

CYBERNETICS

ALGORITHM FOR CONSTRUCTION
OF PARAMETRIC VECTORS FOR SOLUTION
OF CLASSIFICATION PROBLEMS
BY A FEED-FORWARD NEURAL NETWORKA. V. Anisimov,^a E. V. Porkhun,^b and V. Yu. Taranukha^a

UDC 004.912

A universal heuristic algorithm is considered that constructs the vector of properties of an object that are distributed among heterogeneous groups. This algorithm is based on a feed-forward neural network. An automatic system for objects classification is described. The use of the automatic classification system for a literary authorship attribution problem is considered.

Keywords: *input vector of attributes, heterogeneous groups of attributes, feed-forward neural network, criterion of estimation of training, optimal sample.*

PROBLEM STATEMENT

An important problem of object classification is the construction of the attribute space of an object or, in terms of pattern recognition theory, an output description of an object. A successful choice of such a description can reduce the classification procedure to the solution of a trivial problem and, on the contrary, a poor-quality and inefficient choice of an attribute collection can complicate information processing or make the problem completely undecidable. The body of information obtained as a result of investigation of an object should be transformed, and some functions of its output description are found during such transformations. The objective is the obtaining of an output description of an object or an attribute collection in terms of which the classification problem can be aptly solved.

In forming a collection of parameters of an object, the posteriori chosen attributes can be distributed among heterogeneous groups each of which characterizes the object according to the values of the attributes that are contained in it. This essentially complicates the construction of automated systems oriented toward the recognition of specific properties of objects. For example, in problems of recognition and classification of texts in natural languages, morphological, syntactic, semantic, and other characteristics that correspond to definite levels of analysis of a natural language can be used as attributes identifying the text being processed. In automated systems of medical diagnostics, the symptoms of a disease are determined by data obtained with the help of technical devices such as electrocardiographs, X-ray apparatus, blood analyzers, etc. In systems of technical diagnostics, attributes of objects are formed from data obtained from temperature, pressure, strain, vibroacoustic, and other sensors. Thus, in such systems, the attribute collection of an object is constructed from heterogeneous groups of characteristics obtained as a result of using different technical means or programs.

In problems of classification of objects whose attribute collections are distributed over heterogeneous groups, data of such groups should be jointly processed to provide the influence of each attribute on the classification process. The attributes chosen a posteriori to characterize a set of objects can be often inhomogeneously distributed not only over the entire corpus of the objects being investigated but also within one class. Therefore, a relationship between the components of an attribute vector should be found that makes it possible to take into account the importance of each attribute for the entire corpus of objects. In many cases, by virtue of the heterogeneity and nonuniformity of parametric parameters, the search for such a

^aTaras Shevchenko University, Kiev, Ukraine, ava@mi.unicyb.kiev.ua. ^bInternational Scientific-Educational Center of Information Technologies and Systems, Kiev, Ukraine, eillen2003@ukr.net. Translated from *Kibernetika i Sistemnyi Analiz*, No. 2, pp. 3-14, March-April 2007. Original article submitted December 26, 2006.

relationship presents a sufficiently difficult problem whose solution depends on the classification means being used and on the corresponding algorithm of processing input parameters.

In the present article, a universal heuristic algorithm is proposed for construction of a vector of attributes distributed over heterogeneous groups with the help of a feed-forward neural network. The method lies in constructing a function of transformation of an object description obtained as a result of choice of definite attributes. According to the proposed algorithm, such a transformation function for a given vector is constructed during training such a neural network. After each series of training and testing this neural network, the best result of classification at the current stage is determined and the next distribution of the components of the attribute vector is found. On the condition that the algorithm comes to an end, we obtain a function that transforms the input vector components and according to which the output description of the corresponding object is constructed.

A system of automatic classification of objects specified by numerical collections of chosen attributes was developed with the help of a feed-forward neural network and the proposed method of construction of input attribute vectors. The use of the developed automatic classification system for the solution of problems of automatic attribution of literary works is considered.

Along with traditional frequency and morphological characteristics, more sophisticated attributes describing the construction of sentences are introduced into the system. A new method is developed for automatic determination of the parameters that describe the complexity of sentences in texts, namely, the average number of adjectives, participles, and verbal adverbs corresponding to a noun. As a demonstration of capabilities of the system, the solution of the well-known problem of literary criticism, namely, the authorship of V. Nabokov concerning the writing "Novel with Cocaine" published under the name of M. Ageyev is authenticated with a high degree of likelihood.

DEVELOPMENT OF AN AUTOMATIC CLASSIFICATION SYSTEM

Construction of a neural network. As a basic tool for the development of a system of classification of objects whose output description can be specified in the form of numerical vectors, we propose to use a feed-forward neural network and to train it by the method of error backpropagation training. The developed system uses a model of feed-forward neural networks that realizes associative memory in the form of a single-layer perceptron whose hidden layer maps stored images into a code space. To classify a corpus of objects after completion of the process of training or storing etalon examples by the neural network, it suffices to apply the corresponding formed attribute vectors to the input of the network. Using the code-weight table obtained during training, the network classifies the objects applied to it by assigning objects with similar attribute vectors to one class.

The neural network constructed by the authors consists of the input, output, and hidden layers of neurons. For each object that should be classified, its formed attribute vector is applied to the input of the network. The number of neurons of the input layer of the network is equal to the dimension of this vector. To the output layer of neurons corresponds a zero vector with one unit component indicating the number of the class to which belongs the object being classified. The number of neurons in the output layer is equal to the number of existing classes for the objects from the training sample.

The hidden layer is divided into three blocks of neurons that function with different activation functions (Fig. 1). For each block of neurons, the functions $\exp(-x^2)$, $1 - (\exp(-x^2))$, and $\text{th}(x)$, respectively, are used as output functions. For the output layer of neurons, the sigmoid function $1/(1 + \exp(-x))$ is used.

The partition of the hidden layer into blocks with different output functions makes it possible to disclose some distribution of the parameters of an input vector. Taking into account these empirically chosen functions, the network reacts to definite regularities in the distribution of parameters of vectors of similar objects, which allows determining the class to which a given object belongs.

To train the neural network, the classical method of error backpropagation learning was used that includes the following three stages: the direct passage of the input signal of a training example (the feedforward stage), error back propagation (the backpropagation stage), and weights tuning.

Coding network inputs. Despite the fact that an input vector contains only numerical attributes, it is important to normalize data. Theoretically, training is also possible without normalization (then the values of attributes are directly applied to the input) but the network is trained too slowly in this case. This is conditioned by the fact that the values of attributes, as a rule, essentially exceed the values of weights and, hence, the values of neurons get into the saturation domain of threshold functions and, as a result, identical values that do not depend on inputs are obtained at the output.

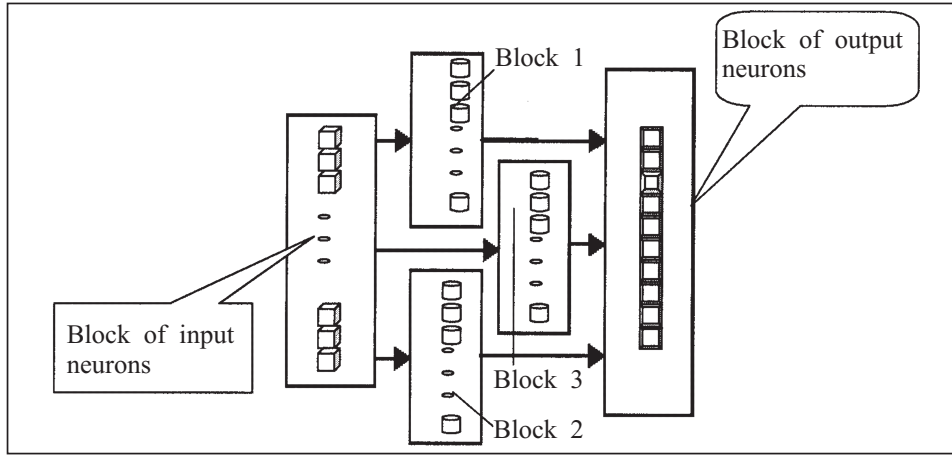


Fig. 1. Network architecture.

Neural networks are characterized by the existence of an interval of input signals within which signals are distinguishable. This is connected with the properties of the threshold functions being used. To a neuron of each type corresponds its interval of admissible input data. As a rule, this range coincides with the range of output signals or is the union of the range of output signals and the segment symmetric to it with respect to zero. We denote this range by $[a, b]$. To maximally use the range of admissible input data, numerical signals should be scaled and shifted so that the whole range of values is in the range of admissible input data.

To normalize the input data of the classification system, numerical signals are processed. A new signal obtained as a result of such a processing is specified by the scaling formula

$$c' = \frac{(c - c_{\min})(b - a)}{(c_{\max} - c_{\min})} + a,$$

where $[a, b]$ is the range of input signals and $[c_{\min}, c_{\max}]$ is the range of values of an attribute c .

Initialization of network weights. The choice of the initial network weights determines whether the network achieves the global (or only a local) minimum error during training and the convergence rate of training.

According to the error back-propagation algorithm, the correction to the weight of the connection between two neurons depends on the values of the derivative of the activation output function of the neuron of the overlying layer and the output signal of the neuron of the underlying layer. Therefore, the choice of initial weights such that at least one of these two values is equal to zero should be avoided. The initial values of weights and corrections must not be too large since, in this case, the initial input signals applied to each hidden or output neuron can get into a domain in which the values of the derivative of the output function are too small (the so-called saturation region). However, if the initial weights are too small, the input signals applied to the neurons of the hidden or output layer can be close to zero, which also essentially slows down training.

In the neural network constructed, weights are initialized using the Nguyen-Widrow weight initialization method (scheme) [1, pp. 289–316].

To improve the ability to train hidden neurons, it is proposed to use this method for initialization of the weights of connections between the neurons of the input and hidden layers. The weights of connections between the hidden and output layers are initialized by random values in the range $[-0.5; 0.5]$. The essence of the method lies in distributing the initial values of weights so that, for each input example, the input signal of the network of one of the hidden neurons can get into the domain in which this neuron is most trainable.

We introduce the following designations: n is the number of input neurons or the dimension of the input vector, p is the number of hidden neurons, β is the following coefficient: $\beta = 0.7(p)^{1/n} = 0.7\sqrt[n]{p}$, v_{ij} is the weight of the connection between the i th neuron of the input layer ($i = 1, \dots, n$) and the j th neuron of the hidden layer ($j = 1, \dots, p$), and v_j is the weight vector for the j th hidden neuron consisting of connection weights v_{ij} .

For each j th hidden neuron, the weight vector v_j is initialized by numbers from a range $[-\gamma, \gamma]$ (here, $\gamma = 0.5$).

Next, the norm of the j th vector v_j is computed as follows:

$$\|v_j\| = \sqrt{v_{1j}^2 + v_{2j}^2 + \dots + v_{nj}^2}.$$

Then the initial weight values are computed by the formula

$$v_{ij}(\text{new}) = \frac{\beta v_{ij}}{\|v_j\|}.$$

METHOD OF FORMATION OF INPUT ATTRIBUTE VECTORS

A distinctive feature of the system developed is that the input attribute vector of an object is formed for the network during training it. In training the neural network, the normalized input vector of attributes that are initially distributed among heterogeneous groups is constructed. To this end, the system search for a relationship that exists between the components of the vector and that allows the system to take into account the influence of each attribute on the classification process and, as a result, to tune the system of recognition of object classes according to these attributes.

We assume that the collection of attributes that characterize object samples consists of two heterogeneous groups α and β . It is proposed to tune the components of the input vector as follows.

We define three types of samples of examples.

Sample A. It consists of vectors that contain only attributes of the group α , and the other components of the vector are zeros,

$$p = (\alpha_1, \alpha_2, \dots, \alpha_i, \dots, \alpha_m, \underbrace{0, 0, \dots, 0, \dots, 0}_k),$$

where p is an example vector from a training sample and $\alpha_i, i=1, \dots, m$, is an attribute of the group α .

Sample B. It contains vectors consisting of the attributes of the group β , and the components corresponding to the attributes of the group α are zeros,

$$p = (\underbrace{0, 0, \dots, 0, \dots, 0}_m, \beta_1, \beta_2, \dots, \beta_j, \dots, \beta_k),$$

where $\beta_j, j=1, \dots, k$, is an attribute of the group β .

Sample C. It includes vectors that consist of weighed values of attributes of the groups α and β ,

$$p = \begin{cases} ((\alpha_1 * \sqrt{b/a}), (\alpha_2 * \sqrt{b/a}), \dots, (\alpha_m * \sqrt{b/a}), \beta_1, \beta_2, \dots, \beta_k), & a < b, \\ (\alpha_1, \alpha_2, \dots, \alpha_m, (\beta_1 * \sqrt{a/b}), (\beta_2 * \sqrt{a/b}), \dots, (\beta_k * \sqrt{a/b})), & a > b, \end{cases}$$

where $a = \alpha_1^2 + \alpha_2^2 + \dots + \alpha_m^2$ and $b = \beta_1^2 + \beta_2^2 + \dots + \beta_k^2$.

The neural network is independently trained using each sample A, B , and C , and then the criterion of estimation of training is checked.

As a test example for the trained network, it makes sense to use the attribute vector of an object from the training sample. By the principle of construction of vectors of the samples A, B , and C , three attribute vectors are formed for this test example.

The criterion of estimation of training the network is understood to be the distance between the following vectors: the answer obtained for the tuned network for the mentioned test example and the answer (class) that is known in advance for this example.

The distance between vectors is determined as usual, namely, by the formula

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2},$$

where $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$ are two vectors. Then the criterion of estimation of training can be specified as follows: $F = \text{Arg min}_{\Omega} D$, where Ω is the set of vectors of samples used for training the network. At the

beginning of the algorithm, the set Ω consists of the samples A, B , and C .

During the functioning of the algorithm, new samples of vectors are continuously formed. By the *termination condition for the algorithm* we understand the attainment of a sufficiently small chosen value ε by the criterion F or the situation when the newly formed sample is no better (as to the value of F) than the previous samples. A sample is considered to be *optimal* if the termination condition of the algorithm is satisfied for it.

Let us define a function φ that assigns some real number to each sample of the set Ω . This function is an injective mapping, i.e., for each number from the domain of the function φ , there exists only one sample corresponding to it. Therefore, to search for an optimal sample, it is proposed to use the segment whose extreme points are numbers corresponding to two samples whose values of F are best at the current stage of the algorithm. For the samples A , B , and C , we have $\varphi(A)=0$, $\varphi(B)=1$, and $\varphi(C)=0.5$, respectively. Thus, the segment in which the optimal sample is searched for is of the form $[0, 1]$.

Let V and W be the samples that correspond to the extreme points of the segment, let G be the sample corresponding to the bisecting point of the segment, and let v_{ij} , w_{ij} , and g_{ij} be, respectively, the components of vectors of the samples V , W , and G , $i=1, \overline{k}$, $j=1, \overline{n}$, where k is the number of vectors in a sample (this number is the same for each sample) and n is the dimension of attribute vectors (this dimension is the same for each vector from the set of samples Ω). We denote by η the point (of the segment) that corresponds to the sought-for optimal sample. Then the algorithm of searching for the optimal sample can be formalized as follows.

Algorithm A1. (Search for the optimal sample.)

INPUT: $\Omega = \{A, B, C\}$ is the set of samples of vectors composed of samples of the types A , B , and C .

RESULT: η is the number corresponding to the optimal sample.

1. Train the network using the samples A , B , and C .
2. Form three vectors for the test example according to the principle of construction of vectors of the samples A , B , and C .
3. Compute the values of the criterion F for each sample A , B , and C .
4. **If** the minimal value of the criterion F is attained for the sample C , **then**
 - 4.1. Check the satisfiability of the criterion F over the set $\Omega = \{A, B\}$.
 - 4.2. **If** the sample A is better than the sample B as to the value of F , **then**
 - 4.2.1. Form the sample V from the vectors of the sample A .
 - 4.2.2. $\varphi(V) \leftarrow 0$.
 - 4.2.3. Form the sample W from the vectors of the sample C .
 - 4.2.4. $\varphi(W) \leftarrow 0.5$.
 - Otherwise**
 - 4.2.5. Form the sample V from the vectors of the sample C .
 - 4.2.6. $\varphi(V) \leftarrow 0.5$.
 - 4.2.7. Form the sample W from the vectors of the sample B .
 - 4.2.8. $\varphi(W) \leftarrow 1$.
- 4.3. **Until** the termination condition of the algorithm is not satisfied
 - 4.3.1. Form the sample G from the vectors with the components

$$g_{ij} = (v_{ij} + w_{ij}) / 2, \quad i = \overline{1, k}, \quad j = \overline{1, n}.$$

- 4.3.2. Train the network using the sample G .
- 4.3.3. $\varphi(G) \leftarrow (\varphi(V) + \varphi(W)) / 2$.
- 4.3.4. Compute the values of the criterion F for the set $\Omega = \{V, W, G\}$.
- 4.3.5. **If** G is the best sample as to the value of F , **then**
 - If** the sample V is better than the sample W as to the value of F , **then**
 - 4.3.5.1. Replace the vectors of the sample W by the vectors of the sample G .
 - 4.3.5.2. $\varphi(W) \leftarrow \varphi(G)$.
 - Otherwise**
 - 4.3.5.3. Replace the vectors of the sample V by the vectors of the sample G .
 - 4.3.5.4. $\varphi(V) \leftarrow \varphi(G)$.
- Otherwise**
 - 4.3.6. **If** the sample V is better than the sample W as to the value of F , **then**
 - 4.3.6.1. Replace the vectors of the sample G by the vectors of the sample V .
 - 4.3.6.2. $\varphi(G) \leftarrow \varphi(V)$.
 - Otherwise**

4.3.6.3. Replace the vectors of the sample G by the vectors of the sample W .

4.3.6.4. $\varphi(G) \leftarrow \varphi(W)$.

4.3.7. $\eta \leftarrow \varphi(G)$.

Otherwise

5. **If** the minimal value of the criterion F is attained for the sample A , then

$\eta \leftarrow \varphi(A)$

otherwise $\eta \leftarrow \varphi(B)$.

6. Result $\leftarrow \eta$.

End A1.

Thus, during functioning the algorithm of searching for the optimal sample, a value of the function φ is obtained according with which the attribute vector of the object being classified is constructed. The algorithm of construction of this vector on the basis of the found value η is described below.

By analogy with the construction of the mapping φ , we define a function ψ that associates some real number from the interval $[0, 1]$ with vector obtained during the functioning of the algorithm of construction of an input attribute vector.

Let R be the attribute vector that should be constructed for the object being classified, let V_1 and V_2 be vectors corresponding to numbers from the vicinity of the value η in the interval $[0, 1]$, and let r_j, v_j^1 , and v_j^2 be components of the vectors R, V_1 , and $V_2, j = \overline{1, n}$, respectively.

Algorithm A2. (Construction of the input attribute vector of an object on the basis of the number η)

INPUT: Number η .

RESULT: R is the attribute vector for the object being classified.

1. Form the vector V_1 from the values of the attributes of the object being classified by the principle of construction of vectors of the sample A .
 2. $\psi(V_1) \leftarrow 0$.
 3. Form the vector V_2 from the values of the attributes of the object being classified according to the principle of construction of vectors of the sample B .
 4. $\psi(V_2) \leftarrow 1$.
 5. **If** $\eta > 0$ and $\eta < 1$, **then**
 - 5.1. Form the vector R from the components of the vectors V_1 and V_2 according to the principle of construction of vectors of the sample C .
 - 5.2. **If** $\eta \neq 0.5$, **then**
 - 5.2.1. $t \leftarrow 2$.
 - 5.2.2. **If** $\eta > 0.5$, **then**
 - 5.2.2.1. Replace the components of the vector V_1 by the components of the vector R .
 - 5.2.2.2. $\psi(V_1) \leftarrow 0.5$.
 - Otherwise**
 - 5.2.2.3. Replace the components of the vector V_2 by the components of the vector R .
 - 5.2.2.4. $\psi(V_2) \leftarrow 0.5$.
 - 5.2.3. **Until** $t \neq \eta$
 - 5.2.3.1. $t \leftarrow (\psi(V_1) + \psi(V_2)) / 2$.
 - 5.2.3.2. Replace the components of the vector R by the averages of components of the vectors V_1 and V_2 : $r_j \leftarrow (v_j^1 + v_j^2) / 2 \forall j = \overline{1, n}$.
 - 5.2.3.3. **If** $t < \eta$, **then**
 - 5.2.3.3.1. Replace the components of the vector V_1 by the components of the vector R .
 - 5.2.3.3.2. $\psi(V_1) \leftarrow t$.
 - 5.2.3.4. **If** $t > \eta$, **then**
 - 5.2.3.4.1. Replace the components of the vector V_2 by the components of the vector R .
 - 5.2.3.4.2. $\psi(V_2) \leftarrow t$.
 6. **If** $\eta = 0$, **then**

Form the vector R from the components of the vector V_1 .
 7. **If** $\eta = 1$, **then**

Form the vector R from the components of the vector V_2 .
 8. Result \leftarrow vector R .
- End A2.

USING THE DEVELOPED AUTOMATIC CLASSIFICATION SYSTEM AND THE METHOD OF FORMATION OF INPUT VECTORS OF PARAMETERS IN PROBLEMS OF TEXT ATTRIBUTION

The developed automatic classification system can be used for the solution of problems of classification of objects specified by any collection of numerical attributes. Selecting quantitative attributes that characterize the chosen corpus of objects and using this system, one can investigate a wide class of classification problems.

A topical classification problem is the classification of texts in a natural language. There exist a series of classifications according to which texts can be sorted into different groups of attributes such as thematic, genre, stylistic, and other ones.

In this article, we illustrate the use of the developed automatic classification system for the solution of the problem of classification of texts by stylistic attributes or attribution of texts (determination of authorship of texts).

Formation of a collection of parameters. At the present time, in the problematics of texts attribution, methods are widespread that are based on the analysis of syntactic structures of a language. The selection of constructions, their arrangement, and integration into sophisticated structures form the individuality of the syntax of a literary work.

The consideration of a style as a structural-syntactic category requires the selection of specialized elements of each style at the lexical, phraseological, morphological, and syntactic levels of a language. The static picture of the organization of all linguistic means (words, wordforms, word-combinations, sentences, etc.) forms the base of each style.

Based on the aforesaid, the characteristics corresponding to the above-mentioned levels of a language and forming the system of expressive means of each author were chosen to form the input vector of parameters of a text.

1. **Frequency dictionary.** The frequency dictionary (frequency word list) S constructed from a text consists of the list of all words $s_i, i = \overline{1, n}$, that are ordered in alphabetic order (n is the dimension of the dictionary in terms of words) and are transformed into word stems. This transformation is realized by special heuristic algorithms that were developed for the lexical-morphological block of text analysis [3]. They assign to each word $s_i \in S, i = \overline{1, n}$, the frequency ch_i of its usage in the text and the part of speech P to which it belongs. The frequency vector γ is formed from the frequencies $ch_i, i = \overline{1, n}$, of the words that belong to the frequency dictionary S of the text being analyzed.

We call the dictionary formed from the entire corpus of training examples the frequency dictionary body. We call the frequency dictionary constructed for a given text without taking into account the frequency dictionary body the working frequency dictionary for this text. Next, we call the dictionary formed as a result of intersection of the frequency dictionary body and the working frequency dictionary of a text the resulting frequency dictionary for the text from which the resulting frequency vector for this example text is constructed. The resulting frequency dictionary contains all the words of the frequency dictionary body but with frequencies from the working dictionary. The frequencies of words that are absent in the working dictionary are equated to zero. The dimension of the resulting frequency dictionary is the same as that of the frequency dictionary body. The dimension of a dictionary is understood to be the number of all words in the dictionary.

2. The portion of occurrences of concrete reserved words (55 reserved words proposed by Fomenko).

This characteristic that is also relevant to the morphological level of a language is used in [6, pp. 768–820]. As investigations showed, this attribute varies from 16 to 30% on the whole and within 1–2% for a concrete author.

In [5], the results of statistical investigations of the influence of characteristics of a text on its perception are presented. According to these investigations, the perception of a text by a reader is greatly influenced by the number of different words, separator symbols, words in a sentence, participles and verbal adverbs connected with a noun (this number characterizes the complexity of a sentence), and letters in words.

Among the selected passages of texts, the greatest influence on readers is exerted (in decreasing order) by texts of Bunin, Chekhov, Tolstoi, Rowling, and Dostoevsky.

The mentioned parameters that characterize the perception of a text by a reader make it possible to determine the style of the text.

3. The portion of occurrences of internal punctuation marks (colons, commas, dashes, semicolons, brackets, inverted commas).

4. The portion of occurrences of external punctuation marks (points, exclamation marks, interrogation marks, dots, and combinations of exclamation and interrogation marks).

5. The average sentence length (the average number of words in a sentence).

6. The portion of words of a definite length (words consisting of one letter, two letters, etc. up to ten).

7. The portion of sentences of a definite length (that consist of one word, two words, etc. up to 30);

8. The average number of adjectives, participles, and verbal adverbs connected with a noun (separately).

TABLE 1

Author	Title
A. Pushkin	"The Captain's Daughter," "Queen of Spades," "Tales of the Late Ivan Petrovich Belkin," "Dubrovskii"
V. Nabokov	"The Encounter," "A Russian Beauty," "Scenes from the Life of a Double Monster," "The Vane Sisters," "The Word"
M. Gorkii	"Evening at Sukhomyatkin's," "Makar Chudra," "Old Woman Izergil'," "Fat-Faced Passions"
M. Bulgakov	"The Crimson Island," "The Extraordinary Adventures of a Doctor," "The Heart of a Dog"
I. Turgenev	"Hamlet of the Shchigrovskii District," "Asya"
L. Tolstoi	"Two Hussars," "The Raid," "The Woodfelling"
A. Kuprin	"Gambrinus," "The Blue Star," "The White Poodle," "The Garnet Bracelet"
A. Chekhov	"The Duel," "The Student," "A Chameleon," "Summer Residents"
A. Tolstoi	"Amena," "Reunion After Three Hundred Years," "The Vampire," "The Vampire Family"
F. Dostoevsky	"White Nights," "Mister Prokharchin," "Uncle's Dream," "The Hostess"

DETERMINATION OF THE RELATION BETWEEN A NOUN AND AN ADJECTIVE

Since an adjective, a participle, and a verbal adverb possess common morphological characteristics (such as gender, number, and case) in the Russian language, we consider this method for adjectives. A variant of determination of a relation between a noun and a participle or a verbal adverb is similarly considered.

To determine the latter mentioned parameter required for recognition, it is necessary to find relations between each noun and adjectives available in a sentence. It is obvious that an adjective must be in the same gender, number, and case as the noun to which it is related. Hence, it is the first condition that must be satisfied by a found relation. This stage is realized with the help of a block developed for lexical-morphological analysis that makes it possible to determine necessary characteristics (gender, number, and case) in two nodes of the sought-for relation (between the corresponding noun and adjective). Parts of speech are also determined during the construction of a frequency dictionary that, together with the number of occurrences of a word in a text, also contains its part of speech [3].

However, a situation can arise in which, according to the condition formulated above, an adjective can be simultaneously related to several nouns. Let us consider the following sentence.

"The moon charmed by the fascination of the evening was still almost full but already noticeably lost its previous golden tint and, emerging over the crowns of tall lindens, increasingly illuminated white thin clouds that sometimes cover it."

According to the first condition described above, the adjective **full** can be simultaneously related (in the original Russian text) to two nouns, namely, "moon" and "tint."

To resolve such situations, it is proposed to use the criterion of a distance between words. It is as follows: the less the distance between a noun and an adjective, the higher the probability of this relation. Such a distance is considered to be the number of the words that are between the noun and the adjective. In particular, the adjective **full** in the example (in Russian) is more close to the noun **moon** than to the noun **tint**. According to this criterion and the condition of the correspondence in gender, number, and case, the program can draw the conclusion concerning the presence of a definite relation. To establish the general parameter of relationship between nouns and adjectives for the entire text, relations are found for all sentences and they are averaged by division by the number of all such relations existing in the text.

ILLUSTRATION OF TESTING THE AUTOMATIC CLASSIFICATION SYSTEM BY AN EXAMPLE OF SOLUTION OF A PROBLEM OF TEXT ATTRIBUTION

As was revealed during testing the neural network trained by examples of texts of ten chosen writers, the network makes an unambiguous decision based on the texts known to it (that are used during training) and also provides a sufficiently exact recognition of unknown texts if they belong to one of these ten authors.

TABLE 2

Author: Product	Distance between the vector of parameters of a text and the vector corresponding to an author
V. Nabokov "Music"	~0.00030
V. Nabokov "Korolek"	~0.00033
V. Nabokov "A Bad Day"	~0.00463
M. Bulgakov "The Khan Fire"	~0.00056
L. Tolstoi "The Snowstorm"	~0.06734
A. Kuprin "Gambrinus"	~0.00447
V. Nabokov "The Encounter"	~0.00002
V. Nabokov "A Russian Beauty"	~0.00001
V. Nabokov "Scenes from the Life of a Double Monster"	~0.00001

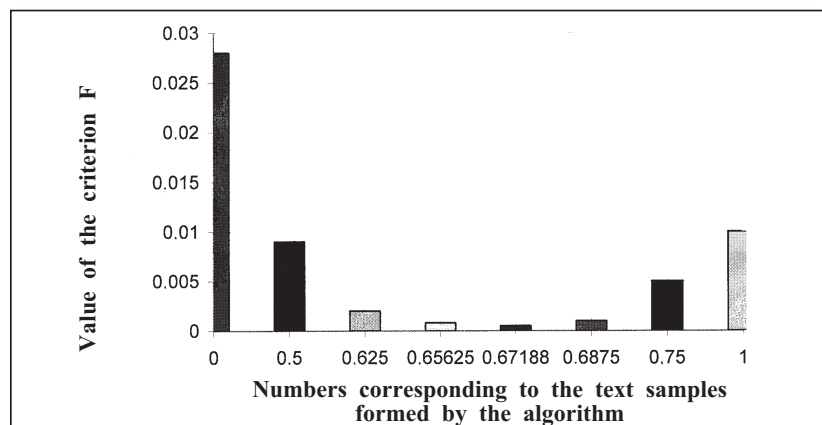


Fig. 2. Formation of an input vector.

Table 1 presents the titles of the texts that were used in training the automatic classification system. In Table 2, the results of testing the system with the help of the chosen texts are given. The writings whose titles are bold-faced are used as training examples from the training sample.

As a test example that was used by the trained network at each stage of the algorithm of formation of the input attribute vector, a passage of the writing "Music" of V. Nabokov was chosen; the attribute vector formed for this text was used by the network as a training example from the training sample. As is seen from the diagram presented in Fig. 2, the best result according to the estimation criterion is attained at the point 0.671875 of the interval [0, 1] (in our case, the prescribed value of $\varepsilon \leq 0.000$).

ATTRIBUTION OF THE WRITING "NOVEL WITH COCAINE"

As an example of application of the developed technique of attribution of anonymous texts, the problem of determination of the authorship of the writing "Novel with Cocaine" that, as well as that of the novel "The Silent Don," was intensively discussed in the literary world of the XXth century.

The novel was printed in the magazine "Chisla" in 1934 under the name of M. Ageyev. At the same year, the story "Nasty People" signed by M. Ageyev was printed in the magazine "Vstrechi." Then there was no sign of the name Ageyev. After a comparative stylistic analysis of the writing "Novel with cocaine," researchers draw the conclusion that the construction and style of the novel of M. Ageyev are very close to the class of writings of V. Nabokov.

Applying methods of the subjective-attributive approach, the researchers of the problem of attribution of this literary work selected several informal distinctive features peculiar to the Nabokov narration and his language, in particular, "fruit tints" (such as the "apricot moon") that are characteristic of Nabokov, reproduction of sounds and noises, twofold repetition for strengthening the emotional effect, a series of key epithets and verbs, coincidence in some sentences and stages and many

other stylistic characteristics of literary works peculiar to Nabokov [7]. The comparisons made allowed us to put forward the hypothesis that the author of this writing is Nabokov who wrote under the pseudonym Ageyev.

We used the developed automatic classification system for solution of this attribution problem. The frequency parameters proposed by the authors for recognition of the authorship of texts make it possible to select a series of special words, including different tints, epithets, and verbs typical of the author that, following the above list of comparisons, are distinctive attributes of writings of Nabokov. The latter attribute used for the formation of the input vector makes it possible to estimate the complexity of sentences, which also is a distinctive characteristic of the writings of V. Nabokov.

The distance between the attribute vector for the writing “Novel with Cocaine” and the output vector corresponding to the authorship of Nabokov in testing the neural network equals ≈ 0.000301 . Thus, this allows us to state with high reliability that the author of the writing “Novel with cocaine” is V. Nabokov.

REFERENCES

1. Laurence V. Fausett, *Fundamentals of Neural Network: Architectures, Algorithms, and Applications*, Prentice Hall, Upper Saddle River, New Jersey (1994).
2. S. Haykin, *Neural Networks: A Comprehensive Foundation*, MacMillan College Publishing Co., New York (1994).
3. A. V. Anisimov, A. N. Romanik, and V. Yu. Taranukha “Heuristic algorithms for determination of canonical forms and grammatical characteristics of words,” *Cybernetics and Systems Analysis*, No. 2, 60–68 (2004).
4. M. A. Marusenko, *Attribution of Anonymous and Pseudoanonymous Literary Works by Methods of Pattern Recognition Theory* [in Russian], Izd. Leningr. Univ., Leningrad (1990).
5. <http://dronnik.h10.ru/>.
6. V. P. Fomenko and T. G. Fomenko, “Author’s invariant for Russian literary texts,” in: *Methods for Quantitative Analysis of Texts of Narrative Sources*, Izd. MGU, Moscow (1996).
7. N. Struve, “Dispute about V. Nabokov and ‘Novel with cocaine’,” *Vestnik RKhD*, No. 146, 156–172 (1986).