Automated Multi-label Text Categorization with VG-RAM Weightless Neural Networks

Alberto F. De Souza [a], Felipe Pedroni [a], Elias Oliveira [b], Patrick M. Ciarelli [c], Wallace Favoreto Henrique [a], Lucas Veronese [a], Claudine Badue [a]

[a] *Departamento de Informática*

[b] *Departamento de Ciência da Informação*

[c] *Departamento de Engenharia Elétrica*
Universidade Federal do Espírito Santo,
Av. Fernando Ferrari, 514
29075-910, Vitória - ES, Brazil

Abstract

In automated multi-label text categorization, an automatic categorization system should output a label set, whose size is unknown a priori, for each document under analysis. Many machine learning techniques have been used for building such automatic text categorization systems. In this paper, we examine Virtual Generalizing Random Access Memory Weightless Neural Networks (VG-RAM WNN), an effective machine learning technique which offers simple implementation and fast training and test, as a tool for building automatic multi-label text categorization systems. We evaluated the performance of VG-RAM WNN on two real-world problems, (i) categorization of free-text descriptions of economic activities and (ii) categorization of Web pages, and compared our results with that of the multi-label lazy learning approach, ML-KNN. Our experimental comparative analysis showed that, on average, VG-RAM WNN either outperforms ML-KNN or show similar categorization performance.

Key words: VG-RAM weightless neural networks, multi-label text categorization, Web page categorization, categorization of economic activities

Email addresses: `alberto@lcad.inf.ufes.br` (Alberto F. De Souza), `elias@lcad.inf.ufes.br` (Elias Oliveira), `claudine@lcad.inf.ufes.br` (Claudine Badue).

1 Introduction

Automatic text categorization is still a very challenging computational problem to the information retrieval communities both in academic and industrial contexts. Most works on text categorization in the literature are focused on single-label text categorization problems, where each document may have only a single label [1]. However, in real-world problems, multi-label categorization is frequently necessary [2–13].

From a theoretical point of view, single-label categorization is more general than multi-label, since an algorithm for single-label categorization can also be used for multi-label categorization: one needs only to transform the multi-label categorization problem into n independent single-label problems, where n is number of possible labels, or categories [1]. However, this equivalence only holds if the n categories are stochastically independent, that is, the association of a category c_i to a document is independent of the association of another category, c_j, to the same document; however, this frequently is not the case. Multi-label categorization systems can take advantage of the correlation between categories in order to improve their performance.

Several techniques for multi-label categorization have been proposed, such as multi-label decision trees [4,6], multi-label kernel methods [5,8,11] or multi-label neural networks [10,12], and many of them specifically for multi-label text categorization [2,3,7,9,10,12]. In this paper we present an experimental evaluation of the performance of Virtual Generalizing Random Access Memory Weightless Neural Networks (VG-RAM WNN [14,15]) on multi-label text categorization. We have chosen an experimental evaluation approach because of the very nature of the task—automatic text categorization relies solely on text semantics, and given that the semantics of a document is a subjective notion, the inclusion of a document in a category cannot be decided deterministically [1].

VG-RAM WNN is an effective machine learning technique which offers simple implementation and fast training and test [14]. We evaluated the categorization performance of VG-RAM WNN on two different real-world multi-label problems: categorization of free-text descriptions of economic activities, and categorization of Web pages. The automation of the categorization of economic activities of companies from business descriptions in free text format is a huge challenge for the Brazilian governmental administration in the present day. So far, this task has been carried out by humans, not all of them properly trained for the job. When this problem is tackled by humans, the subjectivity on their categorization brings a problem: different human categorizers can give different results when working on the same business description. This can cause distortions in the information used for planning, taxation and other governmental obligations of the three Brazilian administrative levels: County, State and Federal. Furthermore, the number of possible categories considered is very large, more than 1000 in the Brazilian scenario, which makes the categorization problem even harder to be solved. Web page categorization, on the other hand, is used by several Web search companies, such as Google and Yahoo, for helping users navigate the Internet, and has significant economic value.

In this work, the performance of VG-RAM WNN on the two categorization problems mentioned was analyzed using four multi-label categorization metrics: *hamming loss*, *one-error*, *coverage*, and *average precision* [3]. We also compared the VG-RAM WNN performance with that of the multi-label lazy learning technique, ML-KNN, i.e., Multi-label k-Nearest Neighbors, proposed by Zhang and Zhou [13]. Their technique achieved higher performance than well-established algorithms in several multi-label problems [13]. Our results show that, in the categorization of free-text descriptions of economic activities, VG-RAM WNN outperforms ML-KNN in terms of the four multi-label evaluation metrics employed, while in the Web page categorization problem, on average, VG-RAM WNN outperforms ML-KNN in terms of *hamming loss*, *coverage* and *average precision*, and show similar categorization performance in terms of *one-error*.

This paper is organized as follows. Section 2 introduces the multi-label text categorization problem and the metrics used to evaluate the performance of the categorization techniques examined. Section 3 and Section 4 briefly introduce VG-RAM WNN and ML-KNN, respectively, and describe how we have used them for multi-label text categorization. Section 5 presents our experimental methodology and analyzes our experimental results. Our conclusions and directions for future work follow in Section 6.

A preliminary version of this paper appeared in the Proceedings of the 7^{th} IEEE International Conference on Intelligent Systems Design and Applications [16].

2 Multi-Label Text Categorization

Text categorization may be defined as the task of assigning categories (or labels), from a predefined set of categories, to documents [1]. In multi-label text categorization, one or more categories may be assigned to a document.

Let $\mathcal{D}$ be the domain of documents, $\mathcal{C} = \{c_1, \ldots, c_{|\mathcal{C}|}\}$ a set of pre-defined categories, and $\Omega = \{d_1, \ldots, d_{|\Omega|}\}$ an initial corpus of documents previously categorized manually by a domain expert into subsets of categories of $\mathcal{C}$. In multi-label learning, the training(-and-validation) set $TV = \{d_1, \ldots, d_{|TV|}\}$ is composed of a number documents, each associated with a subset of categories of $\mathcal{C}$. TV is used to train and validate (actually, to tune eventual parameters of) a categorization system that associates the appropriate combination of categories to the characteristics of each document in the TV. The test set $Te = \{d_{|TV|+1}, \ldots, d_{|\Omega|}\}$, on the other hand, consists of documents for which the categories are unknown to the categorization system. After being (tunned and) trained with TV, the categorization system is used to predict the set of categories of each document in Te.

A multi-label categorization system typically implements a real-valued function $f : \mathcal{D} \times \mathcal{C} \rightarrow \mathbb{R}$ that returns a value for each pair $\langle d_j, c_i \rangle \in \mathcal{D} \times \mathcal{C}$ that, roughly speaking, represents the evidence for the fact that the test document d_j should be categorized under the category c_i. The real-valued function $f(.,.)$ can be transformed into a ranking function $r(.,.)$, which is a one-to-one mapping onto $\{1, 2, \ldots, |\mathcal{C}|\}$ such that, if $f(d_j, c_1) > f(d_j, c_2)$, then $r(d_j, c_1) < r(d_j, c_2)$. If C_j is the set of proper categories for the test document d_j, then a successful categorization system will tend to rank categories in C_j higher than those not in C_j. Those categories that rank above a threshold τ (i.e., $c_k | f(d_j, c_k) \geq \tau$) are then assigned to the test document d_j.

We have used four multi-label evaluation metrics proposed in [17,3] for examining the categorization performance of VG-RAM WNN, namely *hamming loss*, *one-error*, *coverage*, and *average precision*. The metrics *one-error*, *coverage*, and *average precision* evaluate the whole ranking derived from the real-valued function $f(.,.)$, while *hamming loss* evaluates the exact set of categories predicted for the test document d_j. We present each of these metrics below.

Hamming Loss (hloss$_j$) evaluates how many times the test document d_j is misclassified, i.e., a category not belonging to the document is predicted or a category belonging to the document is not predicted:

$$\text{hloss}_j = \frac{1}{|\mathcal{C}|} |P_j \Delta C_j| \tag{1}$$

where $|\mathcal{C}|$ is the number of categories and Δ is the symmetric difference between the set of predicted categories P_j and the set of appropriate categories C_j of the test document d_j.

One-error (one-error$_j$) evaluates if the top ranked category is present in the set of proper categories C_j of the test document d_j:

$$\text{one-error}_j = \begin{cases} 0 \text{ if } [\arg\max_{c \in \mathcal{C}} f(d_j, c)] \in C_j \\ 1 \text{ otherwise.} \end{cases} \tag{2}$$

where $[\arg\max_{c \in \mathcal{C}} f(d_j, c)]$ returns the top ranked category for the test document d_j.

Coverage (coverage$_j$) measures how far we need to go down the rank of categories in order to cover all the possible categories assigned to a test document:

$$\text{coverage}_j = \max_{c \in C_j} r(d_j, c) - 1 \tag{3}$$

where $\max_{c \in C_j} r(d_j, c)$ returns the maximum rank for the set of appropriate categories of the test document d_j.

Average Precision (average-precision$_j$) evaluates the average of precisions computed after truncating the ranking of categories after each category $c_i \in C_j$ in turn:

$$\text{avgprec}_j = \frac{1}{|C_j|} \sum_{k=1}^{|C_j|} precision_j(R_{jk}) \tag{4}$$

where R_{jk} is the set of ranked categories that goes from the top ranked category until a ranking position k where there is a category $c_i \in C_j$ for d_j, and $precision_j(R_{jk})$ is the number of pertinent categories in R_{jk} divided by $|R_{jk}|$. If there is a category $c_i \in C_j$ at the position k and $f(d_j, c_i) = 0$ then $precision_j(R_{jk}) = 0$.

For p test documents, the overall performance is obtained by averaging each metric, that is $\text{hloss} = \frac{1}{p} \sum_{j=1}^{p} \text{hloss}_j$, $\text{one-error} = \frac{1}{p} \sum_{j=1}^{p} \text{one-error}_j$, $\text{coverage} = \frac{1}{p} \sum_{j=1}^{p} \text{coverage}_j$, and $\text{avgprec} = \frac{1}{p} \sum_{j=1}^{p} \text{avgprec}_j$. The smaller the value of *hamming loss*, *one-error*, and *coverage*, and the larger the value of *average precision*, the better the performance of the categorization system. The performance is perfect when $\text{hloss} = 0$, $\text{one-error} = 0$, $\text{coverage} = \frac{1}{p} \sum_{j=1}^{p} (|C_j| - 1)$, and $\text{avgprec} = 1$.

3 Vg-ram wnn

RAM-based neural networks, also known as n-tuple categorizers or weightless neural networks (WNN), do not store knowledge in their connections but in Random Access Memories (RAM) inside the network's nodes, or neurons. These neurons operate with binary input values and use RAM as lookup tables: the synapses of each neuron collect a vector of bits from the network's inputs that is used as the RAM address, and the value stored at this address is the neuron's output. Training can be made in one shot and basically consists of storing the desired output in the address associated with the input vector of the neuron [18] (see Figure 1).

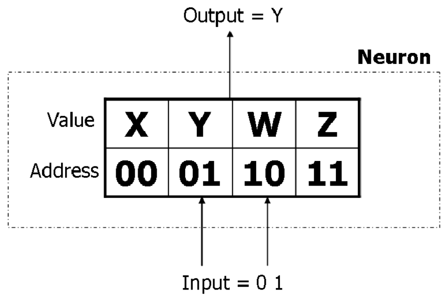


Fig. 1. Weightless neural network.

In spite of their remarkable simplicity, RAM-based neural networks are very effective as pattern recognition tools, offering fast training and test, and easy implementation [14]. However, if the network input is too large, the memory size of the neurons of WNN becomes prohibitive, since it must be equal to 2^n, where n is the input size. Virtual Generalizing RAM (VG-RAM) networks are RAM-based neural networks that only require memory capacity to store the data related to the training set [15]. In the neurons of these networks, the memory stores the input-output pairs shown during training, instead of only the output. In the test phase, the memory of VG-RAM neurons is searched associatively by comparing the input presented to the network with all inputs in the input-output pairs learned. The output of each VG-RAM neuron is taken from the pair whose input is nearest to the input presented—the distance function employed by VG-RAM neurons is the *hamming distance*. If there is more than one pair at the same minimum distance from the input presented, the neuron's output is chosen randomly among these pairs.

Figure 2 shows the lookup table of a VG-RAM neuron with three synapses (X_1, X_2 and X_3). This lookup table contains three entries (input-output pairs), which were stored during the training phase (entry #1, entry #2 and entry #3). During the test phase, when an input vector (input) is presented to the network, the VG-RAM test algorithm calculates the distance between this input vector and each input of the input-output pairs stored in the lookup table. In the example of Figure 2, the *hamming distance* from the input to entry #1 is two, because both X_2 and X_3 bits do not match the input vector. The distance to entry #2 is one, because X_1 is the only non-matching bit. The distance to entry #3 is three, as the reader may easily verify. Hence, for this input vector, the algorithm evaluates the neuron's output, Y, as category 2, since it is the output value stored in entry #2.

To categorize text documents using VG-RAM WNN, we represent a document as a multidimensional vector $V = \{v_1, \ldots, v_{|V|}\}$, where each element v_i corresponds to a weight associated to a specific term in the vocabulary of interest. We use single layer VG-RAM WNN (Figure 3) whose neurons' synapses $X = \{x_1, \ldots, x_{|X|}\}$ are randomly connected to the network's in-

lookup table	X_1	X_2	X_3	Y
entry #1	1	1	0	category 1
entry #2	0	0	1	category 2
entry #3	0	1	0	category 3
	↑	↑	↑	↓
input	1	0	1	**category 2**

Fig. 2. VG-RAM neuron lookup table.

put $N = \{n_1, \ldots, n_{|N|}\}$, which has the same size of the vectors representing the documents, i.e., $|N| = |V|$. Note that $|X| < |V|$ (our experiments have shown that $|X| < |V|$ provides better performance). Each neuron's synapse x_i forms a minchinton cell with the next, x_{i+1} ($x_{|X|}$ forms a minchinton cell with x_1) [19]. The type of the minchinton cell we have used returns 1 if the synapse x_i of the cell is connected to an input element n_j whose value is larger than that of the element n_k to which the synapse x_{i+1} is connected (i.e. $n_j > n_k$); otherwise, it returns zero.

During training, for each document in the training set, the corresponding vector V is connected to the VG-RAM WNN's input N and the neurons' outputs $O = \{o_1, \ldots, o_{|O|}\}$ to the code of one of the categories of the document. All neurons of the VG-RAM WNN are then trained to output this category with this input vector. The training for this input vector is repeated for each category associated with the corresponding document. During test, for each test document, the inputs are connected to the corresponding vector and the number of neurons outputting each category is counted. The network's output is computed by dividing the count of each category by the number of neurons of the network. This output is organized as a vector whose size is equal to the number of categories. The value of each vector element varies from 0 to 1 and represents the percentage of neurons which presented the corresponding category as output (the sum of the values of all elements of this vector is always equal to 1). This way, the output of the network implements the function $f(.,.)$, defined in Section 2. A threshold τ may be used with the function $f(.,.)$ to define the set of categories to be assigned to the test document.

4 ML-KNN

The Lazy Learning Multi-label k-Nearest Neighbors (ML-KNN) categorizer, proposed by Zhang and Zhou [13], is derived from the popular K-nearest neighbor (KNN) algorithm. It is based on the estimate of the probability of a category to be assigned to a test document d_j considering the occurrence of that category on the k nearest neighbors of d_j. If that category is assigned to

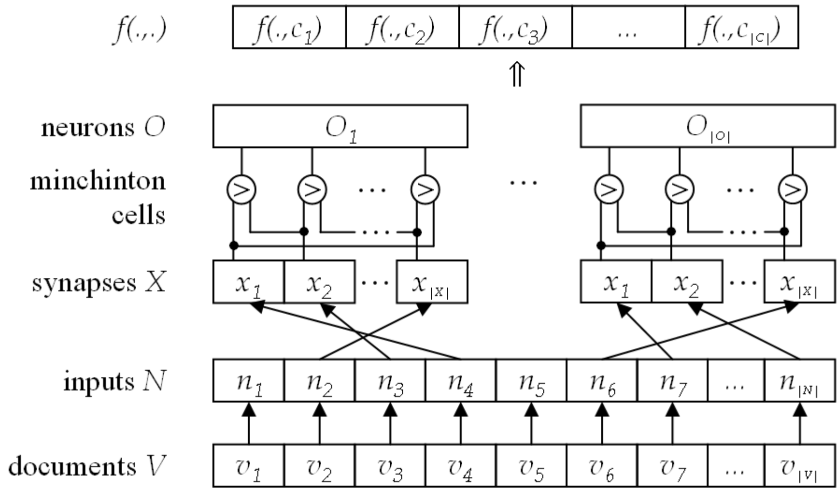


Fig. 3. Vg-ram wnn architecture employed.

the majority (more than 50%) of the k neighbors of d_j, then that category is also assigned to d_j, and not assigned otherwise.

Zhang and Zhou [13] carried out a series of experiments using large data sets, where Ml-knn overcame other categorization approaches, such as Boostexter [3], the multi-label kernel method Rank-svm [5], and the multi-label decision tree Adtboost.MH [6]. This and the fact that Knn is a well known algorithm that can be used to tackle difficult problems in the Information Retrieval area [20,21] have motivated us to use Ml-knn as baseline in the Vg-ram wnn evaluation.

5 Experimental Evaluation

To implement the Vg-ram wnn, we have used the Event Associative Machine (MAE) [22], an open source framework for modeling VG-RAM neural networks developed at the *Universidade Federal do Espírito Santo*. MAE is similar to the Neural Representation Modeler (NRM), developed by the Neural Systems Engineering Group at Imperial College London and commercialized by Novel Technical Solutions [23,24]. However, MAE differs from NRM on three main aspects: it is open source, runs on UNIX (currently, Linux), and uses a textual language to describe WNNs. MAE allows designing, training and analyzing the behavior of modular WNNs whose basic modules are bidimensional neural layers and bidimensional filters. Neural layers' neurons may have several attributes (type, input sensitivity, memory size, etc.) and the user can freely design filters using the C programming language. The MAE user specifies modular WNNs using the MAE Neural Architecture Description

Language (NADL). NADL source files are compiled into MAE applications, which have a built-in graphical user interface and an interpreter of the MAE Control Script Language (CDL). The user can train, test, and alter neural layers contents using the graphical interface or scripts in CDL.

In the following subsections, we evaluate the categorization performance of VG-RAM WNN and compare it against that of ML-KNN using two real-world multi-label learning problems: categorization of free-text descriptions of economic activities, and categorization of Web pages. To perform the experiments with ML-KNN described below, we used a Matlab implementation of it kindly provided by Min-Ling Zhang [13].

5.1 Categorization of Free-text Descriptions of Economic Activities

Currently, most works on automatic text categorization in the literature are focused on categorization of Web pages. However, there are many other important applications to which little attention has hitherto been paid, which are as well very difficult to deal with. One example is the categorization of a company based on its statement of purpose, also called mission statement, which represents the business context of the company activities. Categorization of companies according to their economic activities is an important step of the process of obtaining information for statistical analysis of the economy within a city, state or country.

In many countries, companies must have a contract (Articles of Incorporation or Corporate Charter, in USA) with the society where they can legally operate. In Brazil, this contract is called social contract and must contain the statement of purpose of the company. This mission statement needs to be categorized into a legal business activity by Brazilian government officials. For that, all economic activities recognized by law are cataloged in a table called "*Classificação Nacional de Atividades Econômicas (CNAE)*" (National Classification of Economic Activities) [25]. To perform the categorization, government officials (at the Federal, State and County levels) must find the semantic correspondence between the statement of purpose of the company and one or more entries of the CNAE table. There is a numerical code for each entry of the CNAE table and, in the categorization task, the government official must attribute one or more such codes to the company at hand. This can happen on the foundation of the company or in an eventual change of its social contract, if that modifies its mission statement.

To easy and improve the quality of the categorization of companies according to their economic activities, the Brazilian government is creating a centralized digital library with the statement of purpose of all companies in the country. This library will help the three government levels—the Federal, the 27 States, and the more than 5000 Brazilian Counties—in the task of categorizing Brazilian companies according to the Brazilian law. In order to categorize the mission statement of each company within this digital library into legal economic activities—more than 1000 possible ones—we estimate that data related to more than 5 millions companies will have to be processed. Also, we estimate that at least 300 thousand statements of purpose of new companies, or of companies which are changing their mission statement, will have to be processed every year. It is important to note that the large number of possible categories makes this problem particularly complex when compared with others presented in the literature [1].

To evaluate the performance of VG-RAM WNN on the categorization of economic activities, we used a data set composed of 3264 statements of purpose of Brazilian companies categorized into a subset of 764 CNAE categories. The CNAE codes of each company in this data set were assigned by Brazilian government officials trained in this task. This data set also contains the official brief description of each one of the 764 CNAE categories. We evenly partitioned the whole set of mission statements into four subsets of equal size (816 documents). As the training(-and-validation) set, we adopted the 764 descriptions of CNAE categories and a subset of mission statements with 816 documents, and, as the test set, the other three subsets totalizing 2448 mission statements.

We preprocessed the data set via term selection—a total of 1001 terms were found in the database after removing stop words and trivial cases of gender and plural; only words appearing in the CNAE table were considered. After that, each document in the data set was described as a multidimensional vector using the "Bag-of-Words" representation [26], i.e., each dimension of the vector is a weight associated with one of terms of interest. For this data set, the weight corresponds to the number of times a term in the 1001 terms vocabulary appears in the corresponding document. Table 1 summarizes the characteristics of this data set [1].

To tune the parameters of the VG-RAM WNN categorizer, we divided the training(-and-validation) set into a training set, which was used to inductively build the categorizer, and a validation set, which was used to evaluate the performance of the categorizer in the series of experiments aimed at parameter optimization. The training set is composed of 764 descriptions of CNAE categories and the validation set of 816 mission statements.

Figure 4 shows the MAE application we have built to run the experiments with

[1] Data set available at `http://www.inf.ufes.br/~alberto/vitoria.tar.gz`.

Table 1
Characteristics of the economic activities data set. NC denotes the number of categories, NT denotes the number of terms in the vocabulary, NTD denotes the average number of terms per document, PMC denotes the percentage of documents belonging to more than one category, NCD denotes the average number of categories of each document, and PRC denotes the percentage of rare categories, i.e., those categories associated with less than 1% of the documents.

NC	NT	Training(-and validation) set				Test set			
		NTD	PMC	NCD	PRC	NTD	PMC	NCD	PRC
764	1001	8.11	37.91%	2.63	92.93%	10.79	74.84%	4.31	85.08%

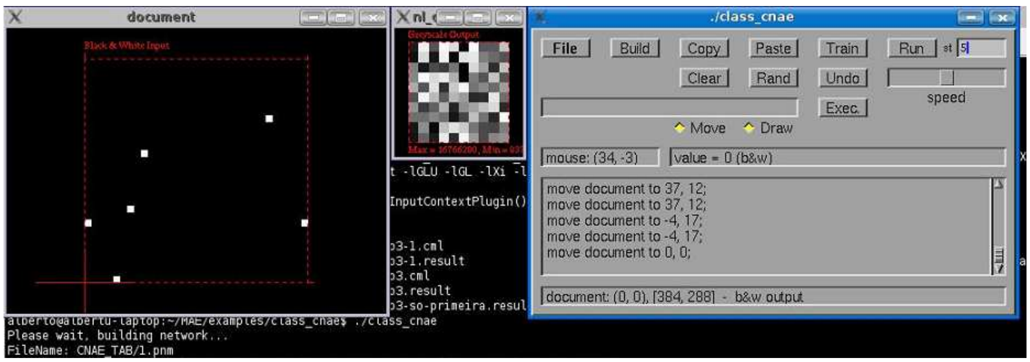


Fig. 4. MAE Application.

the VG-RAM WNN configured 10×10 neurons with 256 synapses each. In the MAE application, the window named document shows the vectors representing the documents been trained or tested. The 1001 elements of these vectors are transformed in a 32×32 input neuron layer (23 of the $32 \times 32 = 1024$ elements of this neuron layer are always filled with zero), which can be shown as a 32×32 pixel image. The outputs of the 100 (10×10) neurons of the network form the 10×10 window shown in the middle of Figure 4, while the left window, named class_cnae, is the graphical user interface of this MAE application.

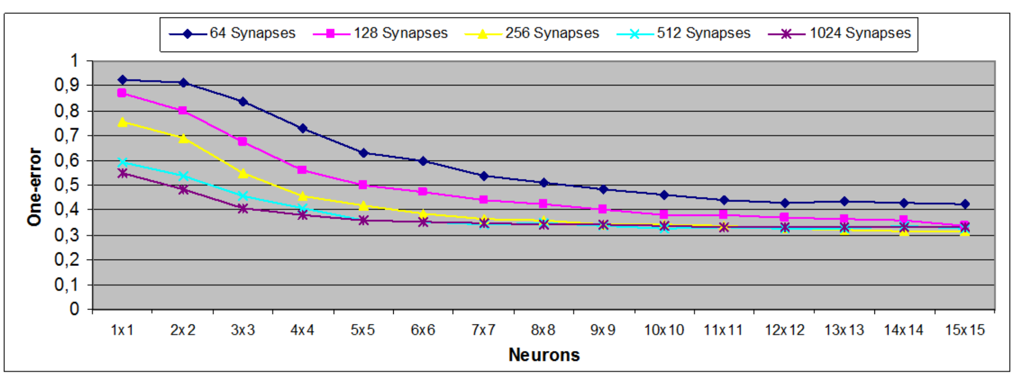


Fig. 5. Results of validation experiments aimed at tuning the number of neurons and synapses per neuron of the VG-RAM WNN.

Figure 5 presents some results of the validation experiments employed for tuning the number of neurons and synapses per neuron of the VG-RAM WNN. The graph in the figure shows the categorization performance in terms *one-error* as a function of the number of neurons and the number of synapses per neuron. As Figure 5 shows, VG-RAM WNN performance increases (*one-error* decreases) with the number of neurons of the network, but levels off when the network has about 144 (12×12) neurons. The reason is that, with a small number of neurons, the network cannot discriminate well between the many CNAE categories, but as the number of neurons increases above 144, additional neurons do not augment the discriminative power of the network. The performance also increases with the number of synapses per neuron, but again levels off at about 256 synapses.

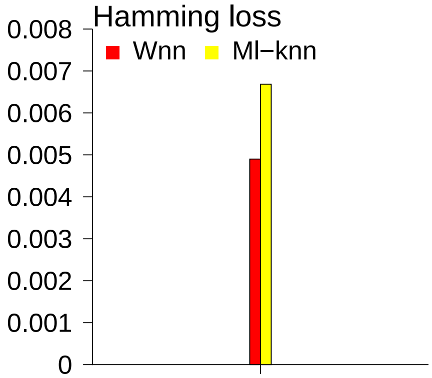
Our validation experiments showed that the values of the parameters of the VG-RAM WNN that yield the best performance for this data set in terms of the four multi-label evaluation metrics adopted (Section 2) are 144 neurons, 256 synapses per neuron, and a threshold equal to 0.08. For ML-KNN, we used the Euclidean metric to measure distances between documents and performed validation experiments to tune the number of neighbors. The best performing ML-KNN encountered has 8 neighbours.

After tunning, the multi-label categorizers were trained with the 1580 documents (764 descriptions of CNAE categories and 816 statements of purposes of companies) of the training(-and-validation) set and tested with the 2448 documents of the test set. Figures 6(a) to 6(d) present the experimental results of each multi-label categorization technique on the economic activities data set in terms of *hamming loss*, *one-error*, *coverage*, and *average precision*, respectively. As Figures 6(a) to 6(d) show, VG-RAM WNN outperforms ML-KNN in terms of the four multi-label evaluation metrics adopted, showing gains of 27%, 14%, 5%, and 5%, in terms of *hamming loss*, *one-error*, *coverage*, and *average precision*, respectively.

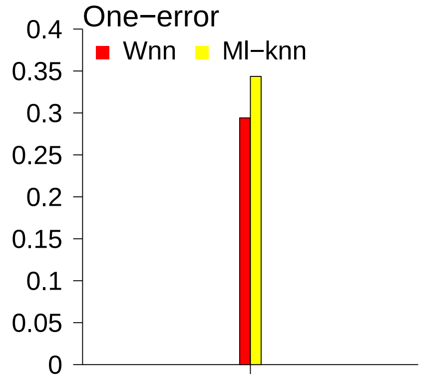
5.2 Categorization of Web Pages

The Web page data employed in our experiments was extracted from the Yahoo directory [2] (`http://dir.yahoo.com`). Currently, the top level of the Yahoo directory consists of 14 Web page categories (i.e.,"Arts", "Business", "Computers", and so on) and each category is further categorized into a number of second-level subcategories. By focusing on these subcategories, one can devise 14 independent text categorization problems. Zhang and Zhou [13] used 11 of these 14 problems to evaluate the performance of ML-KNN. To reduce the dimensionality of each data set, they used a simple term selection method

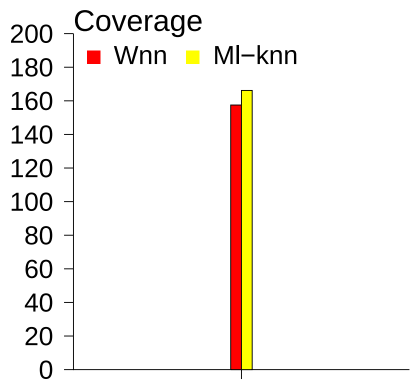
[2] Data set available at `http://www.inf.ufes.br/~alberto/yahoo.tar.gz`.



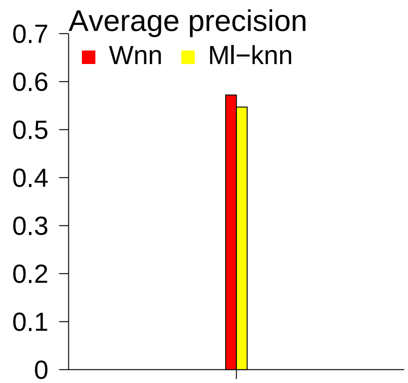
(a) Hamming loss



(b) One-error



(c) Coverage



(d) Average precision

Fig. 6. Experimental results of each multi-label categorizer on the economic activities data set. The smaller the value of *hamming loss*, *one-error*, and *coverage*, and the larger the value of *average precision*, the better.

based on document frequency (the number of documents containing a specific term)—only the top 2% terms with highest document frequency were retained in the final vocabulary. After term selection, each document in the data set was also described as a multidimensional vector using the "Bag-of-Words" representation, i.e. each dimension of the feature vector corresponds to the number of times a word in the vocabulary appears in the document. Table 2 summarizes the characteristics of the Web page data sets [3]. For each data set, the training(-and-validation) set contains 2000 documents while the test set contains 3000 documents.

To tune the parameters of the VG-RAM WNN and ML-KNN categorizers for these data sets, we divided the 2000 documents training(-and-validation) set

[3] The characteristics of the Web page data sets were obtained from the work presented in [13].

Table 2
Characteristics of the Web page data sets (after term selection). NC denotes the number of categories, NT denotes the number of terms in the vocabulary, NTD denotes the average number of terms per document, PMC denotes the percentage of documents belonging to more than one category, NCD denotes the average number of categories of each document, and PRC denotes the percentage of rare categories, i.e., those categories associated with less than 1% of the documents of a the data set.

Data set	NC	NT	Training(-and-validation) set				Test set			
			NTD	PMC	NCD	PRC	NTD	PMC	NCD	PRC
Arts	26	462	34.59	44.50%	1.63	19.23%	34.98	43.63%	1.64	19.23%
Business	30	438	34.27	42.20%	1.59	50.00%	33.69	41.93%	1.59	43.33%
Computers	33	681	47.34	29.60%	1.49	39.39%	46.22	31.27%	1.52	36.36%
Education	33	550	41.20	33.50%	1.47	57.58%	42.85	33.73%	1.46	57.58%
Entertainment	21	640	47.69	29.30%	1.43	28.57%	48.93	28.20%	1.42	33.33%
Health	32	612	42.25	48.05%	1.67	53.13%	42.88	47.20%	1.66	53.13%
Recreation	22	606	43.46	30.20%	1.41	18.18%	44.49	31.20%	1.43	18.18%
Reference	33	793	56.78	13.75%	1.16	51.52%	57.15	14.60%	1.18	54.55%
Science	40	743	59.02	34.85%	1.49	35.00%	58.91	30.57%	1.43	40.00%
Social	39	1 047	65.65	20.95%	1.27	56.41%	63.28	22.83%	1.29	58.97%
Society	27	636	56.05	41.90%	1.71	25.93%	54.49	39.97%	1.68	22.22%

of each problem into a 1500 documents training set, which was used to inductively build the categorizers, and a 500 documents validation set, which was used to evaluate the performance of the categorizers in the series of experiments aimed at parameter optimization. Table 3 shows, for each one of the 11 text categorization problems, the parameters that yield the best VG-RAM WNN performance. For ML-KNN, the number of nearest neighbors that shown best performance was 10 [4].

For each data set, the multi-label categorizers were trained with the 2000 documents of the training(-and-validation) set and tested with the 3000 documents of the test set. Figures 7 to 10 present the experimental results obtained with each multi-label categorization technique on all the Web page data sets in terms of *hamming loss*, *one-error*, *coverage*, and *average precision*, respectively. The graphs in the figures also show the averages for each evaluation metric over all data sets. As Figures 7 to 10 show, on average, VG-RAM WNN performs better than ML-KNN in terms of *hamming loss*, *coverage*, and *average precision*, and shows a comparable performance in terms of *one-error*. When considering the data sets separately (see the results for "Recreation"), VG-RAM WNN shows gains over ML-KNN of up to 11%, 19%, 21%, and 24%, in terms of *hamming loss*, *one-error*, *coverage*, and *average precision*, respectively.

[4] The results for ML-KNN were obtained from [13].

Table 3
Parameters of VG-RAM WNN that yield the best performance.

Data set	Number of neurons	Number of synapses	Threshold τ
Arts	1024	64	0.2
Business	1024	64	0.2
Computers	1024	64	0.4
Education	1024	128	0.4
Entertainment	1024	128	0.3
Health	1024	128	0.2
Recreation	1024	64	0.2
Reference	1024	64	0.5
Science	1024	64	0.2
Social	1024	128	0.4
Society	1024	64	0.3

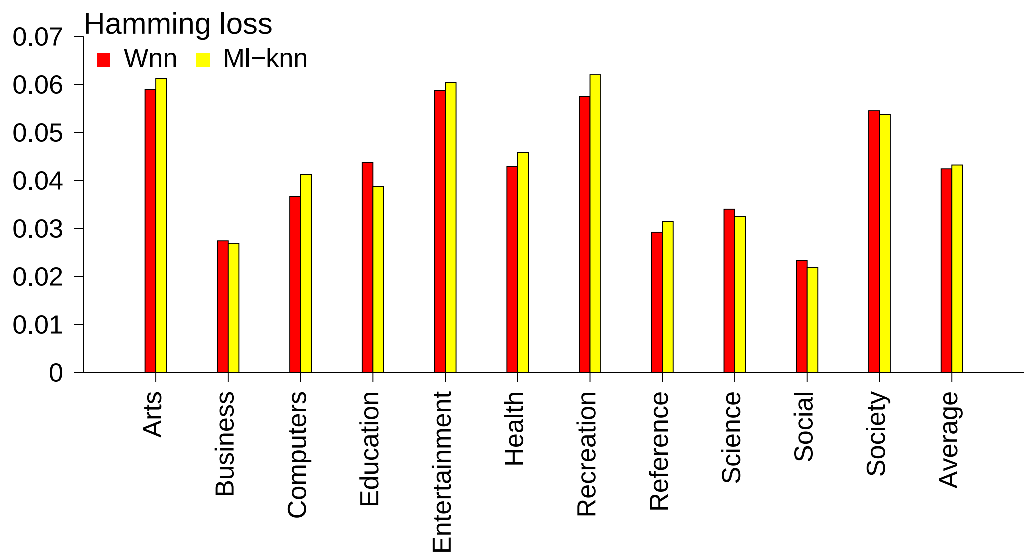


Fig. 7. Experimental results of each multi-label categorizer on the Web page data sets in terms of *hamming loss*. The smaller the value of *hamming loss*, the better.

5.3 Rationale for the Results

In the categorization of free-text descriptions of economic activities, the VG-RAM WNN categorizer outperformed the ML-KNN categorizer in terms of the four multi-label evaluation metrics employed, while in the categorization of Web pages, on average, VG-RAM WNN outperformed the ML-KNN in terms

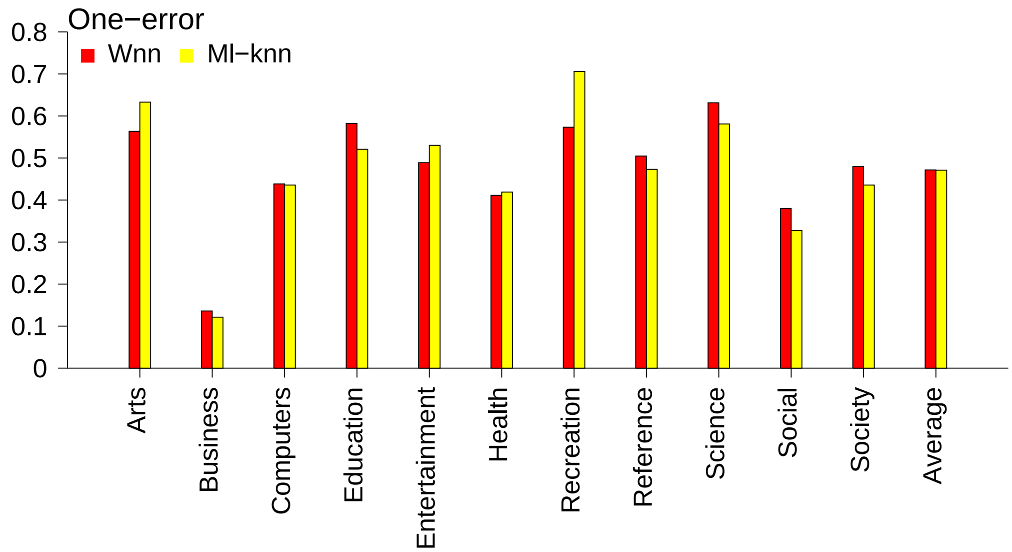


Fig. 8. Experimental results of each multi-label categorizer on the Web page data sets in terms of *one-error*. The smaller the value, the better.

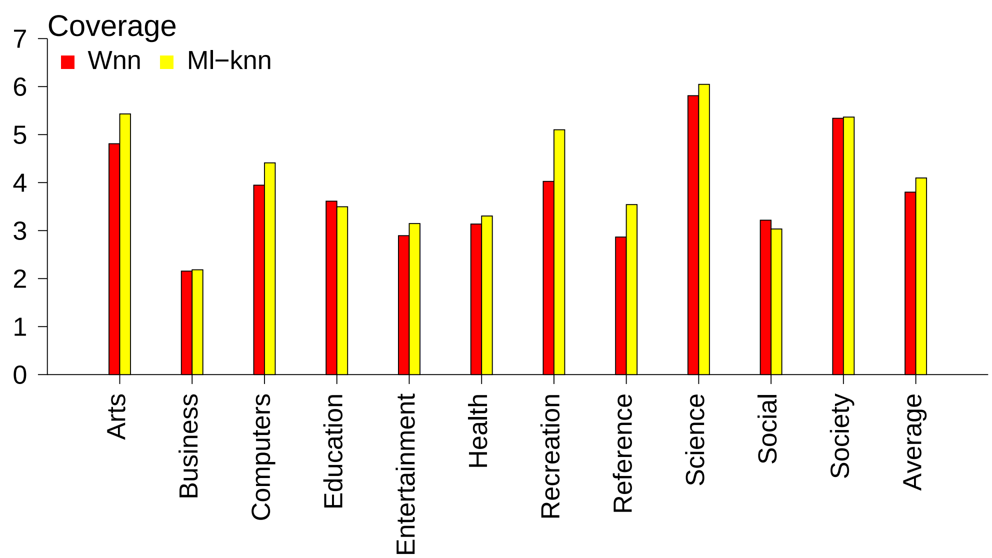


Fig. 9. Experimental results of each multi-label categorizer on the Web page data sets in terms of *coverage*. The smaller the value, the better.

of *hamming loss*, *coverage* and *average precision*, and showed similar categorization performance in terms of *one-error*. The VG-RAM WNN superior performance is the result of two factors that we discuss below.

First, each VG-RAM WNN synapse collects the result of a comparison between two term weights, executed by its corresponding minchinton cell. For the economic activities data set, our best VG-RAM WNN has 256 synapses per neuron and 12×12 neurons. Therefore, during test, 36864 ($256 \times 12 \times 12$) such compar-

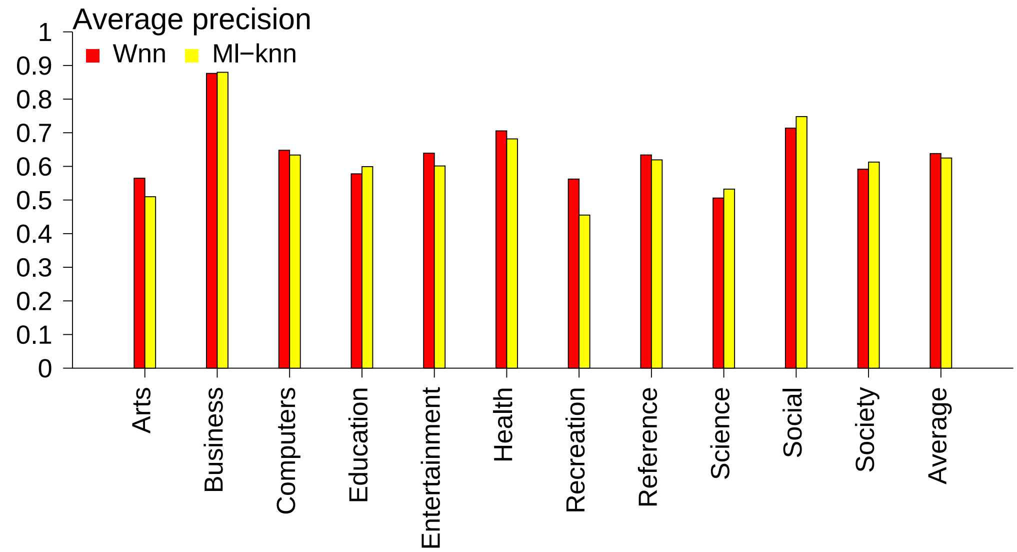


Fig. 10. Experimental results of each multi-label categorizer on the Web page data sets in terms of *average precision*. The larger the value, the better.

isons are executed on an input vector representing a document and the results are checked against equivalent results learned from training documents. For the Web page data sets, the best VG-RAM WNN has 128 synapses per neuron and 32×32 neurons. Thus, the number of comparisons between term weights reaches 131072 ($128 \times 32 \times 32$). This amount of term weight comparisons allows not only high discrimination capability but also generalization.

Second, the large number of neurons of the VG-RAM WNN examined allows an equivalently large number of votes for different categories and each of these neurons samples the documents in different ways. Therefore, the votes help producing a highly discriminative function $f(.,.)$ or ranking function $r(.,.)$ (see Section 2), which are the core of the metrics employed in the evaluation.

6 Conclusions and Future Work

In this work, we presented an experimental evaluation of the performance of Virtual Generalizing Random Access Memory Weightless Neural Networks (VG-RAM WNN [14]) on multi-label text categorization. We performed a comparative study of VG-RAM WNN and the multi-label lazy learning technique ML-KNN [13] using two multi-label problems: categorization of free-text descriptions of economic activities, and categorization of Web pages. In the problem of categorization of free-text descriptions of economic activities, VG-RAM

WNN outperformed ML-KNN in terms of the four multi-label evaluation metrics adopted, while, in the categorization of Web pages, on average, VG-RAM WNN outperformed ML-KNN in terms of *hamming loss*, *coverage*, and *average precision*, and showed similar categorization performance in terms of *one-error*.

A direction for future work is to compare VG-RAM WNN performance against other multi-label text categorization methods. Other direction for future research is to examine correlated VG-RAM WNN [27] and other mechanisms for taking advantage of the correlation between categories. Another direction for further research is to evaluate the categorization performance of VG-RAM WNN using different multi-label categorization problems, such as image annotation and gene function prediction.

7 Acknowledgements

We would like to thank Min-Ling Zhang for all the help with the ML-KNN categorization tool and Web page data sets. We would also like to thank *Receita Federal do Brasil*, *Conselho Nacional de Desenvolvimento Científico e Tecnológico* — CNPq-Brasil (grants 308207/2004-1, 471898/2004-0, 620165/2006-5, 309831/2007-5), *Financiadora de Estudos e Projetos* — FINEP-Brasil (grants CT-INFRA-PRO-UFES/2005, CT-INFRA-PRO-UFES/2006), and *Fundação Espírito Santense de Tecnologia* — FAPES-Brasil (grant 37711393/2007) for their support to this research work.

References

[1] F. Sebastiani, Machine learning in automated text categorization, ACM Computing Surveys 34 (1) (2002) 1–47.

[2] A. McCallum, Multi-label text classification with a mixture model trained by EM, in: Working Notes of the AAAI'99 Workshop on Text Learning, 1999, pp. 1–7.

[3] R. E. Schapire, Y. Singer, BoosTexter: a boosting-based system for text categorization, Machine Learning 39 (2/3) (2000) 135–168.

[4] A. Clare, R. D. King, Knowledge discovery in multi-label phenotype data, in: Lecture Notes in Computer Science, Vol. 2168, 2001, pp. 42–53.

[5] A. Elisseeff, J. Weston, A kernel method for multi-labelled classification, in: Advances in Neural Information Processing Systems, Vol. 14, MIT Press, 2002, pp. 681–687.

[6] F. D. Comité, R. Gilleron, M. Tommasi, Learning multi-label alternating decision tree from texts and data, in: Lecture Notes in Computer Science, Vol. 2734, Springer, 2003, pp. 35–49.

[7] N. Ueda, K. Saito, Parametric mixture models for multi-label text, in: Advances in Neural Information Processing Systems, Vol. 15, MIT Press, 2003, pp. 721–728.

[8] M. R. Boutell, J. Luo, X. Shen, C. M. Brown, Learning multi-label scene classification, Pattern Recognition 37 (9) (2004) 1757–1771.

[9] S. Gao, W. Wu, C.-H. Lee, T.-S. Chua, A MFoM learning approach to robust multiclass multi-label text categorization, in: Proceedings of the 21st International Conference on Machine Learning, 2004, pp. 329–336.

[10] E. Romero, L. Màrquez, X. Carreras, Margin maximization with feed-forward neural networks: a comparative study with svm and adaboost, Neurocomputing 57 (2004) 313–344.

[11] H. Kazawa, T. Izumitani, H. Taira, E. Maeda, Maximal margin labeling for multi-topic text categorization, in: Advances in Neural Information Processing Systems, Vol. 17, MIT Press, 2005, pp. 649–656.

[12] M.-L. Zhang, Z.-H. Zhou, Multi-label neural networks with applications to functional genomics and text categorization, IEEE Transactions on Knowledge and Data Engineering 18 (10) (2006) 1338–1351.

[13] M.-L. Zhang, Z.-H. Zhou, ML-KNN: A lazy learning approach to multi-label learning, Pattern Recognition 40 (7) (2007) 2038–2048.

[14] I. Aleksander, RAM-Based Neural Networks, World Scientific, 1998, Ch. From WISARD to MAGNUS: a Family of Weightless Virtual Neural Machines, pp. 18–30.

[15] T. B. Ludermir, A. C. P. L. F. Carvalho, A. P. Braga, M. D. Souto, Weightless neural models: a review of current and past works, Neural Computing Surveys 2 (1999) 41–61.

[16] A. F. D. Souza, F. Pedroni, E. Oliveira, P. M. Ciarelli, W. F. Henrique, L. Veronese, Automated free text classification of economic activities using vg-ram weightless neural networks, in: Proceedings of the 7th IEEE International Conference on Intelligent Systems Design and Applications, IEEE Computer Society, 2007, pp. 782–787.

[17] R. E. Schapire, Y. Singer, Improved boosting algorithms using confidence-rated predictions, Machine Learning 27 (3) (1999) 297–336.

[18] I. Aleksander, Self-adaptive universal logic circuits, IEE Electronic Letters 2 (8) (1966) 231–232.

[19] R. J. Mitchell, J. M. Bishop, S. K. Box, J. F. Hawker, RAM-Based Neural Networks, World Scientific, 1998, Ch. Comparison of Some Methods for Processing Grey Level Data in Weightless Networks, pp. 61–70.

[20] C. Yu, B. Cui, S. Wang, J. Su, Efficient index-based knn join processing for high-dimensional data, Inf. Softw. Technol. 49 (4) (2007) 332–344.

[21] R. Bueno, A. J. M. Traina, J. Caetano Traina, Genetic algorithms for approximate similarity queries, Data Knowl. Eng. 62 (3) (2007) 459–482.

[22] K. S. Komati, A. F. De Souza, Using weightless neural networks for vergence control in an artificial vision system, Applied Bionics and Biomechanics 1 (1) (2003) 21–31.

[23] I. Aleksander, C. Browne, B. Dunmall, T. Wright, Brain-Like Computing and Intelligent Information Systems, Springer, 1997, Ch. Towards Visual Awareness in a Neural System, pp. 513–533.

[24] Novel Technical Solutions, Neural Representation Modeller, `http://www.ntsu.com` (2007).

[25] CNAE, Classificação Nacional de Atividades Econômicas - Fiscal (CNAE-Fiscal) 1.1, Tech. rep., Instituto Brasileiro de Geografia e Estatística (IBGE), Rio de Janeiro, RJ (2003).

[26] S. T. Dumais, J. Platt, D. Heckerman, M. Sahami, Inductive learning algorithms and representation for text categorization, in: Proceedings of the 7th ACM International Conference on Information and Knowledge Management, 1998, pp. 148–155.

[27] R. Carneiro, S. S. Dias, D. F. Jr., S. Oliveira, A. S. d. Garcez, A. F. De Souza, Improving VG-RAM neural networks performance using knowledge correlation, Lecture Notes in Computer Science 4232 (2006) 427–436.

 **Dr. Alberto Ferreira De Souza** is a Professor of Computer Science and Coordinator of the Laboratório de Computação de Alto Desempenho (LCAD — High Performance Computing Laboratory) at the Universidade Federal do Espírito Santo (UFES), Brazil. He received B.Eng. (Cum Laude) in Electronics Engineering and M.Sc. in Systems Engineering and Computer Science from Universidade Federal do Rio de Janeiro (COPPE/UFRJ), Brazil, in 1988 and 1993, respectively; and Doctor of Philosophy (Ph.D) in Computer Science from the University College London, United Kingdom in 1999. He has authored/co-authored one USA patent and over 60 publications. He has edited proceedings of four conferences (two IEEE sponsored conferences), is a Standing Member of the Steering Committee of the International Conference in Computer Architecture and High Performance Computing (SBAC-PAD), and Coordinator of the Technical Committee on Computer Architecture and High Performance Computing of the Brazilian Computer Society (SBC).

Dr. De Souza held the following positions in UFES and elsewhere: member of the Board of Directors of the Regional Council of Engineering of Espírito Santo — CREA-ES (1995-1996), Vice-Dean of the School of Engineering — UFES (2001-2004), Director Superintendent of the Institute of Technology — UFES (2003-2004), and Pro-Provost of Planning and Development — UFES (2004-2007). At the present time, Alberto is Vice-President of the Administrative Council of the Espírito Santo Science and Technology Foundation (FEST) — FEST, and president of Steering Committee of the Vitória High Speed Metropolitan Area Network (METROVIX) — METROVIX.

Alberto Ferreira De Souza is Comendador of the order of Rubem Braga.


Felipe Pedroni received the B.Eng. degree in Computer Engineering from the Universidade Federal do Espírito Santo (UFES), Brazil, in 2006. From 2003 to 2008, he worked at the Laboratório de Computação de Alto Desempenho (LCAD — High Performance Computing Laboratory) at UFES. Since 2006, he has been pursuing his M.Sc. degree in Computer Science from UFES. His research interests include neural networks, information retrieval, text categorization and computer architecture.


Dr. Elias Oliveira is currently a Lecturer at the Information Science Department at the Universidade Federal do Espírito Santo (UFES), Brazil. He received the B.Sc. degree in Mathematics from Universidade Federal do Rio de Janeiro (UFRJ) and was awarded his Ph.D. in Computer Science in 2001, from School of Computing of the University of Leeds, UK. Dr. Oliveira is a collaborator of the Laboratório de Computação de Alto Desempenho (LCAD – High Performance Computing Laboratory) at UFES. His research interests include machine learning, evolutionary algorithms, information retrieval, text categorization, practical applications of constraint programming, and neural networks applications.


Patrick Marques Ciarelli received his B.Eng. and MSc degrees in Electrical Engineering in 2006 and 2008, respectively, from the Universidade Federal do Espírito Santo (UFES), Brazil. Currently, he is doing doctorate in Electrical Engineering at UFES. His current research interests include neural networks, information retrieval, pattern recognition, and image processing.

**Wallace F. Henrique** is an undergraduate student in Computer Science at the Universidade Federal do Espírito Santo (UFES), Brazil. Since 2006, he has been working at the Laboratório de Computação de Alto Desempenho (LCAD — High Performance Computing Laboratory) at UFES. His research interests include information retrieval and text categorization.

**Lucas Veronese** is an undergraduate student in Computer Science at the Universidade Federal do Espírito Santo (UFES), Brazil. Since 2006, he has been working at the Laboratório de Computação de Alto Desempenho (LCAD — High Performance Computing Laboratory) at UFES. His research interests include information retrieval and text categorization.

 **Dr. Claudine Badue** is an Associate Researcher and member of the Laboratório de Computação de Alto Desempenho (LCAD — High Performance Computing Laboratory) at the Universidade Federal do Espírito Santo (UFES). In 1998, she received the B.Sc. degree in Computer Science from the Universidade Federal de Goiás (UFG), Brazil. She received the M.Sc. degree in Computer Science in 2001 and the Ph.D. degree in Computer Science in 2007, both from the Universidade Federal de Minas Gerais (UFMG), Brazil.

Her research interests are in the areas of information retrieval, text categorization, and performance analysis and modeling. She has been involved in research projects financed through Brazilian research agencies such as the National Research Council (CNPq) and the Espírito Santo Science and Technology Foundation (FEST). She has also been in the program committee and organizing committee of national and international conferences in Computer Science.