

Machine Learning-Based Keywords Extraction for Scientific Literature

Chunguo Wu

(College of Computer Science and Technology, Jilin University, Key Laboratory
of Symbol Computation and Knowledge Engineering of Ministry of Education
Changchun 130012, China

and

The Key Laboratory of Information Science & Engineering of Railway
Ministry, The Key Laboratory of Advanced Information Science and Network
Technology of Beijing, Beijing Jiaotong University
Beijing 100044, China
wucg@jlu.edu.cn)

Maurizio Marchese

(Department of Information and Communication Technology University of
Trento, Via Sommarive 14, 38050 - Povo (TN), Italy
maurizio.marchese@unitn.it)

Jingqing Jiang

(College of Computer Science and Technology, Jilin University, Key Laboratory
of Symbol Computation and Knowledge Engineering of Ministry of Education
Changchun 130012, China

and

College of Mathematics and Computer Science, Inner Mongolia University for
Nationalities, Tongliao 028043, China
tljjq@263.net)

Alexander Ivanyukovich

(Department of Information and Communication Technology University of
Trento, Via Sommarive 14, 38050 - Povo (TN), Italy
a.ivanyukovich@dit.unitn.it)

Yanchun Liang

(College of Computer Science and Technology, Jilin University, Key Laboratory
of Symbol Computation and Knowledge Engineering of Ministry of Education
Changchun 130012, China
ycliang@jlu.edu.cn)

Abstract: With the currently growing interest in the Semantic Web, keywords/metadata extraction is coming to play an increasingly important role. Keywords extraction from documents is a complex task in natural languages processing. Ideally this task concerns sophisticated semantic analysis. However, the complexity of the problem makes

current semantic analysis techniques insufficient. Machine learning methods can support the initial phases of keywords extraction and can thus improve the input to further semantic analysis phases. In this paper we propose a machine learning-based keywords extraction for given documents domain, namely scientific literature. More specifically, the least square support vector machine is used as a machine learning method. The proposed method takes the advantages of machine learning techniques and moves the complexity of the task to the process of learning from appropriate samples obtained within a domain. Preliminary experiments show that the proposed method is capable to extract keywords from the domain of scientific literature with promising results.

Key Words: keywords extraction, metadata extraction, support vector machine, machine learning

Category: H.3.7, H.5.4

1 Introduction

Scientists have communicated and codified their finding in a relatively orderly, well defined way since 17th century through the use of books, serial literature (journals), intellectual property right documents (patents). But many new channels and usages of communication are rapidly developing: electronic publishing, digital libraries, electronic proceedings, and more recently blogs and scientific news streaming are rapidly expanding the amount of available scientific/scholarly digital content related to research and innovation. Recently, we have also witnessed a major shift in the landscape of publishing: the number of open access journals is rising steadily, and new publishing models are rapidly evolving to test new ways to increase readership and access.

In a study carried out in 2003 at the University of California at Berkeley [Lyman et al. 2003], it has been estimated that the world produces between 1 and 2 exabytes (109 GB) of unique information per year, which is roughly 250 megabytes for every man, woman, and child on earth. Printed documents of all kinds comprise only .003% of the total. Digital format is rapidly becoming the universal medium for information storage and sharing.

Scientists benefit much from such quantity of available scholarly resources. However, like all other people, they are flooded with content and find it difficult to search and organize it with traditional methods. The need to provide effective IT platforms for managing and searching such a variety and quantity academic content both on the Web and on local/private repositories (digital libraries) is thus a crucial issue for the advance of scientific knowledge.

A solution proposed within the Semantic Web initiative consists of enriching each digital resource with associated semantics. This means that each digital resource needs to be annotated with terms (i.e. keywords) describing concepts mainly derived from a rich semantic model (i.e. an ontology) of the domain the resource is about. It is clear that, in order to scale to the size of the content under consideration, this approach needs to be supported by appropriate tools

that assist either automatically or semi-automatically the semantic annotation process.

Researchers have been aware of the importance of automatic extraction of semantic information from digital resources and different methodologies have been proposed to fulfill this task. The existing approaches include numerous metadata extraction, document summarization and keywords extraction techniques. Han et al. (2003) proposed an approach to automatically extract metadata of scientific literatures [Han et al. 2003] and the approach has been applied in the CiteSeer.IST project¹. Kiyavitskaya et al. (2005) proposed semi-automatic semantic annotation approach [Kiyavitskaya et al. 2005] based on techniques and technologies traditionally used in software analysis and reverse engineering. Daume et al. (2005) introduced word and phrase alignment-based approaches for document summarization [Daume and Marcu 2005]. Some studies have been performed to extract keywords, but not specific for scientific literatures. José Luis Martínez-Fernández (2003) et al. focused on the automatic keywords extraction for news characterization by using several linguistic techniques to improve the text-based information retrieval [Mart et al. 2004].

These efforts, and related work, can sustain and improve a number of modern scientific/scholarly content services. Both commercial ones like Chemical Abstracts Service®² for chemistry-related articles, Web of Knowledge®³ from ISI-Thomson and Scopus®⁴ from Elsevier B.V.; as well as very popular vertical communities services such as: CiteSeer.IST, DBLP⁵, and more recently Google Scholar⁶.

In this paper we propose a domain-oriented machine learning-based keywords extraction for scientific literature. In Section 2 we describe our motivating use-case where keywords extraction methods and tools are relevant. In Section 3 we present the proposed method based on one of the machine learning methods, namely the least square support vector machines (LS-SVM). In Section 4 we probe our proposed method on a sample of scientific literature documents. Conclusion and future work are given in Section 5.

2 Motivating case study: keywords extractions in a semantic content management system

In our current work, the need for automatic tools for keywords extractions comes within the development, carried out at the University of Trento, of a semantic

¹ <http://citeseer.ist.psu.edu/>

² <http://www.cas.org>

³ <http://www.isinet.com>

⁴ <http://www.scopus.com>

⁵ <http://dblp.uni-trier.de/>

⁶ <http://scholar.google.com/>

content management system for scientific literature. In this system, initially scientific documents are located on the Internet and downloaded to local storage. Then they are converted to textual format. Due to the specifics of text representation in PostScript and PDF formats output textual information may contain different artifacts that do not belong to meaningful content. These artifacts can make further information processing less efficient and can have subsequent negative impact on final results quality.

Several methods have been applied to find and eliminate these artifacts thus assuring the necessary quality level:

- Partial recognition of text structure;
- Pages order detection;
- Pages header/footer detection and elimination;
- Document content and index sections detection and elimination;
- Corrections of the partially recognized text structure (beginnings of abstract, keywords, introduction, conclusion, acknowledgement and reference sections).

Each of the outlined methods is based on statistical data analysis techniques, so they do not require any extra information and ensure high processing speed. Further information processing includes metadata extraction and subsequent metadata correction steps.

Correspondingly we have divided all information processing tasks to several major modules: Parsers, Pre-processors, Metadata Extractors and Post-Processors. The part of the semantic content management system architecture connected to the information processing tasks is represented in Fig. 1. The overall architecture can be described as a “conveyor chain”, where each module is a cluster (“cell”) that spreads corresponding tasks to available distributed processing facilities. The heart of the system is the “distributed file system”, which performs functions of data storage network. Information flow is organized in the way that modules never communicate directly. Instead they operate through distributed file system only. This kind of architecture fulfills three major goals: easy functional extensibility, high performance and scalability.

Because of the modules independency it is possible to easily integrate different keywords extraction techniques, like the one presented in this paper, into the existing information flow chain.

3 Machine learning-based keywords extraction

The proposed method consists of three parts: construction of a keyword database, selection of learning samples and training of a learning machine. Specifically, the

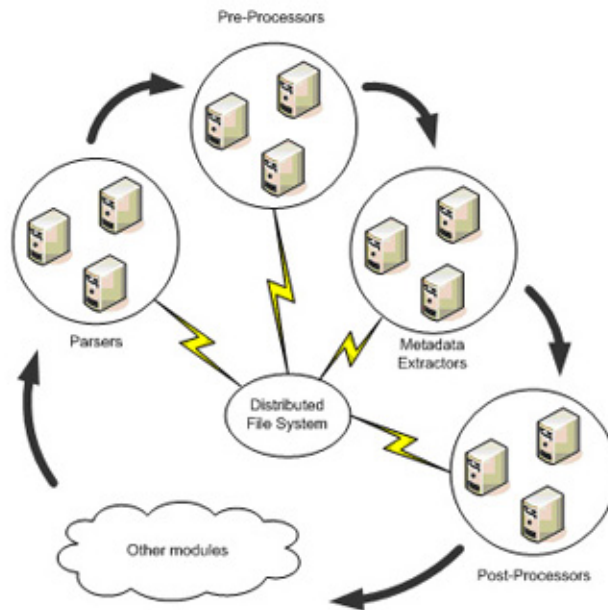


Figure 1: Semantic Content Management System architecture

LS-SVM is used as a model for machine learning. The keyword database is constructed from existing documents in a specific scientific domain with given keywords. Learning samples are drawn from documents with given keywords based on the obtained keyword database. Then the LS-SVM is trained using the samples drawn in the second part. After this process is completed we can use the trained learning machine to extract keywords for unseen documents in the same domain.

3.1 Constructions of keyword database and drawing of learning samples

Keywords database construction is grounded on the data prepared by the distributed semantic content management system designed at the University of Trento. After the Pre-Processors module (see Fig. 1), the scientific documents have already enough information for their classification into two major categories: with and without keywords indicated by document authors. Firstly we process all documents with indicated keywords: we thus collect all the given keywords and populated the keywords database with their unique set. For example, if a line in a pre-processed plain-text file is “*keywords: heuristic search; dynamic programming; markov decision problems*” then the keywords “heuristic

search”, ”dynamic programming” and ”markov decision problems” are put into the keyword database. These items collected in the keyword database are called *candidate keywords*.

Moreover, we observe that the relevance of a keyword can be roughly estimated by its frequency in four parts of the scientific document: title, abstract, body and conclusion (discussion/summary). For a given document, the title and abstract are relatively easy to be identified with heuristic rules implemented in the Metadata Extractors module of our system, since usually the title occupies the first lines of the document and the abstract follows the word ”Abstract”. It is a bit more difficult to determine the conclusion part, because there are some counterparts in scientific literature document, e.g., discussion and summary. Usually we consider the section before the bibliography/reference or acknowledgement (if available) as conclusion part, no matter what the section title is. All sections between abstract or keyword (if available) and conclusion are considered as the body.

Inspired by this observation, we design our samples as 5-dimensional vectors: $(nTitle, nAbstract, nBody, nConclusion, isKeyword)^T$, where $nTitle$, $nAbstract$, $nBody$, and $nConclusion$ are the times that a candidate keyword k appears in title, abstract, body and conclusion of a scientific literature document p , respectively, and $isKeyword$ is a binary variable. If the set of given keywords of document p contains the candidate keyword k , the corresponding $isKeyword$ is set to +1; otherwise, the $isKeyword$ is set to -1. In order to construct the training and testing samples, we scan each line in the plain-text file for each item in the keyword database and count the times that the term appears in each part to compute respectively $nTitle$, $nAbstract$, $nBody$, and $nConclusion$. Hence, if the number of items in keyword database is n and the number of documents in the first category (with keywords) is m , then n -by- m samples can be drawn.

3.2 Training of learning machines

Machine learning methods have demonstrated their relevance, especially, in the fields where the a-priori models are difficult to construct due to uncertainty or complexity. With the emergence of the second generation of statistical learning theory (Vapnik, 1998) [Vapnik 1998], many new powerful models based on support vector machine have been proposed in the machine learning domain: Joachims (1999) *et al.* proposed the SVM^{Light} , which is one of the most popular SVM [Vapnik 1999]. Platt (1999) proposed sequential minimal optimization (SMO) to train SVM, which enabled to analytically compute the coefficient from series of the smallest quadratic programming problems [Platt 1999]. Suykens (1999 and 2000) *et al.* proposed Least squares support vector machine (LS-SVM), which was spread in engineering field in a short time due to its simplicity and efficiency [Suykens and Vandewalle 1999] [Suykens et al. 2000]. Wu (2006) *et*

al. proposed an adaptive iterative training algorithm of LS-SVM, which makes LS-SVM can be trained iteratively and remain the sparseness of support vectors [Jiang et al. 2006]. Jiang (2005) *et al.* proposed a classification method based on function regression [Jiang et al. 2005], which can be used to implement multi-classification efficiently and is entirely different with traditional methods for multi-classification (1-vs-1 or 1-vs-all) [Angulo et al. 2006] [Anguita et al. 2004] [Kressel 1999]. In this paper this regression-based classification method is used to verify the keywords extraction approach.

The regression-based classification method proposed by Jiang (2005) *et al.* is introduced briefly in the following, from [Jiang et al. 2005]:

Let us consider a given training set of N samples $\{x_i, y_i\}$ with the i th input vector $x_i \in R^n$ and the i th output target $y_i \in R$. The aim of support vector machines model is to construct the decision function takes the form:

$$f(x, w) = w^T \varphi(x) + b \quad (1)$$

In least squares support machines for function regression the following optimization problem is formulated

$$\begin{cases} \min_{w, e} J(w, e) = \frac{1}{2} \|w\|^2 + \gamma \sum_{i=1}^N e_i^2 \\ \text{s.t. } y_i = w^T \varphi(x_i) + b + e_i, (i = 1, \dots, N) \end{cases} \quad (2)$$

where γ is a predetermined parameter to balance the precisions between learning and generalization.

$$L(w, b, e, \alpha) = J(w, e) - \sum_{i=1}^N \alpha_i \{w^T \varphi(x_i) + b + e_i - y_i\} \quad (3)$$

with Lagrange multipliers α_i . The solution is given by the following set of linear equations

$$\begin{pmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & \Omega + \gamma^{-1}I \end{pmatrix} \begin{pmatrix} b \\ \alpha \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix} \quad (4)$$

where

$$\Omega_{kj} = \varphi(x_k)^T \varphi(x_j) = \psi(x_k, x_j) \quad (k, j = 1, \dots, N) \quad (5)$$

Let $A = \Omega + \gamma^{-1}I$. Because A is a symmetric and positive-definite matrix, A^{-1} exists. Solving the set of linear Eqs.(6), one can obtain the solution

$$\alpha = A^{-1}(y - b\mathbf{1}) \quad b = \frac{\mathbf{1}^T A^{-1} y}{\mathbf{1}^T A^{-1} \mathbf{1}} \quad (6)$$

Substituting w in Eq. (1) with its expression of α [?], we have

$$f(x, w) = y(x) = \sum_{i=1}^N \alpha_i K(x, x_i) + b \quad (7)$$

The kernel function $K(\cdot)$ is chosen as a radial basis function

$$K(x, x_i) = \exp\{-\|x - x_i\|^2 / (2\sigma^2)\} \quad (8)$$

where σ is a predetermined constant, called as kernel width.

The steps of the regression-based classification method for multi-category problems are as follows [Jiang et al. 2005]:

Step 1. Set class-label for each class. The class-label is usually set as decimal integer, such as $i = 1, 2, \dots, n$.

Step 2. Solve the set of linear Eqs. (4) to get the solutions of α_i and b .

Step 3. Put the solutions of α_i and b into Eq. (7), and obtain the regression function $f(x)$.

When the value of the regression function $f(x)$ is in the specified region of class-label for a given sample x , the sample x is classified by the regression function $f(x)$ correctly.

4 Preliminary experiments

To probe the validity of the proposed method, we selected randomly 332 scientific literature documents with given keywords from our document bibliography database (both DBLP and University of Trento repositories). From these documents, totally 1313 candidate keywords have been collected and put into the keyword database. By using these scientific literature documents with given keywords and candidate keywords, we draw our samples according to the method proposed in 3.1. In these samples there are ca. 11% of positive samples and ca. 89% of negative samples.

With these original samples, 10 experiments of training and testing are performed. The running parameters (γ and σ) are selected as 5000 and 0.01 with 10-fold cross-validation in the space of $[1, 60000]$ -by- $[0.01, 100]$ and the step sizes for γ and σ are 10 and 0.01, respectively. The results are listed in Table 1, where $CR(+)$ and $CR(-)$ represent the correct rates of samples in positive and negative classes, respectively, and CR represents the correct rate of the whole samples. Denote a sample as s_i and the training or testing sample set as S , the formulas for computation of $CR(+)$, $CR(-)$ and CR are as follows:

$$CR(+) = \frac{|\{s_i \mid s_i \in S^+, f(s_i) > 0\}|}{|S^+|} \quad (9)$$

$$S^+ = \{s_i \mid s_i \in S, (s_i)_5 = +1\}$$

Table 1: Training and testing results of original samples

	No.	CR(+)	CR(-)	CT
training (%)	1	54.2936	99.7347	94.2667
	2	48.3871	99.8496	94.0000
	3	55.5891	99.8876	95.0000
	4	58.8235	99.7759	95.3667
	5	48.6957	99.8117	93.9333
	6	48.6726	99.8873	94.1000
	7	55.3517	99.8504	95.0000
	8	53.0612	99.8118	94.4667
	9	53.3537	99.7380	94.6667
	10	52.8409	99.6601	94.1667
	average	52.9069	99.8007	94.4967
testing (%)	1	05.9006	98.8798	88.9000
	2	03.5294	99.3233	88.4667
	3	02.8902	98.7566	87.7000
	4	04.2493	98.3755	87.3000
	5	04.8159	98.8289	87.7667
	6	08.9231	98.6916	88.9667
	7	03.6517	99.1679	87.8333
	8	04.4510	99.3992	88.7333
	9	04.3732	98.8333	88.0333
	10	03.2353	98.9850	88.1333
	average	04.6020	98.9241	88.1833

$$CR(-) = \frac{|\{s_i \mid s_i \in S^-, f(s_i) \leq 0\}|}{|S^-|} \quad (10)$$

$$S^- = \{s_i \mid s_i \in S, (s_i)_5 = -1\}$$

$$CR = \frac{|\{s_i \mid s_i \in S^+, f(s_i) > 0\} \cup \{s_i \mid s_i \in S^-, f(s_i) \leq 0\}|}{|S^+| + |S^-|} \quad (11)$$

where $(s_i)_5$ is the class label for the i th sample in set S .

Generally, we could obtain an LS-SVM with a higher precision. However, as shown in Table 1, it can be seen that the unbalanced data (much more negative samples than positive ones) deteriorates seriously the precision of positive samples in the testing phase. To reduce this disadvantage, the positive samples are duplicated 8 times to balance the ratio of positive and negative samples according to Murphey (2004) [Murphey and Guo 2004]. With the balanced samples,

Table 2: Training and testing results of balanced samples

	No.	CR(+)	CR(-)	CT
training (%)	1	74.7694	87.7193	81.1667
	2	75.9235	88.8140	82.3000
	3	74.6179	87.5585	81.0667
	4	76.5615	88.9477	82.8000
	5	76.5246	86.9831	81.6667
	6	76.5563	86.9128	81.7000
	7	76.4045	88.8366	82.5667
	8	73.9446	90.1617	81.9667
	9	74.8858	89.5024	82.0333
	10	76.0238	90.7133	83.3000
	average	75.6212	88.6149	82.0567
testing (%)	1	73.8710	77.7241	75.7333
	2	74.4997	81.3232	77.8000
	3	72.5581	80.1338	76.3333
	4	72.5426	78.4939	75.4667
	5	74.6571	76.4466	75.5333
	6	72.4483	79.1472	75.8000
	7	76.7320	78.7755	77.7333
	8	69.4301	80.9753	75.0333
	9	72.0430	78.3650	75.0333
	10	72.1799	80.4502	76.4667
	average	73.0962	79.1835	76.0933

the above experiments are repeated and the results are listed in Table 2. The meanings of symbols used in this table are the same as those in Table 1. As shown in Table 2, by introducing the data balance method, the correct rates of positive samples are improved about or more than 20 times, although the whole correct rates are pulled somewhat down (on the average 12%). Maybe this is what we have to accept for the lack of more efficient methods of data balance.

To demonstrate the generalization performance of the proposed method, we randomly selected 116 literatures without given keywords from the same document bibliography repository. Because of the lack of given keywords, the samples construction from these literatures are 4-dimensional vectors, i.e.,

$$(nTitle, nAbstract, nBody, nConclusion)^T$$

and the binary component, isKeyword, is omitted. We present the extracted keywords of 10 documents with corresponding titles in Table 3.

Table 3: Generalization Performance for literatures without given keywords

No.	Title and extracted keywords	
1	Title	Querying Semistructured Heterogeneous Information
	Keywords	Semantics; query; language; meaning
2	Title	Efficient and Flexible Location Management Techniques for Wireless Communication Systems
	Keywords	Graphical; communication; information; search
3	Title	Querying the World Wide Web
	Keywords	world wide web; query; language; distributed
4	Title	On Using a Manhattan Distance-like Function for Robot Motion Planning on a Non-Uniform Grid in Configuration Space
	Keywords	Configuration; extensions; representation; constraints
5	Title	Genetic Algorithms Tournament Selection and the Effects of Noise
	Keywords	genetic algorithms; sampling; noise; evaluation
6	Title	Bayesian Interpolation
	Keywords	complexity; inference; approximation; embodied
7	Title	Acting Optimally in Partially Observable Stochastic Domains
	Keywords	stochastic; belief; search; markov decision planning
8	Title	Deriving Production Rules for Incremental View Maintenance
	Keywords	stochastic; maintenance; information; logic
9	Title	Topography And Ocular Dominance: A Model Exploring Positive Correlations
	Keywords	Logic; pattern; learning; distributed

5 Conclusions and future work

In this paper we propose an offline method for keywords extraction from scientific literature documents. After collecting a proper keyword database, the proposed method can be used to extract keywords from scientific literature documents within a given domain. This method can also be easily extended to online adaptive methods by using adaptive online learning approaches of SVM. When the proposed method is extended to online adaptive version, we expect improvements due to the distributed actions of users interacting with the learning system.

Although the simulated experiments show that the proposed method is prom-

ising, the data unbalance is an inevitable problem in the training process of learning machines by using this approach. To reduce the effect of data unbalance, we expect to obtain better results, than those obtained here, using the current data balance procedure for increasing the number of samples of under-sampled category, which is also one of our incoming work directions.

What should be pointed out is that the quality of initial metadata identifications, i.e. identification of title, abstract, conclusion, acknowledgement, appendix and reference sections, is crucial for improving the efficiency and accuracy of the proposed method for keywords extraction. Current work in regard to the development of a semantic content management system is aiming to provide such quality initial metadata automatic extraction. Moreover, we are working on exploring other keywords extraction strategies and methods and compare their results with the proposed approach.

Acknowledgment

The authors would like to thank for the support of the European Commission for the Erasmus Mundus programme and the project of TH/Asia Link/010 (111084), the National Natural Science Foundation of China (60673023, 60433020), the science-technology development project of Jilin Province of China (20050705-2), the doctoral funds of the National Education Ministry of China (20030183060), the graduate innovation lab of Jilin University (503043), and “985” project of Jilin University.

References

- [Lyman et al. 2003] Lyman, Peter, Varian, H. R.: “How Much Information”; (2003). <http://www.sims.berkeley.edu/how-much-info-2003on15march2006>.
- [Han et al. 2003] Han, H. , Giles, C. L., Manavoglu, E. , Zha, H. Y., Zhang, Z. Y.: “Automatic document metadata extraction using support vector machines”; Proceedings of the 2003 Joint Conference on Digital Libraries, Houston, Texas, USA, (2003), 37-48.
- [Kiyavitskaya et al. 2005] Kiyavitskaya, N., Zeni, N., Cordy, J. R., Mich, L., Mylopoulos, J.: “Semi-Automatic semantic annotations for web documents”; Proc. SWAP 2005, 2nd Italian Semantic Web Workshop, Trento, Italy, 2005.
- [Daume and Marcu 2005] Daume, H., Marcu, D.: “Induction of word and phrase alignments for automatic document summarization”; Computational Linguistics, 31(4), (2005), 505-530.
- [Mart et al. 2004] Martínez-Fernández, J. L., García-Serrano, A., Martínez, P. Villena, J.: “Automatic keyword extraction for news finder”; Lecture Notes in Computer Science, 3094, (2004), 99-119.
- [Vapnik 1998] Vapnik, V. N.: “Statistical Learning Theory”; Springer-Verlag, New York, 1998.
- [Vapnik 1999] Vapnik, T.: “Making large-scale SVM learning practical”; Advances in Kernel Methods-Support Vector Learning, (B. Scholkopf, C. Burges, A. J. Smola, eds.), MIT Press, Cambridge, 1999, 169-184.

- [Platt 1999] Platt, J. C.: "Fast training of support vector machines using sequential minimal optimization"; *Advances in Kernel Methods-Support Vector Learning*, (B. Scholkopf, C. Burges, A. J. Smola, eds.), MIT Press, Cambridge, 1999, 185-208.
- [Suykens and Vandewalle 1999] J. A. K. Suykens, J. Vandewalle: "Least squares support vector machine classifiers"; *Neural Processing Letters*, 9, 3, (1999), 293-300.
- [Suykens et al. 2000] Suykens, J. A. K., Lukas, L., Vandewalle, J.: "Sparse approximation using least squares support vector machines"; *Proceedings of the IEEE International Symposium on Circuits and Systems*, Geneva, Switzerland, 2000, 757-760.
- [Jiang et al. 2006] Jiang, J. Q., Song, C. Y., Wu, C. G., Marchese, M., Liang, Y. C.: "Support vector machine regression algorithm based on chunking incremental learning"; *Lecture Notes in Computer Science*, 3991, (2006), 547-554.
- [Jiang et al. 2005] Jiang, J. Q., Wu, C. G., Liang, Y. C.: "Multi-category classification by least squares support vector regression"; *Lecture Notes in Computer Science*, 3496, (2005), 863-868.
- [Angulo et al. 2006] Angulo, C., Ruiz, F. J., Gonzalez, L., Ortega, J. A.: "Multi-classification by using tri-class SVM"; *Neural Processing Letters*, 23, 1, (2006), 89-101.
- [Anguita et al. 2004] Anguita, D., Ridella, S., Sterpi, D.: "A New Method for Multi-Class Support Vector Machines"; *Proceedings of the IEEE IJCNN 2004*, Budapest, Hungary, 2004.
- [Kressel 1999] Kressel, U.: "Pairwise classification and support vector machine"; *Advances in Kernel Methods-Support Vector Learning*, (B. Scholkopf, C. Burges, A. J. Smola, eds.), Cambridge, MA, MIT Press, (1999), 255-268.
- [Murphey and Guo 2004] Murphey, Y. L., Guo, H.: "Neural learning from unbalanced data"; *Applied Intelligence*, 21, 2, (2004), 117-128.