Evaluation of kernel function modification in text classification using SVMs

Yangzhe Xiao

Motivation

- One of the limitations of SVMs lies in the choice of kernels.
- Good choices of kernels can make a linearly inseparable case become separable in feature space or increase the margin in feature space.

Kernel Modification

- Christianini, Shawe-Taylor & Lodhi (2001) used Latent Semantic Indexing (LSI) to extract the semantic relation matrix ***M*,** and modified radial basis kernel (RBF) as

$$K(x,y)=\exp(-\|\boldsymbol{M}x-\boldsymbol{M}y\|^2/2\sigma^2) .$$

--Documents are implicitly mapped into a "semantic space".

--Better results.

<u>This study is motivated by the same modification idea but with different *M*.</u>

Measurement of Discriminating Power

- For a term *t*
  and a category *c*,

	c	*Non-c*
t	*A*	*B*
No-t	*C*	*D*

$$\chi^2(t,c) = \frac{N \times (AD - CB)^2}{(A+C)\times(B+D)\times(A+B)\times(C+D)}$$

$N=A+B+C+D$.

$\chi^2(t,c)$ is zero if *t* and *c* are independent.

Measurement of Discriminating Power(cont'd)

- For a term *t*
  and a category *c*,

	c	*Non-c*
t	*A*	*B*
No-t	*C*	*D*

$$E(t,c) = \frac{A}{A+C}(1-(-p\log_2 p - q\log_2 q))$$

where $$p = \frac{\frac{A}{A+C}}{\frac{A}{A+C} + \frac{B}{B+D}}$$

$q=1-p$

Matrix *M*

- Assign weights for terms according to the two measurements $\chi^2(t,c)$ or $E(t,c)$ by modifying kernel.
- Let M be a square diagonal matrix with values of $(1+\chi^2(t,c))$ or $(1+E(t,c))$ for every terms in the feature set.

Documents

- Training and testing documents are from Reuters-21578 and have 10 categories.
- Precision (π) and recall (ρ) are defined as:

$$\pi_i = \frac{TP_i}{TP_i + FP_i} \qquad \rho_i = \frac{TP_i}{TP_i + FN_i}.$$

Here, FP_i (false positives wrt category c_i) is the number of test documents incorrectly classified under c_i; TN_i (true negatives wrt c_i), TP_i (true positives wrt c_i), and FN_i (false negative wrt c_i) are defined accordingly.

Results

SampleSize	T	CW	EW	TS
20	0.185	0.211	**0.223**	0.197
100	0.574	0.602	**0.613**	0.593
250	0.690	0.726	**0.728**	0.717
500	0.779	0.790	**0.791**	0.788
1000	0.810	0.819	**0.820**	0.814
1880	0.837	**0.855**	0.851	**0.855**
5100	0.830	0.861	0.867	**0.880**

$$F = \frac{2\pi\rho}{\pi + \rho}$$

F depends equally on precision π and recall ρ.

The End