

Evolving Support Vector Machine Parameters

Anh Tran Quang ⁽¹⁾, Qianli Zhang, Xing Li

Email ⁽¹⁾: qa00@mails.tsinghua.edu.cn

Department of Electronic Engineering

Tsinghua University, China

2002-11-4

Presentation at the 2002 International Conference on Machine Learning and Cybernetics

Contents

1. SVM's training parameters
2. Training model optimization
3. Evolutionary Algorithms
4. Experiments on Intrusion Detection dataset
5. Conclusion

1. SVM training parameters

- Generalization of SVM

1. $S_1 \subset S_2 \subset \dots \subset S$
 $S = \{Q(z, \alpha), \alpha \in \Lambda\}$

Kernel controls the flexibility of S

2. $Q\{z, \alpha_m^k\} \in S_k$
 $R(\alpha_m^k) \leq R_{emp}(\alpha_m^k) + \Omega(1/h_k)$

Upper bound C controls a tradeoff between the Minimal empirical risk And the confidence interval

1. SVM training parameters

- Base kernels

Polynomial: $K_{poly}(\mathbf{u}, \mathbf{v}) = (\sigma_1 * \mathbf{u} \cdot \mathbf{v} + r_1)^d$

RBF: $K_{rbf}(\mathbf{u}, \mathbf{v}) = \exp(-|\mathbf{u} - \mathbf{v}|^2 / \sigma_2)$

Sigmoid: $K_{sig}(\mathbf{u}, \mathbf{v}) = \tanh(\sigma_3 * \mathbf{u} \cdot \mathbf{v} + r_3)$

- Convex combination of kernels

$$K = \lambda_1 K_{poly} + \lambda_2 K_{rbf} + \lambda_3 K_{sig}$$

Where

$$0 \leq \lambda_1, \lambda_2, \lambda_3 \leq 1$$

$$\lambda_1 + \lambda_2 + \lambda_3 = 1$$

1. SVM training parameters

- SVM training model

$$V = \{\lambda_1, \lambda_2, \lambda_3, \sigma_1, r_1, d, \sigma_2, \sigma_3, r_3, C_-, C+\}$$

Where

$$0 \leq \lambda_1, \lambda_2, \lambda_3 \leq 1$$

$$\lambda_1 + \lambda_2 + \lambda_3 = 1$$

$$\sigma_1, r_1, d, \sigma_2, \sigma_3, r_3, C_-, C+ \geq 0$$

2. Training model optimization

- Generalization performance

	Label 1	Label 2
Label 1	A	B
Label 2	C	D

- **Error rate:** $\text{Err} = (B + C) / (A+B+C+D)$
- **Recall:** $\text{Rec} = A / (A+B)$
- **Precision:** $\text{Prec} = A / (A+C)$

2. Training model optimization

- Performance measures
 - Cross validation
 - Simple
 - Leave-one-out
 - Accurate but expensive to compute
 - $\xi\alpha$ -estimator [Joachims T.]
 - Estimate the upper bounds of error rate, recall and precision calculated by leave-one-out method
 - *Need to train just one time.*

2. Training model optimization

- SVM generalization ability

$$G = \rho_1(1 - \text{Err}_{\xi_{\alpha}}) + \rho_2 \text{Rec}_{\xi_{\alpha}} + \rho_3 \text{Prec}_{\xi_{\alpha}}$$

Where

$$\rho_1 + \rho_2 + \rho_3 = 1$$

$$\rho_1, \rho_2, \rho_3 \geq 0$$

2. Training model optimization

- Training model optimization problem

$$\text{Max } G(1 - \text{Err}_{\xi_{\alpha}}(V), \text{Rec}_{\xi_{\alpha}}(V), \text{Prec}_{\xi_{\alpha}}(V))$$

Subject to:

$$V = \{\lambda_1, \lambda_2, \lambda_3, \sigma_1, r_1, d, \sigma_2, \sigma_3, r_3, C-, C+\}$$

$$0 \leq \lambda_1, \lambda_2, \lambda_3 \leq 1$$

$$\lambda_1 + \lambda_2 + \lambda_3 = 1$$

$$\sigma_1, r_1, d, \sigma_2, \sigma_3, r_3, C-, C+ \geq 0$$

3. Evolutionary Algorithms

1. Initialize the population at random
2. Evaluate each individual in the population
3. Save optimal information
4. Update the population
5. Repeat 2 to 4 until 'terminal criterion' is satisfied

4. Experiment

- Data source
 - Network-based attacks
 - Normal and attack-related TCP connections
 - DARPA IDS Evaluation, MIT-Lincoln Lab

4. Experiment

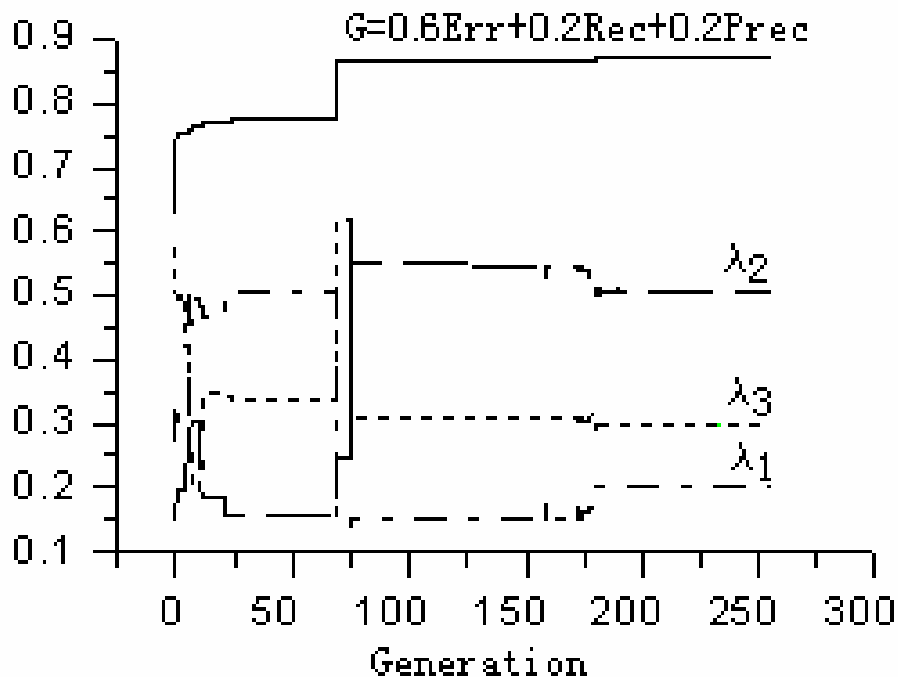
- Feature selection
 - Packet number
 - Time slots
 - Traffic flows
 - Statistics

4. Experiment

- Source Code
 - SVM: A modification of LIBSVM
 - EA: Genetic Algorithm
 - Platform: PC/Linux

4. Experiment

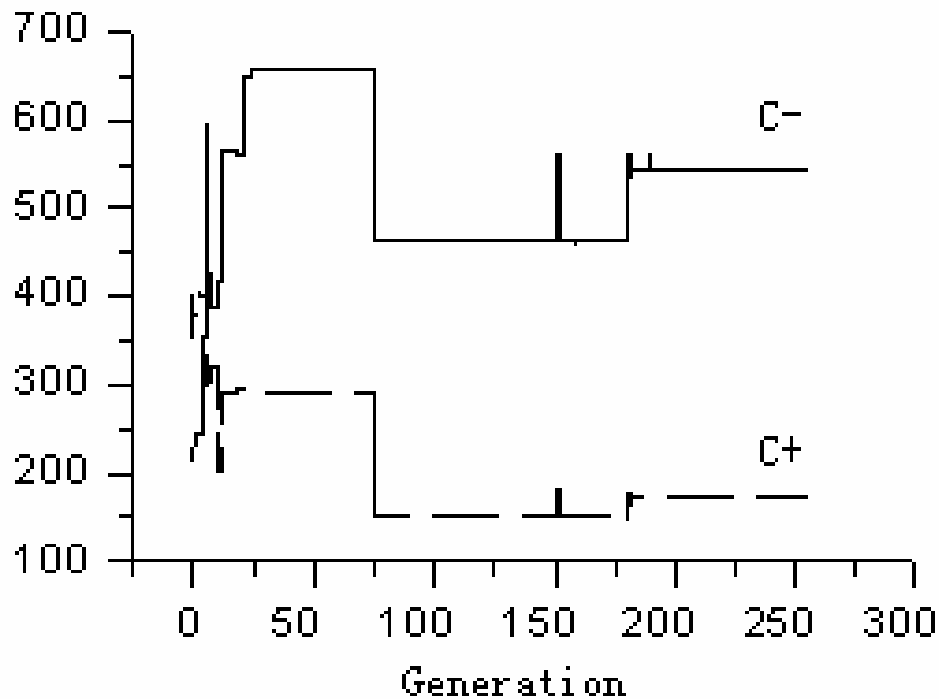
- Evolutionary result (a)



- ✓ The generalization performance is optimized.
- ✓ All parameters converge after 100 criterions.

4. Experiment

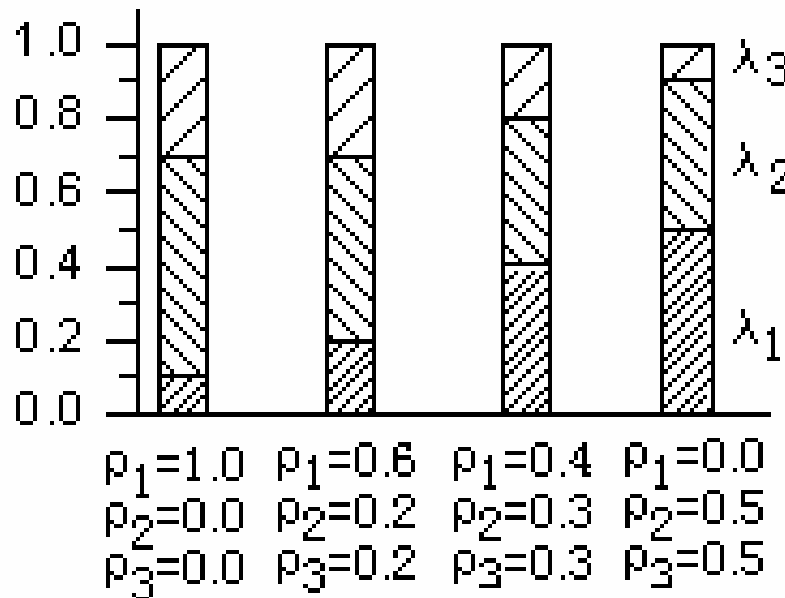
- Evolutionary result (b)



✓ The ratio of C+ and C- seems to constant.

4. Experiment

- Base kernel proportion



- ✓ The proportion of RBF kernel is the highest.
- ✓ Polynomial kernel has ability to improve the recall and precision
- ✓ Sigmod kernel has ability to improve the error rate.

5. Conclusion

- The generalization performance is improved.
- The ratio of C^+ and C^- seems to be constant.
- Each type of base kernel has different ability to improve the generalization performance.
- This method can be used for continuous adaptation to circumstances.

Thank you !

Anh Tran Quang ⁽¹⁾, Qianli Zhang, Xing Li

Email ⁽¹⁾: qa00@mails.tsinghua.edu.cn

Department of Electronic Engineering

Tsinghua University, China

2002-11-4