

Chinese Spam Detection Based on One-class SVM

Donghong Sun, Quang-Anh Tran, Haixin Duan,
Guodong Zhang

CERNET Computer Emergency and Response Team
& Shenyang Institute of Aeronautical Engineering

Email: qa_at_cernet.edu.cn

17 Aug 2004

Presentation at the International Symposium on Computing and
Information, ISC&I 2004

Contents

- Anti-spam techniques
- Chinese spam detection
- One-class Support Vector Machines
- Spam detection by Ham model
- Ham detection by Spam model
- Conclusions

Anti-spam techniques

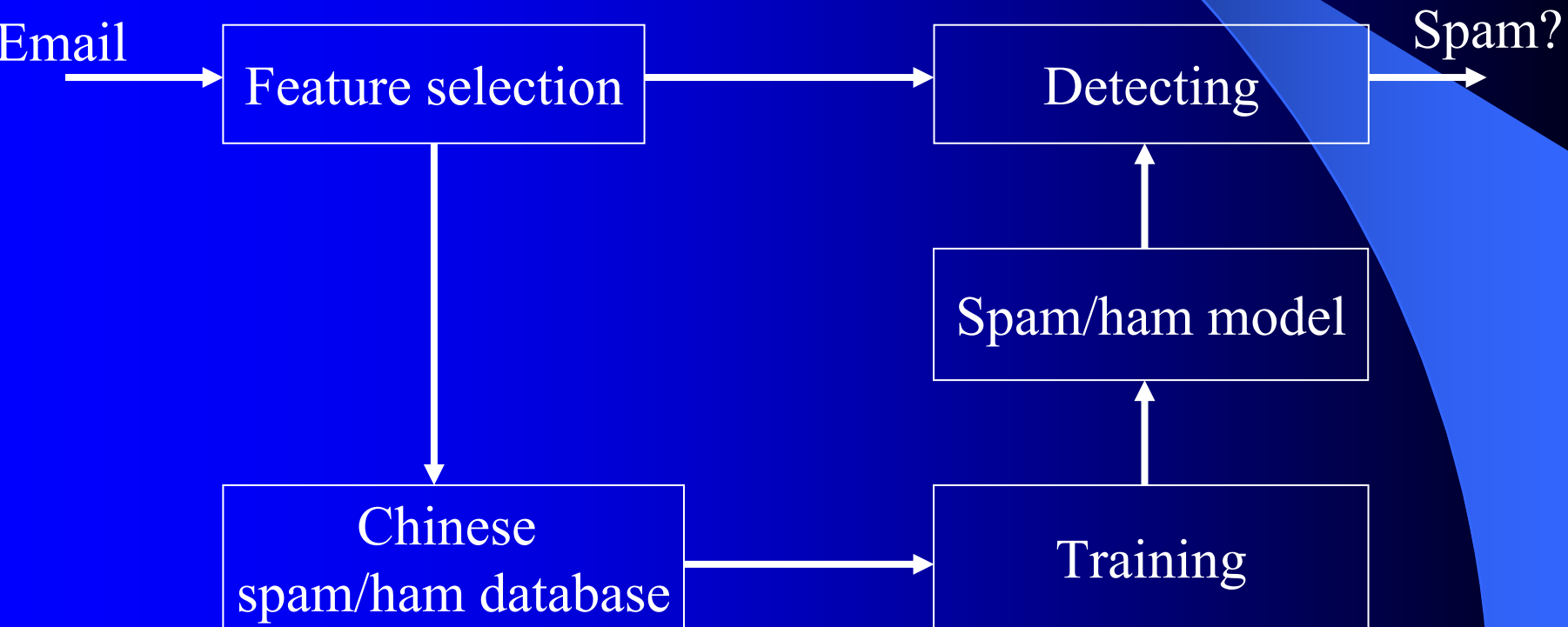
- Anti-spam model
 - Protection (Open relay, charge, law...)
 - Detection
 - Response (Mark, filter, indict...)

Anti-spam techniques

- Spam detection
 - Non-statistical
 - RBL
 - Caller ID
 - Pattern Matching
 - Statistical
 - Ham/Spam Classification (Bayes, NN, SVM...)
 - Ham detection by spam model
 - Spam detection by ham model

Chinese spam detection

- Framework



Chinese spam detection

- Feature selection
 - Decoding
 - Base64
 - QuotedPrint
 - Chinese word segmentation
 - Chinese word dictionary based
 - Right to left
 - Maximum matching
 - TF-IDF formula

$$TF(w_j, d_i)$$

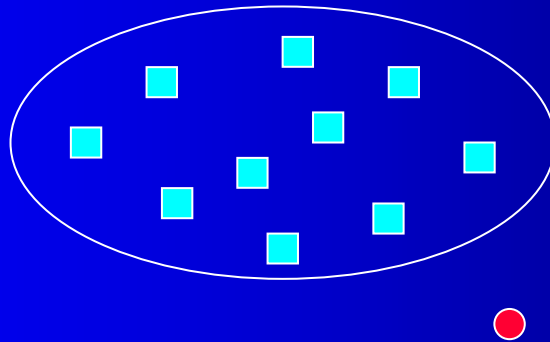
$$IDF(w_j) = \log\left(\frac{n}{2F(w_j)}\right)$$

Chinese spam detection

- Chinese spam/ham database
 - Ham: CCERT working mailinglist (2000)
 - Spam: CCERT spam report (1000)

One-class Support Vector Machines

- Concept
 - Estimating the support of a distribution



One-class Support Vector Machines

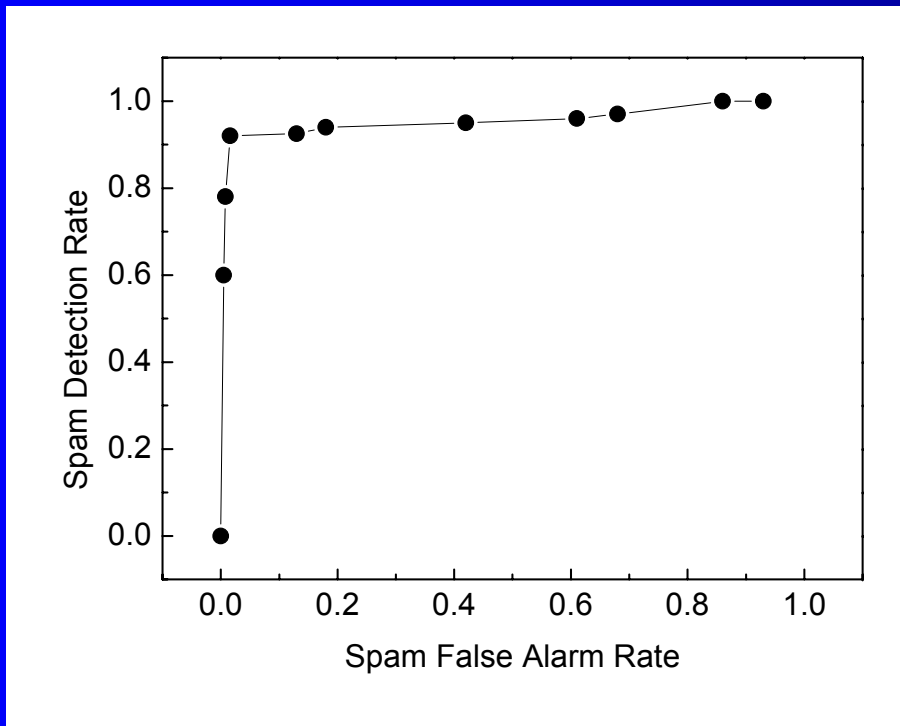
- Key techniques
 - Kernel method
 - RBF, σ
 - Maximal margin hyperplane
 - Data separated from the origin
 - Trading off empirical risk and complexity
 - $0 \leq \nu \leq 1$
- Source code
 - *Svm^{light}*
 - LIBSVM

One-class Support Vector Machines

- Training model selection
 - M: Training model, (σ, ν)
 - G: Generalization performance
 - $\max_M G(M)$
 - $\xi\alpha\rho$ - estimate
 - Genetic Algorithms

Spam detection by ham model

- ROC curve for spam detection



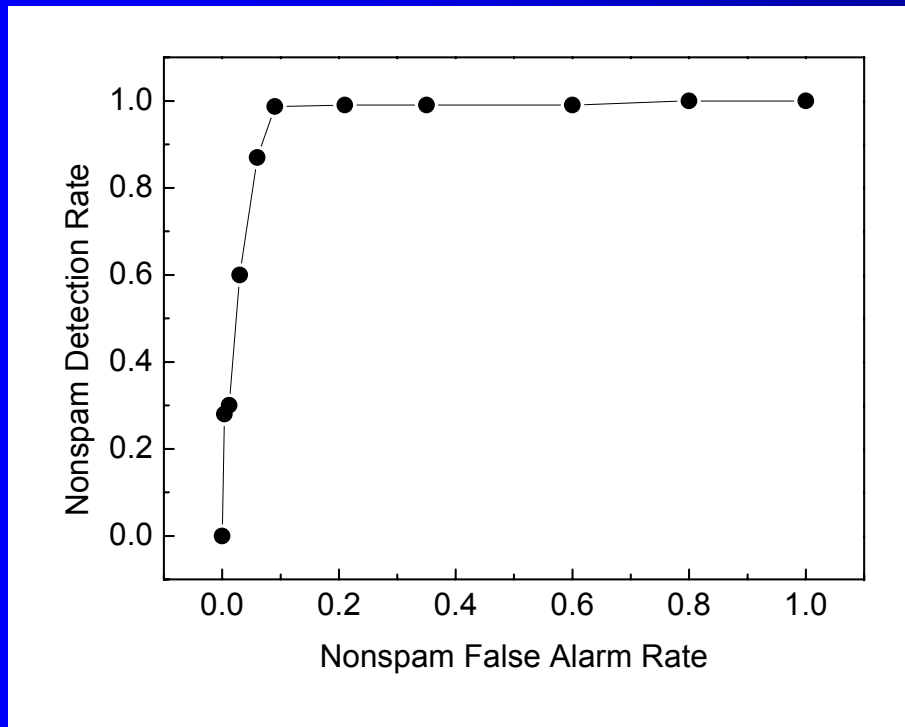
The best working point:

False alarm rate: 0.016

Detection rate: 0.92

Ham detection by spam model

- ROC curve for ham detection



The best working point:

False alarm rate: 0.09

Detection rate: 0.987

Conclusions

- Feature selection: Header and Body
- Non statistical and statistical features
- Using only Header features
- Spam detection by ham model
- Ham detection by spam model

References

- Schölkopf, B., Platt, J, et al. 2001. Estimating the support of a high-dimensional distribution. Neural Computation, 2001.
- Joachims T., Text Categorization with Support Vector Machines: Learning with Many Relevant Features. European Conference on Machine Learning, 1998.
- Tran, Q.A., Zhang Q., Li X. Evolving Training Model Method for One-class SVM. IEEE International Conference on Systems, Man & Cybernetics, 2003.
- LIBSVM, URL: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- CCERT, URL: <http://www.ccert.edu.cn>
- SpamAssassin, URL: <http://www.spamassassin.org>

Thank you!

Quang-Anh Tran

17 Aug 2004