi-class model for additional processing” in place $L_{44}$.

The transition $Z_{10}$ has the form:
$$Z_{10} = \langle \{L_{13}, L_{24}, L_{33}, L_{38}, L_{42}, L_{45}\}, \{L_{46}, L_{47}\}, R_{10}, \lor (L_{13}, L_{24}, L_{33}, L_{38}, L_{42}, L_{47}) \rangle,$$
where:

<table>
<thead>
<tr>
<th>$L_{13}$</th>
<th>$L_{24}$</th>
<th>$L_{33}$</th>
<th>$L_{38}$</th>
<th>$L_{42}$</th>
<th>$L_{47}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>false</td>
<td>true</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>$W_{47,46}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$W_{47,47}$</td>
</tr>
</tbody>
</table>

and:
• $W_{47,46} = \text{"The classifier is validated"}$,
• $W_{47,47} = \neg W_{47,46}$.

The $\alpha$-token that enters in place $L_{46}$ has the characteristic “validated model”.

36
The transition $Z_{11}$ has the form:

$$Z_{11} = \langle \{L_{12}, L_{46}, L_{49}\}, \{L_{48}, L_{49}\}, R_{11}, \lor (L_{12}, L_{46}, L_{49}) \rangle,$$

where:

$$R_{11} = \begin{array}{ccc}
L_{48} & L_{49} \\
\text{false} & \text{true} \\
L_{46} & \text{false} & \text{true} \\
L_{49} & W_{49,48} & W_{49,49}
\end{array}$$

and:

- $W_{49,48} = \text{“The classifier is tested”}$,
- $W_{49,49} = \neg W_{49,48}$.

The $\alpha$-token that enters in place $L_{48}$ has the characteristic “tested model”.

3 Implementation of the algorithm

The implementation of the technique is made using weather datasets [3, 17–19]. The database includes following attributes: Wind, Temperature, Outlook, Humidity, Smoke detector, Infra-red detector and Date/Time. The interest represents the cases when the Infra-red detector has value “yes”. Simple example for analysis based on weather data is given using RapidMiner software. Firstly, it is necessary to have a connection with the database. The preprocessing step is made using “Select Attributes” and “Nominal to Numerical” operators. The classification task demands to have a label attribute - this is the attribute “Infra-red detector”. The following steps are generating a model and its validation.

The kernel model is presented on the Fig. 3. Its accuracy and performance is shown in Fig. 4 and Fig. 5.

![Kernel Model](image1)

**Figure 3. The kernel model**

![Accuracy of the model](image2)

**Figure 4. The accuracy of the model**
4 Conclusion

In the presented paper is constructed a generalized net model for the process of classifying the datasets by support vector machine technique. The classifier is discussed in the cases of two classes and multi-class classification. The version(s) of linearly separable data and the non-linearly separable data is (are) explained. The extensions to the SVM model - soft margin classification and non-linear SVMs are presented. In addition to this article two generalized nets can be constructed: a generalized net for Multiclass SVMs algorithm or algorithms and a generalized net for the process of classification.

Acknowledgment

The first and the third authors are grateful for the support provided by the project DFNI-I-01/0006, funded by the National Science Fund, Bulgarian Ministry of Education, Youth and Science.

References


