SVM Classification Using Sequences of Phonemes and Syllables

Gerhard Paaß[1], Edda Leopold[1], Martha Larson[2],
Jörg Kindermann[1], and Stefan Eickeler[2]

[1] Fraunhofer Institute for Autonomous Intelligent Systems (AIS)
53754 St. Augustin, Germany
[2] Fraunhofer Institute for Media Communication (IMK)
53754 St. Augustin, Germany

Abstract. In this paper we use SVMs to classify spoken and written documents. We show that classification accuracy for written material is improved by the utilization of strings of sub-word units with dramatic gains for small topic categories. The classification of spoken documents for large categories using sub-word units is only slightly worse than for written material, with a larger drop for small topic categories. Finally it is possible, without loss, to train SVMs on syllables generated from written material and use them to classify audio documents. Our results confirm the strong promise that SVMs hold for robust audio document classification, and suggest that SVMs can compensate for speech recognition error to an extent that allows a significant degree of topic independence to be introduced into the system.

1 Introduction

Support Vector Machines (SVM) have proven to be fast and effective classifiers for text documents [6]. Since SVMs also have the advantage of being able to effectively exploit otherwise indiscernible regularities in high dimensional data, they represent an obvious candidate for spoken document classification, offering the potential to effectively circumvent the error-prone speech-to-text conversion. If optimizing spoken document classification performance is not entirely dependent on minimizing word error rate from the speech recognition component, room becomes available to adjust the interface between the speech recognizer and the document classifier.

We are interested in making the spoken document classification system as a whole speaker and topic independent. We present the results of experiments which applied SVMs to a real-life scenario, classifying radio documents from the program Kalenderblatt of the Deutsche Welle radio station. One striking result was that SVMs trained on written texts can be used to classify spoken documents.

T. Elomaa et al. (Eds.): PKDD, LNAI 2431, pp. 373–384, 2002.

2 SVM and Text Document Classification

Instead of restricting the number of features, support vector machines use a refined structure, which does not necessarily depend on the dimensionality of the input space. In the *bag-of-words-representation* the number of occurrences in a document is recorded for each word. A typical text corpus can contain more than 100,000 different words with each text document covering only a small fraction. Joachims [6] showed that SVMs classify text documents into topic categories with better performance than the currently best-performing conventional methods. Similar results were achieved by Dumais et al. [2] and Drucker et al. [1].

Previous experiments [9] have demonstrated that the choice of kernel for text document classification has a minimal effect on classifier performance, and that choosing the appropriate input text features is essential. We assume that this extends to spoken documents and chose basic kernels for these experiments, focusing on identifying appropriate input features. Recently a new family of kernel functions — the so called string kernels — has emerged in the SVM literature. They were independently introduced by Watkins [13] and Haussler [5]. In contrast to usual kernel functions these kernels do not merely calculate the inner product of two vectors in a feature space. They are instead defined on discrete structures like sequences of signs. String kernels have been applied successfully to problems in the field of bio-informatics [10] as well as to the classification of written text [11].

To facilitate classification with sub-word units one can generate n-grams which may take the role of words in conventional SVM text classification described above [Leo02][Joa98][Dum98]. Lodhi et al. [11] used subsequences of characters occurring in a text to represent them in a string kernel. The kernel is an inner product in the feature space consisting of all subsequences of length k, i.e. ordered sequences of k characters occurring in the text though not necessarily contiguously. The subsequences are weighted by an exponentially decaying factor of their full length in text, hence emphasizing those sequences which are close to contiguous. In contrast to our approach they use no classification dependent selection or weighting of features.

We use subsequences of linguistic units — phonemes, syllables or words — occurring in the text as inputs to a standard SVM. We only use contiguous sequences not exceeding a given length. Our approach is equivalent to a special case of the string kernel. Since the focus of this paper is on the representation of spoken documents we go beyond the original string kernel approach insofar as we investigate building strings from different basic units. We employ the word "n-gram" to refer to sequences of linguistic units and reserve the expression "kernel" for traditional SVM-kernels. The kernels that we use in the subsequent experiments are the linear kernel, the polynomial of degree 2 and the Gaussian RBF-kernel. Our experiments consist of 1-of-n classification tasks. Each document is classified into the class which yields the highest SVM-score.

3 Sub-word Unit Speech Recognition

Continuous speech recognition systems (CSR) integrates two separately trained models each capturing a different level of language regularities. The *acoustic model* generates phoneme hypotheses from the acoustic signal whereas the *language model* constrains phoneme sequences admissible in the language. Errors made by a CSR can be roughly attributed to one or the other of these models. The acoustic models are responsible for errors occurring when phonemes in the input audio deviate in pronunciation from those present in the training data or when other unexpected acoustics, such as coughing or background noise from traffic or music intervene. The language model is responsible for errors due to words occurring in the audio input that either were missing or inappropriately distributed in the training data. Missing words are called OOV (Out of Vocabulary) and are a source of error even if the language model includes a 100,000 word vocabulary. A language model which is base! d on sub-word units like syllables rather than on words helps eliminate OOV error and makes possible independence from domain specific vocabularies. A syllable based language model, however, introduces extra noise on the syllable level because of errors due to combinations that would not have been part of the search space in a CSR with a word based language model. So there is a trade off between generality and accuracy of a CSR.

4 Combining SVMs and Speech Recognition

The representation of documents blends two worlds. From the linguistic point of view, texts consist of words which bear individual meaning and combine to form larger structures. ¿From an algorithmic point of view, a text is a series of features which can be modeled by making certain approximations concerning their dependencies and assuming an underlying statistical distribution.

When it comes to the classification of *spoken documents*, the question of the appropriate text features becomes difficult, because the interplay between the speech processing system and the classification algorithm has to be considered as well. Our assumption is that it is desirable to give the SVM fine-grained classes of linguistic elements, and have it learn generalizations over them, rather than try to guess which larger linguistic classes might carry the information which would best assist the classifier in its decision.

Large vocabulary CSR systems were original developed to perform pure transcription and they were optimized to output word for word the speech of the user. Under such a scenario, a substitution of the word 'an' for 'and' — a very difficult discrimination for the recognizer — would be counted as an error. Under a spoken document classification scenario, the effects of the substitution would be undetectable. If recognizer output is to be optimized for spoken document classification instead of transcription, non-orthographic units become an interesting alternative to words.

The potential of sub-word units to enhance the domain independence of a spoken document retrieval system is well documented in the literature. One of

the first systems to experiment with sub-word units, used vectors composed of sub-word phoneme sequences delimited by vowels. In this system the sub-word units, although shorter, perform marginally better than words. At the lowest sub-word level experimenters have used acoustic feature vectors and phonemes. In [4] N-gram topic models are built using such features, and incoming speech documents are classified as to which topic model most likely generated them. A concise overview of the literature on sub-words in speech document retrieval is given in [8].

For spoken document classification we decided that syllables and phoneme strings provide the best potential as text features. Since SVMs are able to deal with high dimensional inputs, we are not obliged to limit the number of input features. The idea is that short-ranged features such as short phoneme strings or syllables, will allow SVM to exploit patterns in the recognition error and indirectly access underlying topic features. Long-ranged features such as longer phoneme strings and syllable bi- and tri-grams will allow the SVM access to features with a higher discriminative value, since they are long enough to be semantically very specific.

5 The Data

In order to evaluate a system for spoken document classification, a large audio document collection annotated with classes is required. It is also necessary to have a parallel text document collection consisting of literal transcriptions of all the audio documents. Classification of this text document collection provides a baseline for the spoken document system.

The Deutsche Welle *Kalenderblatt* data set consists of 952 radio programs and the parallel transcriptions from the Deutsche Welle *Kalenderblatt* web-page http://www.kalenderblatt.de. Although the transcriptions are not perfect, they are accurate enough to provide a text classification baseline for spoken document classification experiments. The transcriptions were decomposed into syllables for the syllable experiments and phonemes for the phoneme based experiments using the transcription module of the BOSSII system [7].

Each program is about 5 minutes long and contains 600 running words. The programs were written by about 200 different authors and are read by about 10 different radio reporters and are liberally interspersed with the voices of people interviewed. This diversity makes the Deutsche Welle *Kalenderblatt* an appealing resource since it represents a real world task. The challenge of processing these documents is further compounded by the fact that they are interspersed with interviews, music and other background sound effects.

In order to train and to evaluate the classifier we needed topic class annotations for all of the documents in the data set. We chose as our list of topics the International Press Telecommunications Council (IPTC) subject reference system. Annotating the data set with topic classes was not straightforward, since which topic class a given document belongs to is a matter of human opinion. The

Fig. 1. Agreement of human Annotators in classifying the Kalenderblatt Documents

DW Kalenderblatt data from year	top choice of both annotators the same	one annotator choosing top others less	complete disagreement between annotators
1999	67 %	22 %	11 %
2000	74 %	17 %	9 %
2001	70 %	10 %	20 %

agreement of the human annotators about the class of documents represents an upper bound for the performance of the SVM classification system (table 1).

6 Experiment: Design and Setup

In the standard bag-of-words approach texts are represented by their type-frequency-vectors. Here we examine the usefulness of type-frequency-vectors constructed from the following linguistic units: word-forms of the written text, syllables derived from written text using BOSSII, phonemes derived from written text using BOSSII, syllables obtained from spoken documents by CSR, and phonemes obtained from spoken documents by CSR.

To derive syllables and phonemes from *written text* we use the transcription module of the second version of the Bonn Open Source Synthesis System (BOSSII) developed by the Institut für Kommunikationsforschung und Phonetik of Bonn University to transform written German words into strings of phonemes that represent their pronunciations and their syllable decompositions. A more detailed description of this system is given in [7].

In order to obtain Phonemes and Syllables the *spoken documents* We used the simplest acoustic models — monophone models — which have been trained on a minimal amount of generic audio data and have not been adapted to any of the speakers in the corpus. Additionally, we train a simple bigram model as the language model for the speech recognition using data from a completely different domain. We use syllables as the basic unit for recognition. System tests showed that the syllable recognition accuracy rate hovers around 30% for this configuration. Phonemes of the spoken documents are drawn from the syllable transcripts by splitting the syllables into their component phonemes parts.

As there is a large number of possible n-grams in the text we used statistical test to eliminate unimportant ones. First we required that each term must occur at least twice in the corpus. In addition we check the hypothesis that there is a statistical relation between the document class and the occurrence of a term w_k. Let $f(w_k, y)$ denote the number of documents of class y containing term w_k and let N_1 and N_{-1} be the number of documents of class 1 or -1 respectively. Then we obtain the table

	number of documents where ...	
	class $y = 1$	class $y = -1$
w_k in document	$f(w_k, y = 1)$	$f(w_k, y = -1)$
w_k not in document	$N_1 - f(w_k, y = 1)$	$N_{-1} - f(w_k, y = -1)$

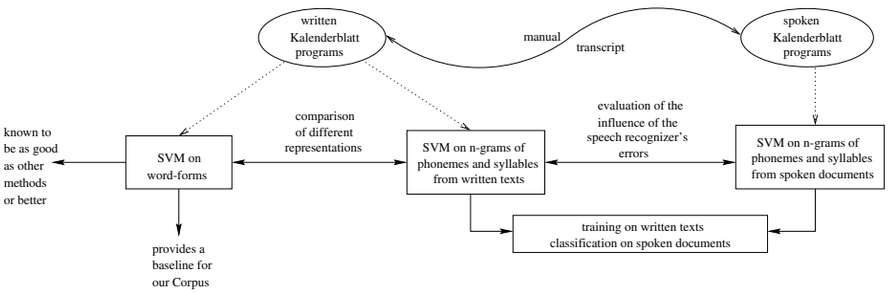

Fig. 2. The logical structure of our experimental design

If the rows and columns were independent then we would have $f(w_k, y = 1) = N * p(w_k) * p(y = 1)$ where $p(w_k)$ is the probability that w_k occurs in a document and $p(y = \pm 1)$ is the probability that $y = \pm 1$. We may check by a significance test if the first table originates from the distribution which obeys the independence assumption. We use a Bayesian version of the likelihood ratio test assuming a Dirichlet prior distribution. The procedure is discussed in [3]. The resulting test statistic is used to perform a preliminary selection of promising input terms to reduce the number of many thousand inputs. In the experiments different threshold values for the test statistic are evaluated.

We consider the task of deciding if a previously unknown text belongs to a given category or not. Let c_{targ} and e_{targ} respectively denote the number of correctly and incorrectly classified documents of the target category $y = 1$ and let e_{alt} and let c_{alt} be the same figures for the alternative class $y = -1$. We use the *precision* $prec = c_{targ}/c_{targ} + e_{alt}$ and the *recall* $rec = c_{targ}/c_{targ} + e_{targ}$ to describe the result of an experiment. In a specific situation a decision maker has to define a loss function and quantify the cost of misclassifying a target document as well as a document of the alternative class. The F-measure is a compromise between both cases [12]

$$F_{val} \quad = \quad \frac{2}{\frac{1}{prec} + \frac{1}{rec}}, \tag{1}$$

If recall is equal to precision then F_{val} is also equal to precision and recall.

7 Experiments with the Kalenderblatt Corpus

Our corpus poses a quite difficult categorization task. This can be seen from figure 1 where the discrepancies of the different human annotators are shown. If we assume that one annotator provides the "correct" classification then precision and recall of the other annotator is about 70%. As the final classification was defined by a human this is some upper limit of the possible accuracy that can be achieved. In our experiments we compare the properties of three different representational aspects and their effect on classification performance: (1)

Representation of a document by words. (2) Representation by simple terms or n-grams of terms, where 'non-significant' n-grams are eliminated. (3) Terms generated from the written representation or terms produced by CSR. As all representations are available for the same documents this allows to compare the relative merits of the representations. The setup is shown in figure 2.

We used five-fold cross-validation to get large enough training sets. As the F-value seems to be more stable because it is not affected by the tradeoff between recall and precision we use it as our main comparison figure. We utilized the SVM^{light} package developed by Joachims [6]. We performed experiments with respect to two topic categories: 'politics' of about 230 documents and 'science' with about 100 documents. This should give an impression of the possible range of results. Experiments with smaller categories led to unsatisfactory results. In preliminary experiments RBF-kernels turned out to be unstable with fluctuating results. We therefore concentrated on linear kernels.

7.1 Experiments with Written Material

We observed as a general tendency that precision increases and recall decreases with the size on n-grams. This can be explained by the fact, that longer linguistic sign-aggregates have a more specific meaning than shorter ones.

As can be seen in the upper part of table 1 topic classification using simple words starts with an F-value of 67.6% and 60.5% for 'politics' and 'science' respectively.

For both classes the syllables yield better results than words. For 'politics' syllables reach an F-value of 71.4% which is 3.8% better than the best word figure. There is a gain by using n-grams instead of single syllables which nevertheless reach an F-value of 70.1%. Longer n-grams ($n = 5, 6$) reduce accuracy. This can be explained by their low frequency of occurrence.

For the smaller category 'science' there is a dramatic performance increase to an $F_{val} = 73.1\%$ compared to an optimal 60.5% for words. Here n-grams perform at least 8.7% worse than simple terms, perhaps as they are more affected by the relatively large noise in estimating syllable counts. The good performance of syllables again may be explained by more stable estimates of their frequencies in each class. It is interesting that in the larger 'politics' class n-grams work better in contrast to the smaller 'science' class.

The best results are achieved for phonemes. For 'politics' there is no significant difference F-values compared to syllables, whereas for the small category 'science' there is again a marked increase to an F-value of 76.9% which is 3.8% larger than for syllables. The average length of German syllables is 4 to 5 phonemes, so phoneme trigrams in average are shorter and consequently more frequent than syllables. This explains the high F-value of phoneme trigram in the small category. Note that for both categories we get about the same accuracy which seems to be close to the possible upper limit as discussed above.

The effect of the significance threshold for n-gram selection can be demonstrated for bigrams, where the levels of 0.1 and 4 were used. The selection of

features according to their significance is able to support the SVMs capability to control model complexity independently of input dimension.

Table 1. Classification results on spoken and written material. Linear kernels and ten-fold cross-validation are applied

linguistic		n-gram		**politics**			**science**		
units	source	degree	thresh.	prec.	recall	F_{val}	prec.	recall	F_{val}
words	written	1	0.1	65.5	69.1	67.3	69.1	53.8	**60.5**
		1	4.0	66.1	69.1	**67.6**	71.6	55.8	62.7
		2	0.1	69.9	62.3	65.9	76.8	41.3	53.8
		2	4.0	69.5	63.2	66.2	85.2	44.2	58.2
		3	0.1	71.1	63.6	67.1	80.0	38.5	51.9
		3	4.0	71.5	60.5	65.5	84.9	43.3	57.3
syllables	written	1	0.1	63.0	80.5	70.7	70.5	76.0	**73.1**
		1	4.0	58.7	78.2	67.1	68.1	77.9	72.6
		2	0.1	69.4	72.3	70.8	78.1	54.8	64.4
		2	4.0	66.5	72.3	69.3	72.8	56.7	63.8
		3	0.1	71.2	70.9	71.1	78.7	46.2	58.2
		3	4.0	68.7	67.7	68.2	75.7	51.0	60.9
		4	0.1	71.9	70.9	**71.4**	80.0	46.2	58.5
		4	4.0	70.2	66.4	68.2	79.0	47.1	59.0
		5	4.0	71.1	65.0	67.9	79.3	44.2	56.8
		6	4.0	70.6	64.5	67.5	79.3	44.2	56.8
phonemes	written	2	0.1	55.2	84.5	66.8	59.5	90.4	71.8
		2	4.0	57.3	85.9	68.7	59.0	88.5	70.8
		3	0.1	60.6	79.5	68.8	72.8	72.1	72.5
		3	4.0	60.0	79.1	68.2	74.1	79.8	**76.9**
		4	0.1	65.9	76.4	70.7	81.2	66.3	73.0
		4	4.0	63.9	78.2	70.3	76.3	68.3	72.1
		5	4.0	65.0	75.0	69.6	77.9	57.7	66.3
		6	4.0	68.6	73.6	**71.1**	80.6	51.9	63.2
syllables	spoken	1	0.1	58.2	75.9	**65.9**	39.6	36.5	38.0
		1	4.0	57.6	75.5	65.4	40.2	45.2	**42.5**
		2	0.1	71.8	48.6	58.0	80.0	3.9	7.3
		2	4.0	69.0	52.7	59.8	60.0	5.8	10.5
		3	4.0	75.2	34.5	47.4	33.3	1.0	1.9
		4	4.0	76.5	29.5	42.6	33.3	1.0	1.9
		5	4.0	77.5	28.2	41.3	33.3	1.0	1.9
		6	4.0	77.5	28.2	41.3	33.3	1.0	1.9
phonemes	spoken	2	0.1	43.5	84.1	57.4	28.7	65.4	39.9
		2	4.0	47.9	79.5	59.8	30.2	59.6	**40.1**
		3	4.0	58.4	71.4	**64.2**	42.6	27.9	33.7
		4	4.0	64.8	61.8	63.3	63.2	11.5	19.5
		5	4.0	67.5	49.1	56.8	80.0	3.9	7.3
		6	4.0	73.4	41.4	52.9	50.0	1.0	1.9

Table 2. SVM classification of spoken documents when trained on written material. Only the topic category 'politics' is considered. Linear kernels are applied and ten-fold cross-validation is performed

linguistic units	n-gram degree	thresh.	results for **politics** prec.	recall	F_{val}
syllables	1	0.1	57.5	57.7	57.6
	1	4.0	55.6	54.1	54.8
	2	0.1	72.5	33.6	46.0
	2	4.0	64.5	53.6	**58.6**
	3	0.1	79.1	30.9	44.4
	3	4.0	71.2	42.7	53.4
	4	0.1	79.3	31.4	45.0
	4	4.0	74.3	38.2	50.5
	5	4.0	74.5	34.5	47.2
	6	4.0	74.5	33.2	45.9
phonemes	2	0.1	48.8	82.3	61.3
	2	4.0	55.5	69.1	61.5
	3	0.1	57.6	77.7	**66.2**
	3	4.0	59.5	74.1	66.0
	4	0.1	65.4	60.9	63.1
	4	4.0	59.8	69.5	64.3
	5	4.0	60.8	70.5	65.3
	6	4.0	62.2	67.3	64.6

7.2 Experiments with Spoken Documents

As discussed above the language model of the speech recognizer was trained on a text corpus which is different from the spoken documents to be recognized. Only 35% of the syllables produced by CSR were correct. With the experiments we can investigate if there are enough regularities left in the output of the CSR such that a classification by the SVM is possible. This also depends on the extent of systematic errors introduced by the CSR.

Again we performed experiments with respect to the two topic categories 'politics' and 'science'. In the next section we evaluate the results for spoken documents and compare them to the results for written material. As before the SVM was trained on the output of the CSR and used to classify the documents in the test set. The results are shown in the lower part of table 1.

For 'politics' simple syllables have an F-value of 65.9%. This is only 5% worse than for the written material. The effect of errors introduced by the CSR is relatively low. There is a sharp performance drop for higher order n-grams with $n > 3$. A possible explanation is the fact that the language model of the CSR is based on bigrams of syllables.

For 'science' classifiers using syllables for spoken documents yield only an F-value of 42.5% and perform far worse than for written documents (73.1%).

Table 3. Optimal F-values for experiments discussed in this paper

topic category	data used for training	test	optimal F-values words	syllables	phonemes
'politics'	written	written	67.6	71.4	71.1
'politics'	spoken	spoken	—	65.9	64.2
'politics'	written	spoken	—	58.6	66.2
'science'	written	written	60.5	73.1	76.9
'science'	spoken	spoken	—	42.5	40.1

Probably the errors introduced by CSR together with the small size of the class lead to this result.

Surprisingly phonemes yield for the topic category 'politics' on spoken documents an F-value of 64.2% which is nearly as good as the results for syllables. This result is achieved for 3-grams. For the small category 'science' phonemes yield 40.1% which is about 1.5% worse than the result for syllables.

7.3 Classification of Spoken Documents with Models Trained on Written Material

To get insight into the regularities of the errors of the speech recognizer we trained the SVM on synthetic syllables and phonemes generated for the written documents by BOSSII and applied these models to the output of the CSR.

The results for this experiment are shown in table 2. Whereas models trained on phonemes arrive at an F-value of 45.0% the syllables get up to 63.4%. This is nearly as much as the maximum F-value of 65.9% resulting from a model directly trained on the CSR output. This means that — at least in this setting — topic classification models may be trained without loss on synthetically generated syllables instead of genuine syllables obtained from a CSR.

We suppose that in spite of the low recognition rate of the speech recognizer the spoken and written dataset correspond to each other in terms of those syllables which consist the most important features for the classification procedure. On may argue, that those syllables are pronounced more distinctively which makes them better recognizable.

8 Discussion and Conclusions

The main results of this paper are summarized in table 3.

- On written text the utilization of n-grams of sub-word units like syllables and phonemes improve the classification performance compared to the use of words. The improvement is dramatic for small document classes.
- If the output of a continuous speech recognition system (CSR) is used for training and testing there is a drop of performance, which is relatively small

for larger classes and substantial for small classes. On the basis of syllable n-grams the SVM can compensate errors of a low-performance speech recognizer.
- In our setup it is possible to train syllable classifiers on written material and apply them to spoken documents. This is important since written material is far easier to obtain in larger quantities than annotated spoken documents.

An interesting result is that a spoken document classifier can be trained on written texts. This means that no *spoken* documents are needed for training a spoken document classifier. One can instead rely on *written* documents which are much easier to obtain.

The advantage of using for example syllables instead of words as input for the classification algorithm is that the syllables occurring in a given language can be well-represented by a finite inventory, typically containing several thousand forms. The inventory of words, in contrast, is infinite due to the productivity of word formation processes (derivation, borrowing, and coinage).

The results were obtained using a CSR with a simple speaker-independent acoustic model and a domain-independent statistical language model of syllable bigrams, insuring that recognizer performance is not specific to the experimental domain. Both models were trained on a different corpus which shows that the CSR may be applied to new corpora without the need to retrain. Syllables help circumvent the need for a domain-specific vocabularies and allows transfer to new domains to occur virtually unhindered by OOV considerations. The syllable model serves to control the complexity of the system, by keeping the inventory of text features to a bare minimum.

It is difficult to judge the significance of results. Since the tables demonstrate the F-value shows a stable behavior for different experimental setups, we think that the tendencies we have discovered are substantial. Future experiments will seek to further substantiate our results by evaluating topic categories additional to the two focused as well as investigating different kernels.

Acknowledgment

This study is part of the project Pi-AVIda which is funded by the German ministry for research and technology (BMFT) (proj. nr. 107). We thank the Institute for Communication and Phonetics of the University of Bonn for contributing the BOSSII system and we thank Thorsten Joachims (Cornell University) who provided the SVM-implementation SVM^{light}.

References

1. Drucker, H, Wu, D., Vapnik, V. Support vector machines for spam categorization. IEEE Transactions on Neural Networks, 10 (5): 10048-1054, 1999. 374
2. Dumais, S., Platt, J., Heckerman, D., Sahami, M. (1998): Inductive learning algorithms and representations for text categorization. In: *7th International Conference on Information and Knowledge Management,* 1998. 374

3. Gelman, A., Carlin J. B., Stern, H. S., Rubin, D. B.: Bayesian Data Analysis. Chapman, Hall, London, 1995. 378
4. Glavitsch, U., Schäuble, P. (1992): A System for Retrieving Speech Documents, SIGIR 1992. 376
5. Haussler, David (1999): Convolution Kernels on Discrete Structures, UCSL-CRL-99-10. 374
6. Joachims, T. (1998). Text categorization with support vector machines: learning with many relevant features. *Proc. ECML '98*, (pp. 137–142). 373, 374, 379
7. Klabbers, E., Stöber, K. , Veldhuis, R. Wagner, P., Breuer, S.: Speech synthesis development made easy: The Bonn Open Synthesis System, EUROSPEECH 2001. 376, 377
8. Larson, M.: Sub-word-based language models for speech recognition: implications for spoken document retrieval, Proc. Workshop on Language Modeling and IR. Pittsburgh 2001. 376
9. Leopold, E., Kindermann, J.: Text Categorization with Support Vector Machines. How to Represent Texts in Input Space? Machine Learning, 46, 2002, 423–444. 374
10. Leslie, Christa, Eskin, Eleazar, Noble, William Stafford (2002): The Spectrum Kernel: A String Kernel SVM Protein Classification. To appear: Pacific Symposium on Biocomputing. 374
11. Lodhi, Huma, Shawe-Taylor, John, Cristianini, Nello & Watkins, Chris (2001) Text classification using kernels, NIPS 2001, pp. 563-569. MIT Press. 374
12. Manning, Christopher D., Schütze (2000): Foundations of Statistical Natural Language Processing, MIT Press. 378
13. Watkins, Chris (1998): Dynamic alignment Kernels. Technical report, Royal Holloway, University of London. CSD-TR-98-11. 374