Silesian University of Technology

Memetic evolution of training sets with adaptive radial basis kernels for support vector machines

Jakub Nalepa, Wojciech Dudzik and Michal Kawulok

{jakub.nalepa, wojciech.dudzik, michal.kawulok}@polsl.pl

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

European Funds
Republic of Poland
European Union
European Social Fund

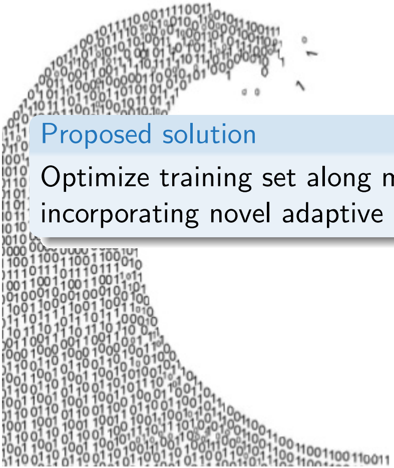
Presentation outline

What are the problems

Main problems

- High computational complexity of training $O(t^3)$
- High memory complexity of training $O(t^2)$
- Need to fine-tune the models with hyperparameters
- Growing number of features
- Bigger datasets
- Classification time linearly depends on number of SV ($O(S)$)

Solution



Proposed solution

Optimize training set along model optimization by incorporating novel adaptive radial basis function kernel.

Proposed method

Adaptive RBF kernel

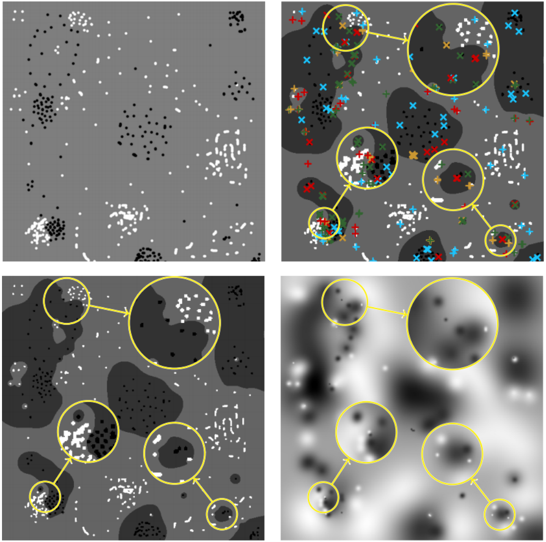
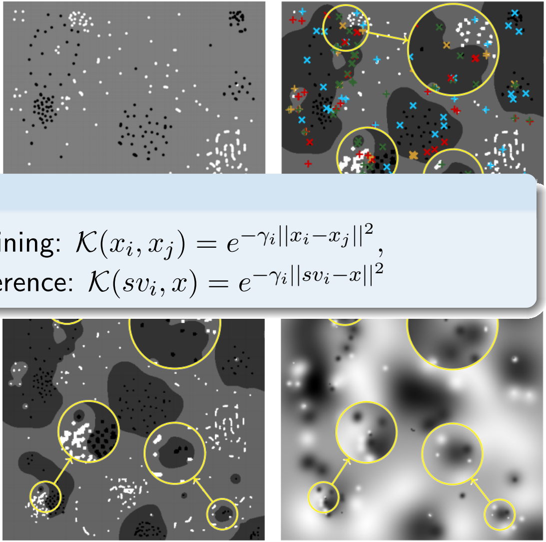


Figure: Assigning different γ's in the RBF kernel to different $\boldsymbol{T}$ vectors can help better "model" the SVM hyperplane

Adaptive RBF kernel



Figur
differ
RBF
differ
can h
"model" the SVM
hyperplane

Kernel function

$$\text{For training: } \mathcal{K}(x_i, x_j) = e^{-\gamma_i ||x_i - x_j||^2},$$

$$\text{For inference: } \mathcal{K}(sv_i, x) = e^{-\gamma_i ||sv_i - x||^2}$$

Memetic algorithms

Algorithm 1 Memetic evolution of SVM training sets.

Select $\vec{\gamma}$
for all γ_i in $\vec{\gamma}$ **do** ▶ γ's are sorted (ascendingly)
    $P, P'_{\text{best}} \leftarrow$ Generate population(N, C, γ_i, $\boldsymbol{T}$)
    if $\eta(P'_{\text{best}}) > \eta(P_{\text{best}})$ **then**
        $P'_{\text{best}} \leftarrow$ RunEvolution()
        if $\eta(P'_{\text{best}}) > \eta(P_{\text{best}})$ **then**
            Add $\mathrm{SV}(P'_{\text{best}})$ to $\mathcal{S}_{\text{best}}$
            $\boldsymbol{T} \leftarrow$ Shrink($\boldsymbol{T}$, P)
            $P_{\text{best}} \leftarrow P'_{\text{best}}$
        end if
    end if
end for
return P_{best}

Memetic algorithms

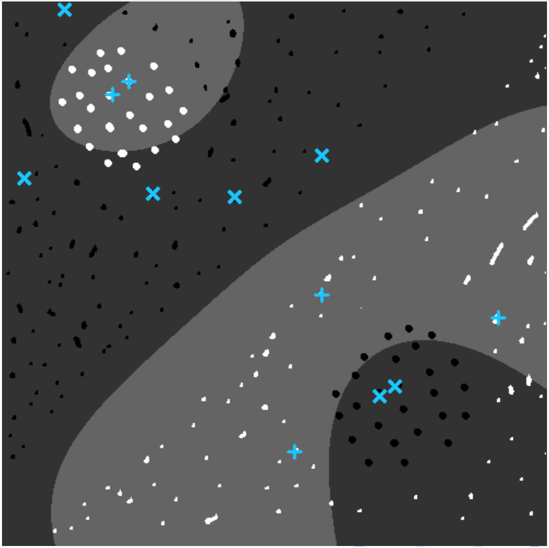


Figure: The best solution from initial population.

Memetic algorithms

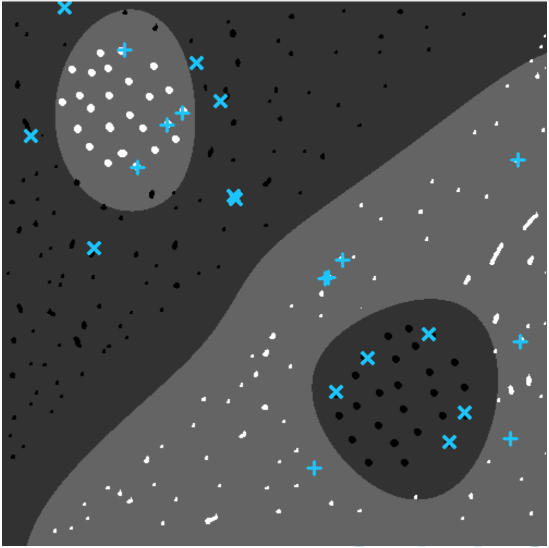


Figure: The best solution after finishing evolution with first γ from $\vec{\gamma}$.

Memetic algorithms

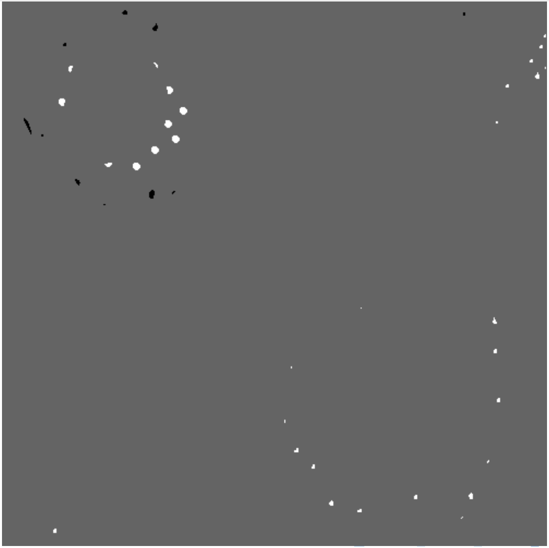


Figure: Shrinked training set that will be used in next iteration with subsequent γ. Shrinking procedure is based on whole population.

Memetic algorithms

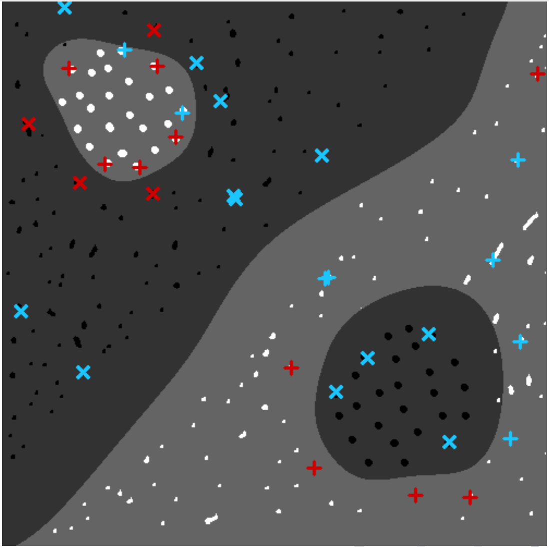


Figure: Solution after second evolution has ended. Added new support vectors marked with red color crosses.

Memetic algorithms

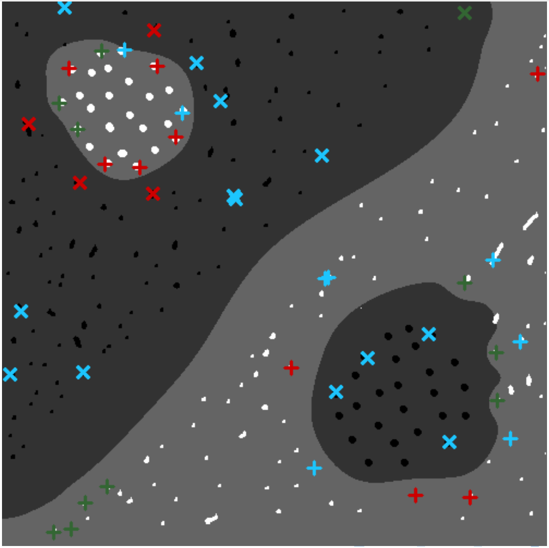
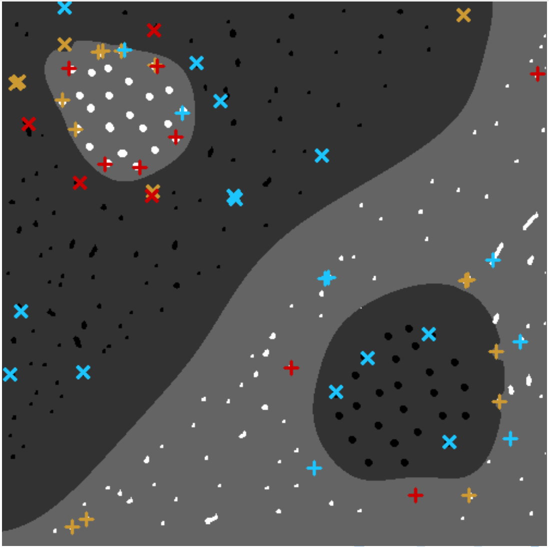


Figure: Adding next γ value marked with green vectors provided worse classification performance, these support vectors will be removed.

Memetic algorithms

Figure: Final solution for given dataset containing three different γ values.



Experiments

Implementation details

Setup:

- Windows 10 machine equipped with i9-7900X CPU and 64 GB of RAM

Settings:

- Objective function AUC with thresholding for optimal accuracy over validation set
- 5-fold cross-validation
- All results shows average from 50 runs (10 times for each fold),
- Folds contains training, validation and test set in 3:1:1 proportion, respectively

Table: The results of all methods. The best results are boldfaced.

Algorithm	AUC	Acc	F1	Pr	Re	MCC
GNB	0.839	0.736	0.471	0.454	**0.708**	0.397
LR	0.862	0.912	0.599	0.691	0.567	0.545
k-NN(3)	0.857	0.930	0.674	0.729	0.644	0.626
k-NN(5)	0.872	0.930	0.632	0.716	0.591	0.590
k-NN(7)	0.879	0.928	0.591	0.695	0.548	0.554
SVM(Linear)	0.848	0.912	0.584	0.648	0.560	0.528
SVM(Poly)	0.869	0.924	0.632	0.717	0.605	0.592
SVM(RBF)	0.807	0.877	0.201	0.303	0.198	0.162
MASVM	0.880	0.936	0.651	0.763	0.605	0.622
MASVM(MG)	0.877	0.936	0.676	0.793	0.617	0.644
Ours	**0.888**	**0.941**	**0.696**	**0.810**	0.640	**0.667**

Table: The ranking test over MCC (together with the statistical importance of the differences between our MA and the corresponding approach), for various IR ranges. The meanings of ns, *, **, ***, and ****: $p > 0.05$, $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$, and $p \leq 0.0001$. The best results are boldfaced.

Algorithm	Q1	Q2	Q3	Q4	All
GNB	7.63^{**}	9.04^{***}	8.58^{****}	6.50^{****}	7.94^{****}
LR	6.08^{ns}	5.79^{ns}	5.75^{ns}	5.29^{ns}	5.73^{****}
k-NN(3)	5.33^{ns}	4.88^{ns}	4.67^{ns}	5.50^{ns}	5.09^{*}
k-NN(5)	5.50^{ns}	5.21^{ns}	5.67^{ns}	5.67^{ns}	5.51^{***}
k-NN(7)	5.21^{ns}	5.63^{ns}	6.58^{**}	6.25^{**}	5.92^{****}
SVM(Linear)	6.67^{ns}	5.63^{ns}	6.42^{**}	5.71^{**}	6.10^{****}
SVM(Poly)	5.92^{ns}	4.88^{ns}	5.38^{ns}	4.75^{ns}	5.23^{**}
SVM(RBF)	7.92^{**}	10.71^{****}	9.83^{****}	8.13^{****}	9.15^{****}
MASVM	5.17^{ns}	4.92^{ns}	4.79^{ns}	5.04^{ns}	4.98^{ns}
MASVM(MG)	4.75^{ns}	4.46^{ns}	3.83^{ns}	4.00^{ns}	4.26^{ns}
Ours	**3.75**	**4.42**	**2.92**	**3.21**	**3.57**

Conclusions and outlook

Conclusions

- Our technique outperforms SVMs optimized using other evolutionary methods and other supervised learners with grid-searched hyperparameters.
- It delivers consistent results across sets of various characteristics.
- Our technique can be easily applied in imbalanced classification, where it outperformed all other methods.
- Assigning different γ's to different training vectors is useful in heterogeneous ("difficult") parts of the input space, as visually shown for our synthetic datasets.

Thank you

{jakub.nalepa, wojciech.dudzik, michal.kawulok}@polsl.pl

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

European Funds
Republic of Poland
European Union
European Social Fund
AIDA

ICPR 2020, Milan January 10-15, 2021