Frames in Hilbert Spaces: A Tool for Artificial Intelligence

A.D. Linkevich

Polotsk State University
P.O. Box 105, 211440 Novopolotsk-3, BELARUS

(Received 2 October 2001)

A mathematical framework is presented for development of systems of AI intended to be endowed with ability to understand the meaning of information. The crux of the approach is representation of information by vectors of some Hilbert space treated as a semantic space and construction of frames whose vectors are associated with semantic units corresponding to particular relevant meanings. Two definitions and main properties of frames in a vector space are outlined. It is shown that frames can be used to solve such tasks of information processing as least squares approximation of experimental data, forecasting time series and probability density estimation. Frames are also applied to tackle some problems of AI pertaining to semantic retrieval and classification of (hyper) texts, medical diagnostics, control of operation of the Internet and manufacturing enterprises, etc.

Key words: artificial intelligence, information processing, frame, vector space, Hilbert space, functional analysis, data processing, approximation, forecasting, probability estimation, meaning of information, semantic retrieval, text classification, Internet search, medical diagnostics, neural field, learning, mathematical modelling, distributed complex system, Internet traffic, manufacturing enterprise, intelligent agent.

PACS numbers: 05.45.- a, 0.590. + m, 07.05 Mh, 07.05 Kf, 89.70. + c, 89.80. + h, 89.90. + n

Contents

1 Introduction 109
2 Frames in Hilbert Spaces: A Sketch of Theory 109
  2.1 Preliminary Remarks 109
  2.2 Motivation for Frames 110
  2.3 Definition and Main Properties of Frames 111
  2.4 Construction of Frames 111
3 Some Applications of Frames to Data Processing 112
  3.1 Least Squares Approximation of Experimental Data 112
  3.2 Forecasting Time Series 113
  3.3 Probability Density Estimation 113
4 Application of Frames to Problems of Artificial Intelligence 115
  4.1 Semantic Retrieval and Classification of (Hyper) Texts Using Frames 115
  4.2 Medical Diagnostics, Text Classification, and Frames 116
  4.3 Neural Fields, Dynamical Implementation of the Meaning of Information, and the Learning Problem 118
  4.4 Internet Economics, Competition Between Web Sites, and Control of the Internet Traffic Using Frames 118
  4.5 Application of Frames to Predictive Agents Controlling Manufacturing Enterprises 120
6 Conclusions 121

Appendix A. Some Examples of Frames 121
  A1 Frames in Finite-Dimensional Vector Spaces 121
  A2 Frames in Hilbert Spaces: Wavelets and Signal Processing 122
  A3 Frames in Hilbert Spaces: Coherent States of a Quantum Oscillator 123

Appendix B. Bases 124

Appendix C. Calculation of a Frame Approximately Reciprocal to a Given Frame 125

Appendix D. Knowledge Extraction from the WWW: Some Features 126

References 126
1 Introduction

It is apparently the crux of studies in the field of artificial intelligence (AI) how to develop computer systems capable to cope with the issue of semantic treatment of information (for an introduction see, e.g., [41, 63, 58]). This problem occurs of paramount importance because it is just the meaning of information that makes it useful, and actually vital, for survival and evolution of humans. Language provides a rich, elaborated system of meanings that enables people to perform communication, elaboration and implementation of plans of behavior, maintainance of self-identity and self-awareness and so on. Accordingly, it is obvious that artificial systems could appear effective being endowed with ability to understand information to be processed.

To advance in semantic treatment of information, an approach is put forward in [63, 61, 62] in which frames are defined in Hilbert spaces and used to represent semantic factors. This occurs beneficial because analysis of meanings becomes possible in terms of semantic primitives that depend on one another.

Frames appear also useful in a much more wide area of applications due to their distinguishing characteristic, namely they are composed of vectors which could be linearly dependent (though they should meet, of course, certain requirements). In particular, frames may be employed to solve such widespread tasks of information processing as approximation of experimental data, prediction of time series, estimation of probability distributions, etc. In turn, this opens the door to develop new means to handle such problems pertaining, in essence, to AI as semantic retrieval and classification of (hyper) texts, medical diagnostics, forecasting future states of a system in order to perform control of the Internet traffic and manufacturing enterprises and so on.

The above-mentioned issues determine the basic contents of the present overview which consists of two parts. The first part (Chapter 2) brings a sketch of a theory, namely it provides two definitions and a concise description of main properties of frames (some introductory and technical issues are delegated to Appendixes A, B and C). The second part is devoted to applications of frames to data processing (Chapter 3) and to the above-mentioned problems of AI (Chapter 4 and Appendix D).

The main body of the first part rests on the presentation of frames given in [63, 61], while results presented in the second one were delivered, in a preliminary form, at meetings (see [65, 66, 67, 68]). The paper contains also some new findings.

2 Frames in Hilbert Spaces: A Sketch of Theory

2.1 Preliminary Remarks

The notion of frames apparently originates from the paper [31] and lays in the foundation of the wavelet analysis which seems to be one of the most powerful and promising approaches to signal processing (for an introduction see, e.g., [22, 23, 24, 29, 45, 76, 100]). That is way we begin this chapter with a brief exposition of some issues pertaining to the wavelet analysis. It is apparently the most remarkable property of wavelets that they may be used instead of basis functions, though they generally appear not orthogonal to one another. A feature is also that all wavelets involved in analysis of a particular problem are produced starting with a unique 'mother' wavelet by translation and dilation operators.

Frames could obviously occur advantageous because to get a reliable approximation of data is only one of tasks that may arise. Instead, it can appear desirable to represent the acquired data in terms of functions chosen in advance, independent of obtained results. So, such a case naturally emerges if some models stand behind the formulas employed to analyze signals.

In this way we meet the important problem of how to expand a function of interest \( f \) that will be represented by a vector \(| f >\) of some relevant Hilbert space \( H \) over a set of given functions \( \{ h_a, a \in A \} \) that can generally be linearly dependent and, accordingly, cannot serve as a basis. Under proper conditions, such a family of functions constitutes however a frame (or a generalized frame if the label \( a \) may take on continuous values), which
2.2 Motivation for Frames

As is known, a set of vectors \( \{|h_a > \in H, \ a \in A\} \) can constitute a basis in a vector space \( H \) over a field \( K \) if they are linearly independent, i.e. their linear combination

\[
\int_A d\mu(a) \ e^a |h_a >, \ e^a \in K
\]

can occur zero only in the case when all the coefficients \( e^a \) appear zero: \( e^a = 0, \ \forall a \in A \).

(Here and below one can admit, in general, that the index \( a \) can take on values from finite, countable infinite and uncountable infinite sets so that the collection of all the feasible values appears a Borel measurable set \( A \) from the \( \sigma \)-algebra \( \sigma(A^*) \) on some appropriate space \( A^* \). Accordingly, the quantity

\[
\Phi[F,\mu] = \int_A d\mu(a) \ F(a)
\]

means the integral of a Borel measurable function \( F(a) \) defined on the set \( A \) with a Borel measure \( \mu(a) \) supported on \( A \). It is an important particular case when the measure \( \mu(a) \) is absolutely continuous with respect to a "natural" Lebesgue measure defined on the space \( A^* : d\mu(a) = \nu(a) \ da \). Here \( \nu(a) \) is the so-called Radon-Nikodym derivative (density). Then instead of \( \Phi[F,\mu] \) we meet the functional of a more special type:

\[
\phi[F,\nu] = \int_A da \ \nu(a) \ F(a).
\]

In a simple case when the variable \( a \) can take on only discrete (finite or countable infinite) values the above integral is actually nothing but a sum:

\[
\phi[F,\nu] = \sum_{a \in A} \nu(a) \ F(a).
\]

If, instead, the variable \( a \) takes on (piecewise) continuous values and there exists a usual Riemann integral then the functional \( \phi[F,\nu] \) is just this integral.

The property of linear independence is used to obtain the formula of the inverse transform of vectors. Namely, if a vector \( |u > \in H \) is expanded over a set of basis vectors \( \{|e_a > \in H, \ a \in A\} \) so as

\[
|u > = \int_A d\mu(a) \ u^a |e_a >
\]

then the coefficients \( u^a \) are uniquely expressed through this vector \( |u > \) and the vectors \( \{< l^a, a \in A\} \) of the dual basis so as \( u^a = < l^a |u > \), and they are referred to as the components (coordinates) of the vector \( |u > \) with respect to the basis \( \{|e_a >, a \in A\} \). This interrelation may be interpreted as a faithful (one-to-one) map (transform) \( g : H \rightarrow K \) so that the transition from a vector \( |u > \in H \) to the set of coefficients \( \{u^a \in K, a \in A\} \) is treated as a transform of the vectors defined by the prescription

\[
g : |u > \mapsto \{u^a, a \in A\},
\]

while construction of a vector through the set of its components is accordingly viewed as the inverse transform:

\[
g^{-1} : \{u^a, a \in A\} \mapsto |u >.
\]

It is remarkable however that the requirement of linear independence of vectors appears only sufficient, but not a necessary condition for a set of vectors to allow a similar transform. Namely, the above-mentioned condition can be weakened so that so-called frames may be used instead of bases for such transforms of vectors. In other words, the notion of frame is a generalization of that of basis, and this opens the door to introduce new kinds of vector transforms in addition to such known...
ones as constructed by Fourier, Laplace, etc. In particular, the currently popular wavelet analysis of signals just rests upon a class of frames [22, 23, 24, 29, 31, 45, 76, 100].

(Notice that the so-called coherent states of a quantum oscillator [12, 35, 49] used in quantum mechanics and quantum field theory appear a particular case of frames.

Strictly speaking, we use not proper frames, but rather generalized frames as they are described in [45], or even their slight generalizations. However, for short, all of them are referred to as frames in what follows.)

2.3 Definition and Main Properties of Frames

We accept the following definition which is close to that of [45, 63, 61].

A frame in a Hilbert space $H$ is a set of vectors $\{|h_a > \in H, \ a \in A\}$ such that there exists a set of vectors $\{|h_a > \in H, \ a \in A\}$ constituting a reciprocal (dual) frame with respect to the first one in the sense that jointly these sets provide resolution of unity in the form

$$I = \int_A d\mu(a) \ |h_a > < h^a|$$

$$= \int_A d\mu(a) \ |h^a > < h_a|.$$ 

(2.1a)

(2.1b)

Here $\mu(a)$ is a Borel measure supported on a set $A$.

The two forms of the latter decomposition (1) enable one to introduce the two kinds of the transform of vectors. Namely, if we expand a vector $|u > \in H$ over the reciprocal frame $\{|h_a >, \ a \in A\}$

$$|u > = \int_A d\mu(a) \ u_a |h_a >,$$ 

(2.2a)

then the components $u_a = < h_a|u >$ are expressed through the vectors of the original frame $\{|h_a >, \ a \in A\}$. And vice versa, if we take the representation

$$|u > = \int_A d\mu(a) \ u^a |h_a >$$

(2.2b)

then the coefficients are $u^a = < h^a|u >$. (In other terms, the resolution of unity makes it possible to obtain the reconstruction formula (inverse transform) for the expansion of vectors of the space $H$ over the frame vectors.)

These equations enable us to establish the transform

$$|u > \mapsto \{u_a, \ a \in A\}$$

(2.3)

of vectors $|u > \in H$ determined by the frame $\{|h_a >, \ a \in A\}$ and interpret the quantities $u_a = < h_a|u >$ as projections of the vector $|u >$ to the frame vectors. One can regard that the above transformation (3) performs analysis of the vector $|u >$ in terms of the frame vectors so that the inner product $u_a = < h_a|u >$ separates from the vector $|u >$ such a "detail", or a "building block", that is specified by the frame vector $|h_a >$. Namely, the quantity $|u_a|^2$ measures how much of the element described by $|h_a >$ is contained in the vector $|u >$.

The later interpretation can be clarified as follows [29, 45]. Let us normalize the vector $|u >$ so that $||u||^2 = < u|u > = 1$. Then the Cauchy-Schwartz-Bunyakovsky inequality yields

$$|u_a|^2 = |< h_a|u >|^2 \leq ||h_a||^2 \equiv M(a).$$

Furthermore, the equality is reached here if and only if the vector $|u >$ is a multiple of the frame vector $|h_a >$. In other words, the quantity $|u_a|^2$ is dominated (majorized) by the function $M(a)$ which is uniquely determined by the frame. For any given fixed value $a_0$, the quantity $|u_{a_0}|^2$ attains its maximum value $M(a_0)$ if and only if the vector $|u >$ is a multiple of $|h_{a_0} >$ (and accordingly $|u_{a_0}|^2 < M(a)$ for any $a \neq a_0$, if we exclude the trivial case when some $|h_a >$ is a multiple of the $|h_{a_0} >$).

Similarly, synthesis, or reconstruction, of vectors represents the inverse map

$$\{u_a, \ a \in A\} \mapsto |u >.$$ 

(2.4)

2.4 Construction of Frames

It is the question however how to construct such sets of vectors $\{|h_a >, \ a \in A\}$ and $\{|h^a >, \ a \in A\}$ that constitute frames in Hilbert space $H$ reciprocal (dual) to each other (in the sense that they jointly
ensure the resolution of unity (1)). Suppose that there exists an operator $T$ such that

$$|h^a > = T|h_a >, \quad \forall a \in A. \quad (2.5)$$

Then the above problem amounts to finding this operator given a set of vectors $\{h^a, \quad a \in A\}$.

Let us introduce the so-called metric operator

$$G = \int_A d\mu(a) |h_a > < h_a|, \quad (2.6)$$

Then the required resolution of unity (1) is achieved if the operator $T$ satisfies the equation

$$TG = I, \quad (2.7)$$

or symbolically $T = G^{-1}$. Thus the problem is reduced to finding the inverse operator $G^{-1}$ (provided that the latter exists, otherwise a kind of generalization or approximation, such as pseudo-inverse matrix, may be employed).

Here we can introduce a new, more concise, definition of frames as follows.

A frame in a Hilbert space $H$ is a set of vectors $\{h_a \in H, \quad a \in A\}$ such that the metric operator $G$ (6) is invertible, i.e. there exists an operator $T$ such that $TG = I$.

Let us consider in more detail the prevailing case when the label variable $a$ can take on only discrete values from some set $A$ and vectors $|h_a >$ are actually determined by their coordinates $h_{ai} = < e_i|h_a >$ with respect to a countable orthonormal basis composed of vectors $\{|e_i >, \quad i = 1, 2, \ldots\}$ that satisfy the condition

$$< e_i|e_j > = \delta_{ij}, \quad i, j = 1, 2, \ldots$$

Then eqs. (5)-(7) yield the following expression for components $h_i^a = < e_i|h^a >$ of the reciprocal frame

$$h_i^a = \sum_j T_{ij} h_{aj}, \quad (2.8)$$

where the matrix elements $T_{ij} = < e_i|T|e_j >$ obey the equation

$$\sum_j T_{ij} G_{jk} = \delta_{ik} \quad (2.9)$$

with

$$G_{jk} = \sum_{a \in A} h_{aj} \bar{h}_{ak}. \quad (2.10)$$

Thus one can sum up the above consideration as follows. Given components $\{h_{ai}, \quad a \in A, \quad i = 1, 2, \ldots\}$ of vectors $\{|h_a > \in H, \quad a \in A\}$, it is required to calculate components $\{h_i^a, \quad a \in A, \quad i = 1, 2, \ldots\}$ of vectors $\{|h^a > \in H, \quad a \in A\}$ so as to provide the resolution of unity

$$\delta_{ij} = \sum_{a \in A} h_i^a \bar{h}_{aj} \quad (2.11a)$$

$$= \sum_{a \in A} h_{ai} \bar{\delta}_{j}^a. \quad (2.11b)$$

An algorithm of direct computing these $h_i^a$ consists of the following steps:

1) compute matrix elements of the metric operator $G_{jk}$ using eq. (10);
2) solve eq. (9) so as to find matrix elements $T_{ij}$ of the operator $T$;
3) find components $h_i^a$ using eq. (8).

3 Some Applications of Frames to Data Processing

3.1 Least Squares Approximation of Experimental Data

In this section we are going to demonstrate how the technique presented in Chapter 2 can be used to provide a description of experimental data in terms of functions that are taken independent of results of an experiment and may be treated as elements of a frame. As is mentioned in Section 2.1 such a task arises, e.g., if some model is intended to be applied to the subject under investigation. Namely, the question to be answered is how a variable $y$ depends on another variable $x$.

Let $(x_1, y_1), \ldots, (x_N, y_N)$ be experimental data to be processed. It is required to find such a function $y(x)$ that yields the best approximation of the data in the sense that the functional

$$E = \frac{1}{2} \sum_{i=1}^N \left( \frac{y_i - y(x_i)}{\sigma_i} \right)^2 \quad (3.1)$$

takes the minimum [44, 55, 92]. Here \( \sigma_i \) is the standard (mean square) deviation corresponding to \( y_i \).

Let us expand the function \( y(x) \) over a frame \( \{ h_a(x), \ a \in A \} : 
\[
y(x) = \int_A d\mu(a) \ y^a h_a(x).
\] (3.2)

Then the so-called normal equations \( \partial E/\partial y_a = 0 \) yield the coefficients of the above expansion:
\[
y^a = < h^a | y / \sigma_i^2 > = \sum_{i=1}^{N} \tilde{h}^a(x_i) \ y_i / \sigma_i^2.
\] (3.3)

### 3.2 Forecasting Time Series

In this section we are going to tackle the problem of prediction of time series (for an introduction see, e.g., [18, 2]). We suppose that a signal \( x(t) \) is sampled at discrete times \( t_0, t_1, \ldots \), which yields a time series to be analysed. In practice, only a scalar variable \( x(t) \) is often measured with a uniform sampling so that \( t_k = k\tau, \ k = 0, 1, \ldots \). Thus the dynamics is described by a map \( G(\cdot) \) such that \( x(t + \tau) = G(x(t)) \), or \( x_{k+1} = G(x_k) \) where \( x_k \equiv x(k\tau) \equiv x(t_k) \).

As \( x(t) \) is a scalar whereas the state of the underlying system is determined by a \( D \)-dimensional vector \( u(t) \), the proper degrees of freedom should be taken into account. Familiar time derivatives of \( x(t) \) are obviously inaccurate because the data on \( x(t) \) are acquired only at discrete times. Instead, one can use a \( d \)-dimensional vector in the embedding space: \( y_i \equiv y(i) \equiv y(t_i) = (x(i), x(i-1), \ldots, x(i-d+1)) \). The condition \( d < 2D + 1 \) ensures a faithful representation of attractors of the dynamics [72, 91]. Then it is required to find such a map \( F(\cdot) \) that describes the time evolution of the vector \( y_i \).

Thus we meet the following mathematical problem. Let \( y \in \mathbb{R}^d \) be a \( d \)-dimensional vector composed of time delays of a variable \( x \). We are interested in a map \( F(\cdot) \) which is defined on some appropriate Hilbert space \( H \) and describes time evolution of the vector \( y \):
\[
y_{i+1} = F(y_i), \ i = 0, 1, \ldots
\] (3.4)

Let us expand the function \( F(y) \) over a frame \( \{ h_a(y), a \in A \} \) so as
\[
F(y) = \int_A d\mu(a) \ F^a h_a(y).
\] (3.5)

Using the resolution of unity (2.1) we find
\[
F^a = \langle h^a | F \rangle = \sum_{j=0}^{N} \tilde{h}^a(y_j) \ y_{j+1} / \sigma_j^2
\] (3.6)

where \( N_i \leq i - 1 \). Substitution of (6) into (5) yields
\[
y_{i+1} = \sum_{j=1}^{N_i} \Delta(y_i, y_j) \ y_{j+1} / \sigma_j^2
\] (3.7)

where
\[
\Delta(y, y') = \int_A d\mu(a) \ h_a(y) \ \tilde{h}^a(y').
\] (3.8)

The latter two expressions constitute a straightforward generalization of the corresponding equations considered in [63, 61] for the case when a function is expanded over a basis composed of orthonormal functions.

A similar form for the map \( F(y) \) with a specific function taken as \( \Delta(y, y') \) is used in [1] because this is suggested by the kernel density estimation approach [86, 2]. Let us notice also that our framework leads to an equation for the probability density that is similar to the form accepted in the kernel estimation method as is shown in the next section.

### 3.3 Probability Density Estimation

Let \( f \) be a function represented by vector \( |f > \in H \) of some Hilbert space \( H \). Suppose that the following properties of the space \( H \) take place:

(i) there exists a countable infinite basis in \( H \) composed of vectors \( \{|x >, \ x \in X \} \) that provides the resolution of unity of the form
\[
I = \int_X d\Omega(x) \ |x > < x |
\] (3.9)

where \( \Omega(x) \) is a measure supported on a space \( X \);
(ii) one can construct a frame and reciprocal (dual) frame composed of vectors \(|h_a|, a \in A\) and \(|h^a|, a \in A\) respectively such that the resolution of unity (2.1) holds as well.

Then using eq.(9) we can expand vectors \(|h_a|\) and \(|h^a|\) over the basis

\[
|h_a| = \int_X d\Omega(x)\ h_a(x)|x| >,
\]

\[
|h^a| = \int_X d\Omega(x)\ h^a(x)|x| >,
\]

and employ these representations to obtain the equation

\[
\int_X d\Omega(x')\ \Delta(x, x') f(x') = f(x)
\]

(3.10)

that occurs for any function \(f \in H\). (Here \(\Delta(x, x')\) is defined by eq.(8).)

Generally, the Hilbert space \(H\) is infinite-dimensional, while in practice one has to restrict one self to a finite-dimensional approximation of the space \(H\). Assume also that vectors of the frame and reciprocal frame can be approximated by \(M\)-dimensional vectors such that the resolution of unity (2.1) is replaced by the following approximate expression

\[
I \approx I_K = \sum_{k=1}^{K} |h_k| < h_k|\quad (3.11a)
\]

\[
= \sum_{k=1}^{K} |h^k| < h^k|.
\]

(3.11b)

(Obviously that the relation \(M \leq K\) must take place.) Then eq.(10) turns to the equation

\[
\int_X d\Omega(x')\ \tilde{\Delta}_{K}(x, x') f(x') = f(x)
\]

(3.12)

with

\[
\Delta_{K}(x, x') = \sum_{k=1}^{K} h_k(x)\ h^k(x')
\]

which is satisfied for any function \(f \in H_K\) where \(K \leq \tilde{K}\).

Passing to the limit \(\tilde{K} \to \infty\), we observe that the relation (12) holds for any value \(K\) and consequently for any function \(f \in H\). Thus we conclude that \(\Delta_{K}(x, x') \to \delta(x - x')\) as \(K \to \infty\) where \(\delta(\cdot)\) is the Dirac delta function. Accordingly, \(\Delta_{\tilde{K}}(x, x')\) can be viewed as an incomplete delta function (see, e.g., [6, 34]).

The property (12) suggests introducing the following approximation of the probability density:

\[
\rho_{\tilde{K}}(x) = \frac{1}{N} \sum_{i=1}^{N} \Delta_{\tilde{K}}(x, x_i).
\]

(3.14)

Then passage to the limit \(\tilde{K} \to \infty\) yields \(\rho_{\tilde{K}}(x) \to \rho(x)\) where

\[
\rho(x) = \frac{1}{N} \sum_{i=1}^{N} \delta(x - x_i)
\]

(3.15)

is the so-called natural (or box-counting) invariant probability measure defined on the phase space [32]. Equation (14) looks like a kind of the kernel density estimation [86, 2]

\[
\rho(x) = \sum_{i=1}^{N} K(||x - x_i||)
\]

(3.16)

which enables one to get a smooth probability distribution from a set of discrete data points. To this end, each point \(x_i\) is provided with the contribution \(K(||x - x_i||)\) taken to be, as a rule, a non-increasing function of the distance.

It seems promising however to take not accidental functions as kernels but incomplete delta functions instead. This is essential to obtain a consistent estimate of the probability distribution (see, e.g., [82]). Namely, the estimate

\[
\rho(x) = \frac{1}{N} \sum_{i=1}^{N} K_N(x - x_i)
\]

of the probability density \(\rho(x)\) is consistent if at any point of continuity of \(\rho(x)\)

\[
\int K_N(x - x')\ \rho(x')\ dx' \to \rho(x)\quad \text{as}\quad N \to \infty
\]

and

\[
\frac{1}{N} \sup K_N(x) \to 0\quad \text{as}\quad N \to \infty.
\]
Thus the function \( K_N(x) \) should approach the delta function \( \delta(x) \) as \( N \to \infty \), but the \( \delta(x) \) must not emerge, which is ensured if the convergence \( \Delta_N(x) \to \delta(x) \) is slow enough so that \( K_N(x)/N \to 0 \) as \( N \to \infty \) at any point \( x \).

4 Application of Frames to Problems of Artificial Intelligence

4.1 Semantic Retrieval and Classification of (Hyper) Texts Using Frames

4.1.1 Classification of (Hyper) Texts

The problem of classification of documents (texts, hypertexts, pieces of data, etc.) can be posed as follows (for an introduction into the subject see, e.g., [30, 77]): given a set of classes \( C = \{c_1, \ldots, c_K\} \) and a document \( d \), it is required to assign this document to such a class \( c^* \in C \) that is most probable. To this end, some features are used as the ground of classification. In principle, one can admit any function of the document and the class, \( f_i(d, c) \), to be a feature.

It is reasonable actually to deal with a slightly more general problem: given a set of classes \( C = \{c_1, \ldots, c_K\} \) and a document \( d \), it is required to estimate the conditional probability \( P(c|d) \) that this document \( d \) belongs to the class \( c \in C \). Then the decision will naturally be to assign the document \( d \) to such a class \( c^* \in C \) that has the maximal probability \( P(c^*|d) \), i.e. the value \( c^* \) is determined by the condition \( P(c^*|d) = \max_{c \in C} \{P(c|d)\} \).

Two regimes of operation of a classifier are commonly distinguished: (1) the learning (training) of the classifier when its parameters are adjusted on the basis of some training data, (2) the proper classification of documents when the classifier is provided, as its input, with new documents that were not used for the learning and must be classified. Training data comprise a set of labeled documents \( D = \{d_1, d_2, \ldots\} \), i.e. during the learning one deals with a set of pairs \( (d_1, c_1), (d_2, c_2), \ldots \) where \( d_i \) is a number (or title, label, name, etc.) of the document and \( c_i \) is a label (number, etc.) of the class to which the document \( d_i \) belongs.

The classification of (hyper) texts mainly rests on the so-called bag-of-words or unigram representation of documents when each word in the document is regarded as a separate feature and the classifier relies on statistics about single words [77]. Most of learning algorithms are based on the multi-variate Bernoulli model which uses a Boolean variable to encode the presence of a particular word in the document. The multinomial model deals with integer word counts so that words are treated as "events" and a document is viewed as a collection of such events. The second method outperforms the first one on several data sets Cravenetal1998.

There are a number of supervised learning algorithms used for text classification, namely naive Bayes, k-nearest neighbor, vector support machines, boosting, rule learning algorithms, maximum entropy and so on. (For an introduction into a general framework of learning algorithms for classification see, e.g., [30, 77].) However, not a single technique consistently outperforms the others throughout the variety of particular domains.

The aim of this section is to show that semantic retrieval and classification of texts can be treated with the help of frames. The general idea of analysis of the meaning of information in terms of some semantic factors represented in a Hilbert space by vectors of a frame has been put forward in [56] and developed in [58, 62, 65].

In our approach any document \( d \) is represented by a vector \( |d \rangle \in H \) of some Hilbert space \( H \). To specify such vectors, various ways, may, in principle, be used. So, we can define \( |d \rangle \) as follows:

\[
|d \rangle = (p(\tilde{w}_1|d), \ldots, p(\tilde{w}_M|d))^T.
\] (4.1)

Here \( p(\tilde{w}_i|d) \) is the probability for "word" \( \tilde{w}_i \) to appear in the document \( d \), and \( \tilde{w}_1, \ldots, \tilde{w}_M \) are words from a special vocabulary \( \tilde{W} \). (More precisely, \( p(\tilde{w}_i|d) \) is not a probability, but its estimate determined by the average frequency of appearance of the word in the document. Generally, by word \( \tilde{w}_i \) one can imply a proper word of any language and any word group that has a certain particular meaning.)

Each class \( c_k \) is similarly associated with the vector \( |c_k \rangle \). If we choose the representation (1) then

\[
|c_k \rangle = (p(\tilde{w}_1|c_k), \ldots, p(\tilde{w}_M|c_k))^T.
\] (4.2)
where \( p(\tilde{w}_i|c_k) \) is the probability for "word" \( \tilde{w}_i \) to appear in a "typical" document that belongs to the class \( c_k \).

Generally, the vectors \( \{|c_k > \in H, \ k = 1, \ldots, K\} \) may occur linearly dependent and accordingly can not be taken as basis vectors of the space. However, under proper conditions we can treat them as elements of a frame and construct the reciprocal frame \( \{|c^k > \in H, \ k = 1, \ldots, K\} \), which allows analysis and synthesis of documents.

Namely, one can expand vector \( |d > \in H \) over the reciprocal frame so as

\[
|d > = \sum_{k=1}^{K} s_k(d) |c^k > .
\]

Here the coefficient \( s_k(d) = < c_k|d > \) measures how much the class \( c_k \) is represented in the document \( d \). Being appropriately normalized, these coefficients admit the probabilistic interpretation. Namely, the quantity \( P(c_k|d) = \frac{1}{Z(d)} |s_k(d)|^2 \) yields the probability that the document \( d \) belongs to the class \( c_k \) provided that the normalization factor

\[
Z(d) = \sum_{k=1}^{K} |s_k(d)|^2 .
\]

In this way we have a solution of the problem of text classification.

### 4.1.2 Semantic Retrieval of (Hyper) Texts

For semantic retrieval of information pertaining to a given specific subject, we employ a set of some semantic units \( \{|\sigma_a > \in H, \ a \in A\} \) that represent all relevant particular meanings of information in all feasible forms and manners. For humans, any semantic unit can be explained, e.g., with the help of definitions and example sentences. For computer systems, special ontologies are produced (for an introduction see, e.g., [69] and references therein).

Each semantic unit \( \sigma_a \) is represented by a vector \( |\sigma_a > \in H \) of a Hilbert space \( H \) called the semantic space. This can be achieved in a way similar to the treatment of the issue of text classification given in the previous subsection. Namely, we can compose vector \( |\sigma_a > \) as follows:

\[
|\sigma_a > = (p(\tilde{w}_1|a), \ldots, p(\tilde{w}_M|a))^T .
\]

Here \( p(\tilde{w}_i|a) \) is the probability for word \( \tilde{w}_i \) to occur in a "typical" text corresponding to the unit \( \sigma_a \), and \( \tilde{w}_1, \ldots, \tilde{w}_M \) are words from a special vocabulary \( \tilde{W} \).

Any document \( d \) is represented in the same way by a vector \( |d > \in H \).

The vectors \( \{|\sigma_a > \in H, \ a \in A\} \) can be linearly dependent, which prevents using them as basis vectors of the space, while under proper conditions they constitute a frame. This enables us to accomplish analysis and synthesis of meanings. So, we can represent vector \( |d > \in H \) in the form

\[
|d > = \int_A d\mu(a) |\sigma_a > ,
\]

and then the coefficient \( s_a(d) \) shows how much of the semantic unit \( \sigma_a \) is contained in the document \( d \). Therefore this quantity can be referred to as the (semantic) capacity of the (semantic) unit \( \sigma_a \) in the document \( d \).

The task of a particular semantic retrieval of information is specified then by indicating desired values for the semantic capacities as follows: find the set \( D_a = \{vd \in D : s_a(d) \in S_a, \forall a \in A\} \) of all documents \( d \) from a repository \( D \) such that their semantic capacities \( s_a(d) \) take values from given prescribed sets \( S_a \). It is worth noting that employment of frames enables us to pose the problem of semantic retrieval of documents in a mathematically rigorous form in contrast to prevailing approaches.

### 4.2 Medical Diagnosing, Text Classification, and Frames

The human body exhibits a reach variety of dynamical phenomena (for an introduction see, e.g., [36, 81]) and making a medical diagnosis appears a notoriously difficult problem. Therefore it seems reasonable to look for new approaches to the problem of medical diagnostics in addition to traditional ones (see, e.g., [38, 70, 94] for an introduction into making diagnoses of diseases of the cardiovascular system). So, artificial neural networks [10, 13, 14, 42, 96] and methods of nonlinear dynamics [11, 15, 28, 33, 52, 53, 54, 83, 90, 89, 103, 104, 105] turn out to be useful to increase diagnostic efficacy.

In essence, diagnostics is nothing but the identification of a kind of illness, which implies that classification can appear of paramount importance.
for medical diagnostics. This makes it natural to try methods and algorithms of classification and pattern recognition (for an introduction see, e.g., [30, 77]).

Let us consider an object (a patient in our case) whose state at time moment \( t \) can be specified by a set of variables \( u_\kappa(t) \) where the index \( \kappa \) may, generally speaking, take on discrete and/or continuous values. We will investigate in what follows only finite-dimensional systems whose states are completely determined by a finite number of degrees of freedom. (Actually, for the human body, this is but an approximation because a number of continuous media (fields) could turn out to be relevant and should be taken into account. Now we have to suppose yet that all of them can be discretized and thereby incorporated in our setting. Notice nevertheless that the formalism under consideration might be generalized to the case of infinite-dimensional systems (fields) as well.) Dynamics of the system, i.e. time evolution of the state vector \( u(t) = (u_1(t), \ldots, u_D(t)) \) can be described by a map \( g \) which takes an initial state \( u(t_0) \) to a state \( u(t) \).

We assume that the patient is examined with the help of some measuring instruments which yields \( m \) observables (signals) \( x^1(t), \ldots, x^m(t) \) (these can be results of analysis of blood, pressure, breathing, electrocardiograms, electroencephalograms, electrooculograms, electroretinograms, etc.). Thus we deal with \( m \)-dimensional signal vector \( x(t) = (x^1(t), \ldots, x^m(t)) \), and one can regard that this vector belongs to a set \( X \) from the \( m \)-dimensional Euclidean space \( \mathbb{R}^m \), i.e. \( x(t) \in X \subseteq \mathbb{R}^m \).

As usual, the observable signal \( x(t) \) is supposed to be related with the state of the underlying dynamical system through some measurement function contaminated by additive noise. Moreover, a number of other sources of stochasticity should be, strictly speaking, taken into account.

Usually, it is supposed that the signal \( x(t) \) is sampled at discrete times \( t_0, t_1, \ldots, \) which yields a time series to be analysed. In practice, only a scalar variable \( x(t) \) is often measured with a uniform sampling so that \( t_k = k\tau, \ k = 0, 1, \ldots \). Then to take the proper degrees of freedom into account, vectors in the embedding space can be used to provide a faithful representation of attractors of the underlying dynamics (see, e.g., [2, 18, 32, 72, 91] and references therein).

It is worth noting that the above assumption that time is discrete is neither necessary for a theoretical description of the human body nor quite sufficient to provide a complete coherent picture of the underlying dynamics. Also, experimental studies can, generally speaking, give us analog signals, and it is digital computers and digital measuring equipment that make reasonable to treat time as a discrete variable. Moreover, the case of continuous time appears more tractable theoretically (or, at least, more convenient), and we will use this opportunity, though our approach can be extended to the case of discrete time using an approach such as developed in [104].

Suppose that a medical examination of a patient provides some \( m \)-dimensional signal \( x(t) = (x^1(t), \ldots, x^m(t)) \in \mathbb{R}^m \) sampled at discrete times \( t_1, t_2, \ldots, \), which yields a time series \( x_1 = (x^1_1, \ldots, x^m_1), (x^1_2, \ldots, x^m_2), \ldots \) to be handled. Let us introduce the partition of the phase space by cells \( \Gamma_e \) with sizes \( \Delta_e^1, \ldots, \Delta_e^m \) along the axes \( x^1, \ldots, x^m \) respectively. All such cells can be enumerated using, e.g., some modification of the known Cantor diagonal procedure.

In general, possible values of the index \( e \) can constitute a countable infinite set of numbers. However, it is reasonable to suppose that the human body is a dissipative system and any trajectory is confined within a finite region. Then, as a result, we have that the index \( e \) takes on values from the finite set \( \{1, \ldots, M\} \).

Any cell \( \Gamma_e \) (or a group of cells) can be treated as a word encoded by the coordinates of its center (or its certain corner). Then time evolution of the system during a time interval generates a sequence of words that can be referred to as a text, or a document, associated with the given sample of the dynamics.

Thus the problem of making a medical diagnosis of a patient amounts to the issue of classification of dynamical regimes of the body and is reduced to the task of classification of documents, which allows application of methods of text classification.

In terms relevant to the issue of making a medical

\[ \text{Nonlinear Phenomena in Complex Systems Vol. 5, No. 2, 2002} \]
diagnosis, the problem of text classification formulated in the previous section amounts to the following task: given a set of kinds of illness \( C = \{c_1, \ldots, c_K\} \) and a record \( d \) of measurements produced by the body of a patient (a set of symptoms expressed with help of some special "alphabet"), it is necessary to find the probability \( P(c|d) \) that this patient suffers from the disease \( c \in C \). (The case of good health can easily be included into the set \( C \) as an element, e.g., \( c_1 \).)

Thus the problem of the diagnostics is reduced to a task of text classification and frames can be used to solve it.

4.3 Neural Fields, Dynamical Implementation of the Meaning of Information, and Employment of Frames to Solve the Learning Problems

Neural networks prove to be powerful tools of information processing. It is significant that a neural network has a finite number of degrees of freedom. In contrast, the state of a neural field (NF) is described by a vector of some infinite-dimensional topological space such as Hilbert one [64]. Observe that the meaning of any thought is generally inexessible completely through any finite number of words (or signs of another kind) and appear infinite-dimensional. It struck therefore that the NF perfectly suits to implement meanings of information.

We suppose that (1) a NF can occur in states specified by settled neural field functions \( \psi_k, \ k \in I_\psi \), (2) there exists an operator \( \Lambda : \psi_k \rightarrow \Gamma_k \) that transforms any of such \( \psi_k \) into a Borel measure \( \Gamma_k \) supported on \( X \), (3) each measure \( \Gamma_k(x), \ x \in X \), is absolutely continuous with respect to a "natural" Lebesgue measure defined on the same space \( X \), i.e. \( d\Gamma_k(x) = \rho_k(x)dx \) where \( \rho_k(x) \) is the Radon-Nikodym derivative (density), (4) to read out information one uses a family of measurable functions \( f^* = \{f^\nu(x), \ x \in X, \ \nu \in I_f\} \).

Then the main assumption is that the NF contains pieces of information whose codes are \( u^\nu_k = <f^\nu|\rho_k>, \ \nu \in I_f, \ k \in I_\psi \). Here \( <f|g> = \int_X d\Omega(x) f(x) g(x) \) denotes the inner product of functions \( f, g \in H \). The learning problem is posed as the inverse task: given a set of codes \( u^\nu_k, \ \nu \in I_f, \ k \in I_\psi \), find measure densities \( \rho_k \) and probing functions \( f^\nu \) such that \( <f^\nu|\rho_k> = u^\nu_k \) for all \( \nu \) and \( k \).

To solve the learning problem let us expand the measure densities \( \rho_k(x) \) and probing functions \( f^\nu(x) \):

\[
\rho_k(x) = \int_A d\mu(a) \rho_{ka} h^a(x),
\]

\[
f^\nu(x) = \int_A d\mu(a) f^{\nu a} h_a(x)
\]

where \( \rho_{ka} = <h_a|\rho_k>, \ f^{\nu a} = <f^\nu|h^a> \).

Substitution of these expansions into the equation \( <f^\nu|\rho_k> = u^\nu_k \) yields

\[
\int_A d\mu(a) f^{\nu a} \rho_{ka} = u^\nu_k.
\]

It is customary when the domains of indexes \( I_f \) and \( I_\psi \) are finite sets that may be written so as \( I_f = \{1, \ldots, N\}, \ I_\psi = \{1, \ldots, K\} \), and the index variable \( a \) can take on discrete values \( \{1, \ldots, L\} \). Then the latter equations reduces to the algebraic system

\[
\sum_{a=1}^L f^{\nu a} \rho_{ka} = u^\nu_k, \ \nu = 1, \ldots, N, \ k = 1, \ldots, K,
\]

considered in [63, 61, 64, 60].

4.4 Internet Economics, Competition Between Web Sites, and Control of the Internet Traffic Using Frames

The Internet radically changes ways of representation, retrieval, access and treatment of information. However, it is a problem how to ensure that any user will obtain any document from the WWW as quickly as possible. This task is customarily referred to as control (or routing) of the Internet traffic.

Nowadays, the main communication mechanism of the Internet is the hyper-text transfer protocol (HTTP) (though other protocols can be used instead such as FTP, Gopher, etc.). HTTP is a simple
client-server interconnection that provides exchange of messages: (1) a user (client) with the help of a browser connects with a server (provider), (2) the browser sends to the server a request to retrieve a certain document, (3) the server sends the requested document, (4) the server closes the connection. (A browser and a server may also interact through an intermediate machine called a proxy.)

Current models (see, e.g., [75, 99]) deal with highly simplified networks. In particular, (1) traffic originates only at routers of a certain kind, (2) ultimate destinations are also specialized routers, (3) bottlenecks can appear only at yet another set of routers. A number of other features of these models seem to be not realistic enough, and we attempted to overcome some their limitations and shortcomings. Below we are going to sketch our model and show that a problem of forecasting time series appears rather naturally in attempts to increase effectiveness of performance of the Internet traffic mechanisms.

A problem of prediction of future events arises also in the Internet economics. As is stated in recent papers (see, e.g., [75] and references therein), the traditional theory of competitive equilibrium can not be applied to the Internet economics because of a "frictionless", or "massless", nature of electronic markets when (1) the price of a Web page is essentially zero, (2) supply will always match demand, i.e. demand can be instantly satisfied by supply at a negligible cost to the supplier, (3) competition between Web sites is mainly based not on the cost, but rather on the advertising and differentiation in the provided services, (4) the only relevant variable quantity is the aggregate demand, i.e. the number of customers willing to visit a site or download information or software. As a result, there emerge new kinds of behavior, strategies, and phenomena.

To explore effects of competition among Web sites, a dynamical model has been suggested and investigated in [75]. The most striking property observed in this model is apparently the following: as the competition level between sites increases there happens a sudden, sharp transition from the fair market share (when many sites thrive simultaneously), to a "winner-take-all" market in which one or a few sites grab almost all the users, whereas the other sites get no market share. Such a phenomenon is indeed empirically observed in electronic markets. Also, a very small change in the parameter values can crucially affect the stability of the equilibrium. All this entails tremendous difficulties to predict which equilibrium the system will ultimately converge to.

In contrast to prevalent models such as [73, 99], we put forward a unified, slightly generalized framework specified as follows. A network constitutes a directed graph composed of \( N \) uniformized, functionally single-type nodes. Each node can act as (1) a source of packets to be delivered (a starting point of traffic), i.e. a server (provider) which sends a requested document, (2) a receiver of packets (an endpoint of traffic), i.e. a browser which has sent a request to retrieve the document, (3) a proper router (an intermediate fork point), i.e. a transfer node that has more than one outgoing channels).

Thus we suppose that a generalized node \( i \) at any time moment \( t \) \( (t = 0, 1, \ldots ) \) has \( m + 1 \) inputs \( x_{i0}^t, x_{i1}^t, \ldots , x_{im}^t \) and \( n + 1 \) outputs \( y_{i0}^t, y_{i1}^t, \ldots , y_{in}^t \) where \( x_{ij}^t \) is a packet received at time \( t \) by node \( i \) from input channel \( j \), while \( y_{ik}^t \) is a packet sent at time \( t \) by node \( i \) through output channel \( k \). Observe that input \( x_{i0}^t \) is intended to put packets in the network (if this node can function as a provider), and output \( y_{i0}^t \) takes a packet if this node is the target of the packet. Each packet contains the address of the destination and data to be delivered.

Any node operates at each time moment as follows. (1) The node takes one of packet to be arrived through input channels of the node. (2) The received packet is placed at a storage. (3) The node picks one packet out from its storage, selects an outgoing line, and sends the packet through the chosen output channel. (4) The node eliminates from its storage any packet successfully delivered if an appropriate acknowledgement has been received.

A convenient quantity is the probability \( P_{ij}(l,T) \) that a packet of maximal length \( l \) will be successfully delivered from node \( i \) to node \( j \) within maximal time \( T \) provided that the channel between these nodes is kept open during the time \( T \) (nodes \( i \) and \( j \) are implied to be adjacent). (Notice that the customarily
defined bandwidth of a channel between nodes $i$ and $j$ can be determined as the maximal value of $l$ for which $P_{ij}(l, T = 1) \neq 0$.) Also, it is advantageous to introduce the binary variable $s_{ij}^t$ to indicate the decision made by node $i$ at time moment $t$ either to switch on ($s_{ij}^t = +1$) or to switch off ($s_{ij}^t = -1$) the channel from node $i$ to node $j$. Thus the architecture and operational characteristics of a network are described by such probabilities $P_{ij}(l, T)$, while the state of the network at a given time moment $t$ is determined by the "switching" variables $s_{ij}^t$ at previous time moments $t' < t$. It is worth noting also the non-Markovian character of the network.

Let us suppose that control of operation of a node is performed by an intelligent agent endowed with abilities to gather, collect, and process relevant information so as to elaborate a decision for the node. Then a task of any such agent is to predict values of all the switching variables $s_{ij}^t$ for future time moments given values of these quantities in the past. Really, not all of these variables are relevant for operation of a given node. It is reasonable therefore to determine, for any node $i$, vector $\sigma_i = (\sigma_{i1}\beta_1, \sigma_{i2}\beta_2, \ldots, \sigma_{iL_i}\beta_{L_i})$ whose components constitute an ordered subset of all quantities $s_{ij}$ arranged according to attainability and importance for the agent (node) $i$ to make and exploit forecasting. Then the task to be achieved by agent $i$ reduces to the following: given values $\sigma_i^{t'}$ for $t' = t, t - 1, \ldots, t - T_p$, find values $\sigma_i^{t''}$ for $t'' = t + 1, \ldots, t + T_f$.

It seems promising to treat any $\sigma_i$ as an integer number, or even as an approximation of a rational number. In the latter case, one natural way to improve (if necessary) predictions is to enhance the accuracy of the approximation of these rational numbers by taking into account the states of more nodes.

Thus the problem of control of the Internet traffic includes the task of forecasting time series and frames could be useful for this purpose.

### 4.5 Application of Frames to Predictive Agents Controlling Intelligent Manufacturing Enterprises

A number of economical, social, biological, and other dynamical systems constitute ensembles of interacting agents that exhibit a rich variety of collective phenomena including deterministic chaos, fractals, self-organization, etc. Agents can have different beliefs, expectations, desires, intentions, strategies, and objectives. (The term *agent* is also used in another sense in computer science - see [69].)

An important particular class is distributed dynamical systems that have little centralized control and communication between their agents, but still should behave in a prescribed manner. The seminal paper [8] has introduced a model of economic agents and posed the question when they can achieve a prescribed global goal avoiding frustration. Treatment of actions in terms of the minority game [19, 21, 17, 106] provides a fruitful approach to model market mechanisms and investigate self-organization and other collective phenomena in various systems.

The main aim of this section is to demonstrate that frames defined in Hilbert spaces might appear a promising means to forecast time series that should be handled to achieve efficient control of modern manufacturing enterprises (MEs). To this end, we have put forward in [67] a general framework intended to model MEs and to broaden thereby the area of application of methods of mathematical physics in addition to social and biological systems considered, e.g., in papers [3, 5, 8, 9, 19, 21, 17, 74, 79, 84, 85, 88, 98, 102, 101].

Our treatment of a ME is close to the minority game in the sense that agents simultaneously and adaptively compete for limited resources of the ME, have heterogeneous strategies, expectations, knowledge, objectives, etc.

The key idea of our approach is to provide each node of a ME with an intelligent agent to control operation of the node. Such a controlling agent should predict the state of the ME to elaborate decisions for the node. (The model considered below is a gen-
eralization of that presented in the previous section and we have to repeat some points.)

Let us represent a manufacturing enterprise (ME) as a directed graph consisting of \( N \) generalized nodes that operate in discrete time \( t = 0, 1, \ldots \). It is convenient to enumerate all relevant utilities (products, services, workers, materials, information, energy, and other resources) that can, in principle, appear as an input or an output of a node and collect them in a set \( U = \{ u_\mu, \mu \in M \} \) where (multi) index \( \mu \) may take on values from a set \( M \).

Each node gets some utilities from other nodes, transforms them (or, in particular, keeps intact), and gives some available utilities to other nodes. Thus a generalized node \( i \) evolves in discrete time \( (t = 0, 1, \ldots) \), has \( m + 1 \) inputs \( x^t_{i0\mu_0}, x^t_{i1\mu_1}, \ldots, x^t_{in\mu_m} \) and \( n + 1 \) outputs \( y^t_{i0\nu_0}, y^t_{i1\nu_1}, \ldots, y^t_{i\nu_n} \) where \( x^t_{ij\mu} \in U \) is the utility of kind \( \mu \) obtained at time \( t \) by node \( i \) from node \( j \); while \( y^t_{ik\nu} \in U \) is the utility of kind \( \nu \) transferred at time \( t \) by node \( i \) to node \( k \). Notice that input \( x^t_{i0\mu} \) denotes utilities produced by node \( i \) by itself, and output \( y^t_{i0\nu} \) corresponds to utilities qualified as end-products of the ME as a whole.

Control of operation of a node is assumed to be performed by an agent able to predict values of switching variables. These variables indicate decisions made by node \( i \). Namely, the value \( s^t_{ij\mu} \) corresponds to the decision made by node \( i \) at time moment \( t \) to give (starting with the next time moment \( t + 1 \)) utility \( \mu \) to node \( j \), while \( s^t_{ij\mu} = -1 \) means the decision to try to get utility \( \mu \) from node \( j \) (though this intention can appear not successful). We introduce also the probability \( P^t_{ij\mu}(T) \) that the decision to transfer utility \( \mu \) from node \( i \) to node \( j \) will be accomplished within time interval \( T \).

It is reasonable to deal with vector \( \sigma_i = (\sigma_{\alpha_1\beta_1\mu_1}, \sigma_{\alpha_2\beta_2\mu_2}, \ldots, \sigma_{\alpha_L\beta_l\mu_l}) \) whose components are quantities \( s^t_{ij\mu} \) arranged according to attainability and importance for the agent (node) \( i \). Then agent \( i \) again should make predictions: given values \( \sigma^t_\nu \) for \( t = t, t - 1, \ldots, t - T_\mu \), find values \( \sigma^t_\nu \) for \( t = t + 1, \ldots, t + T_f \).

Thus the problem of control of operation of a ME is reduced, in part, to the standard task of forecasting time series, and controlling agents should be provided with ability to accomplish forecasting relevant events. In particular, frames in Hilbert spaces can turn out to be an effective means for making such predictions.

### 5 Conclusions

The strategy behind this paper is that we generalize first particular tasks concerning information processing and cast them in an abstract form relevant for treatment by making use of techniques based on functional analysis and other mathematical disciplines. After such an ascent from the ground of specific problems to the top level of abstract consideration, we go through mathematics and acquire appropriate knowledge. Then we come up with new ways of handling our original problems being armed with powerful mathematical tools. The cardinal point of the developed approach is that all quantities of interest are treated as elements of a Hilbert space endowed with a Borel measure. It is worth noting that though we chiefly need just Hilbert spaces being guided by our specific applications, the notion of frame seems to be of the utmost importance and generality, can already be introduced in any vector space and may be successfully employed to handle various problems. Also, one can mention that methods of functional analysis are applied to the best advantage to a wide variety of problems of AI in a number of papers, just to wit such articles as [16, 25, 34, 46, 51, 59, 104].

It is worth noting also that any dynamical system can be treated as a generator of texts, which enables application of methods of text classification for investigation of its properties. In particular, one can address the issue how to predict what kind of behavior of the system will occur under given conditions because the problem of classification of dynamical regimes is easily reduced to the standard task of classification of documents. So, a maximum entropy technique may be used to cope with the latter problem.

Summarizing, the main acquisitions of the paper may be phrased as follows. Frames are defined in a vector space as a generalization of the notion of...
basis. A special emphasis is put on Hilbert space because it enables analysis and synthesis of functions that represent signals to be processed and are treated as vectors of the space. In particular, there appears possible to expand a function of interest over a set of functions that are linearly dependent. Two definitions and main properties of frames are considered, and an algorithm of the direct construction of a frame is presented. Frames are used to solve traditional tasks of information processing including least square approximation of experimental data, prediction of time series and estimation of probability distributions. It occurs beneficial to apply frames to such problems of AI as semantic treatment of information, text classification, medical diagnostics, control of the Internet traffic and manufacturing enterprises and so on.


A Appendix. Some Examples of Frames

A1 Frames in Finite-Dimensional Vector Spaces

Here we consider a simple case of a finite-dimensional vector space. More precisely, let the Euclidean space $\mathbf{R}^2$ be taken as the space $X$ considered in Chapter 2. We choose the vectors $< h_1 | = (2, 0), < h_2 | = (1, 1)$ and $< h_3 | = (\sigma, \lambda)$. Here $\sigma$ and $\lambda$ are real parameters to be determined so as these vectors could constitute a frame in $\mathbf{R}^2$, and $< h_a | = (|h_a>)^T$, $a = 1, 2, 3$, where the sign $T$ denotes the transposition. (The above vectors have been chosen as an illustration of the general consideration because the book [45] contains in Chapter 1 a detailed discussion of the basis composed of $|h_1>$ and $|h_2>$ along with related issues.)

In accord with the algorithm presented in Section 2.4 we compute the metric operator

$$G = diag\{(5 + \sigma^2), (1 + \lambda^2)\},$$

solve eq.(2.9) and find thereby the operator

$$T = G^{-1} = diag\{(5 + \sigma^2)^{-1}, (1 + \lambda^2)^{-1}\}.$$ 

(Remember that the notion $diag\{a, b\}$ stands for the matrix $\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$.) Using eq.(2.8) we calculate the vectors of the reciprocal frame:

$$< h_1 | = \frac{1}{2}(5 + \sigma^2)^{-1}, 0),$$

$$< h_2 | = \frac{1}{2}((5 + \sigma^2)^{-1}, -(1 + \lambda^2)^{-1}),$$

$$< h_3 | = \frac{1}{2}(\sigma(5 + \sigma^2)^{-1}, \lambda(1 + \lambda^2)^{-1}),$$

where the parameters $\sigma$ and $\lambda$ should meet the condition $\sigma\lambda = 1$.

Yet another finite-dimensional example can be found in [45] (see Chapter 4).

A2 Frames in Hilbert Spaces: Wavelets and Signal Processing

It is pointed out already in Section 2.1 that the currently popular wavelet analysis of signals rests on a class of frames (for an introduction into wavelet analysis see, e.g., [22, 23, 24, 29, 45, 76, 100]). Here we are going to demonstrate that the wavelets occur nothing but a particular case of frames whose general properties are examined in Chapter 2.

Indeed, the resolution of unity (2.1) in a special case when each $|h^a| > = |h_a| >$ takes the form

$$I = \int_A d\mu(a) |h_a| < h_a|.$$  \hspace{1cm} (A2.1)

The next step is the assumption that there exists an operator $\Psi$ which transforms vector $|a>$ of an appropriate space to the vector $|h_a| >$ such that $|h_a| > = \Psi|a| >$, $\forall a \in A$. Then eq.(1) yields

$$I = \int_A d\mu(a) \Psi^\dagger|a| < a|\Psi.$$  \hspace{1cm} (A2.2)

A particular case arises further when vector $|a>$ is actually specified by two discrete variables $j,k$ such that the latter expression takes the form

$$I = \sum_{j,k=-\infty}^{\infty} \Psi^+|j,k> <j,k|\Psi.$$  \hspace{1cm} (A2.3)

In the approach under consideration (see, e.g., Chapter 2 above or [34]) a signal $u(t)$ is represented as the inner product of a vector $|u> \in H$ and a basis vector $|t>$, i.e. $u(t) = <t|u>$. Then in view of eq.(3) we find

$$u(t) = <t|u> = \sum_{i,k=-\infty}^{\infty} <t|\Psi^+|j,k> <j,k|\Psi|u>.$$  \hspace{1cm} (A2.4)

It proves convenient and efficient to deal with a Hermitian operator $\Psi$ whose matrix elements are determined as follows:

$$\psi_{j,k}(t) \equiv <t|\Psi|j,k> = 2^{j/2}\psi(2^j t - k).$$  \hspace{1cm} (A2.5)

The latter function $\psi_{j,k}(t)$ is nothing but a kind of discrete wavelets investigated in [29]. A square integrable function $\psi(t)$ is known as the "mother", or "basic", wavelet, and accordingly $\Psi$ can be referred to as the generating operator.

Other types of wavelets may be treated similarly (see, e.g. [34]).

A3 Frames in Hilbert Spaces:

Coherent States of a Quantum Oscillator

As is mentioned in Section 2.1, yet another particular class of frames constitute, as a matter of fact, coherent states of a quantum harmonic oscillator which turns out to be an extremely beneficial model of various systems treated within quantum theories. These states introduced apparently in [12, 35] are defined as eigenvectors of the annihilation operator which occurs non-Hermitian (for more detail see, e.g., [40, 48, 49, 80]).

B Appendix. Bases

As the notion of frame is a generalization of the notion of basis, we suggest below a concise presentation of main properties of bases.

Let $X$ be a vector space over a field $K$. A basis in $X$ is a set of vectors $\{e_\alpha \in X, \alpha \in A\}$ such that (i) any vector $x \in X$ can be represented as a linear combination of these $e_\alpha$, i.e.

$$x = \sum_{\alpha \in A} x^\alpha e_\alpha, \forall x \in X,$$

and (ii) there exists just one set of coefficients $\{x^\alpha\}$ for which this expansion can be done. The coefficients $x^\alpha$ are called the components of the vector $x$ with respect to the basis $\{e_\alpha\}$.

Introduce the space $L$ of linear functionals $l : X \rightarrow K$ defined on $X$. This $L$ is called the dual space of $X$. For any $l \in L$ and $x \in X$, the value of the functional $l$ at the point (vector) $x$ is usually denoted as $l(x), <l|x>, <l,x>$, or $(l,x)$.

Let us construct a basis $\{l^\alpha \in L, \alpha \in A\}$ for the dual space $L$ as follows. For any $x = \sum_{\alpha \in A} x^\alpha e_\alpha$, define the functional $l^\alpha$ by means of the relation $<l^\alpha|x> = x^\alpha$, i.e. the functional $l^\alpha$, being applied to a vector $x \in X$, simply gives the (unique) $\alpha$-th component of $x$ with respect to the basis $\{e_\alpha\}$. It is easy to observe that each $l^\alpha$ is indeed linear. The set of all $\{l^\alpha \in L, \alpha \in A\}$ is called the dual basis of the basis $\{e_\alpha \in X, \alpha \in A\}$. Since, for basis vectors, one can write the expression

$$e_\beta = \sum_{\alpha \in A} \delta^\beta_\alpha e_\alpha, \forall \beta \in A,$$

the components of the basis vector $e_\beta$ itself are determined by the relation

$$<l^\alpha|e_\beta> = \delta^\alpha_\beta.$$

This duality relation can equally be employed to define the dual basis (instead of the definition given above). Also, it provides the so-called resolution of unity

$$I = \sum_{\alpha \in A} |e_\alpha> <l^\alpha| = \sum_{\alpha \in A} |l^\alpha> <e_\alpha|$$

so that one can write the following expansion of a vector \( x \in X \) over the basis \( \{ e_\alpha \in X, \ \alpha \in A \} \)

\[
x = \sum_{\alpha \in A} |e_\alpha \rangle < l^\alpha |x > = \sum_{\alpha \in A} x^\alpha |e_\alpha >
\]

with \( x^\alpha = < l^\alpha |x > \).

For a given vector space \( X \), a basis can appear (i) finite, (ii) countable infinite, or (iii) uncountable infinite. A vector space \( X \) is called finite-dimensional if it has a finite basis, otherwise it is an infinite-dimensional space. For any finite-dimensional vector space, the number of elements of a basis does not depend on the basis, and it is called the dimension of the space, \( \dim X \) or \( |X| \). If \( \dim X = n \) then \( X \) is called \( n \)-dimensional.

Basis vectors are customarily numbered by natural numbers from 1 (or 0) to \( n \), though this is not necessarily required. One can think of a basis of a vector space \( X \) simply as a subset of \( X \) with no labels for its elements. Numbering (or, more precisely, the order) of elements of a basis becomes significant as the matrix formalism is used when vectors are represented through columns or rows. It is essential however that elements of a basis should anyway be linearly independent.

An infinite-dimensional vector space is usually supplied with a kind of topology, and a basis is defined taking into account this topology and the available possibility to construct infinite-dimensional linear combinations.

### C Appendix. Calculation of a Frame Approximately Reciprocal to a Given Frame

In Section 2.3 and 2.4 we gave two equivalent, in essence, definitions of frames. Here we are going to show how to construct a frame which is approximately reciprocal (dual) to a given frame.

Let us suppose that we have some frame, i.e. a set of vectors \( |\psi_\alpha > \) ensuring resolution of unity

\[
I = \sum_{\alpha, \beta} |\psi_\alpha > \Delta_{\alpha \beta} < \psi_\beta |
\]

with known quantities \( \Delta_{\alpha \beta} \). Then using the operator \( T \) defined by eq.(2.5) we can write

\[
|h^a > = \sum_{\alpha, \beta, \gamma, \delta} |\psi_\alpha > \Delta_{\alpha \beta} T_{\beta \gamma} \Delta_{\gamma \delta} < \psi_\delta |h_a >
\]

where \( T_{\beta \gamma} = < \psi_\beta |T|\psi_\gamma > \). The latter matrix elements can be found as follows. After multiplication of eq.(2.7) by \( < \psi_\alpha | \) and some auxiliary \( |\chi^\mu > \) we obtain

\[
\sum_{\beta} T_{\alpha \beta} v^\mu_{\beta} = u^\mu_{\alpha}
\]

where \( v^\mu_{\beta} = \sum_{\gamma} \Delta_{\beta \gamma} < \psi_\gamma |G|\chi^\mu >, \)

\( u^\mu_{\alpha} = < \psi_\alpha |\chi^\mu > \).

Thus the vectors \( |h^a > \) of the reciprocal frame are determined by eq.(2) where the quantities \( T_{\beta \gamma} \) are solutions of eqs.(3) with some appropriate vectors \( |\chi^\mu > \).

Observe that if the Hilbert space \( H \) under consideration is infinite-dimensional, then only an infinite set of vectors may constitute a frame. Hence the system of equations (14) turns out to be infinite. In practice, however, one has to restrict oneself by finite-dimensional approximations. As a result, the system (3) becomes finite as well, though we can meet here a case when the system appears underdetermined, overdetermined or inconsistent because of numeric values of the coefficients so that there does not exist an inverse matrix for the matrix composed of the values \( v^\mu_{\beta} \). Such systems are considered in, e.g., in [63, 61].

### D Appendix. Knowledge Extraction from the WWW: Some Features

In this Appendix we address some issues of semantic treatment of (hyper) texts that appears of great concern as the WWW becomes the information source of paramount importance. The approach under consideration might be useful for development of Web spiders and search engines, Web-based knowledge bases, interface assistants and so on.
For ordinary text documents, the ground of classification is, as usual, the content of the document (e.g., the set of words it contains). In contrast, the WWW provides diverse sources of information: (1) the full text of pages, (2) the text in titles and headings, (3) the text associated with the hyperlinks, (4) the text in neighboring pages, (5) the file organization provided by URLs. Accordingly, different kinds of classifiers can be designed. However, not a means appears able to cope with Web pages with sufficient accuracy (see, e.g., [26]), and combining different approaches could be promising. Obverse that the vector associated in our approach with a document can incorporate data obtained from any source of information about the contents of the document (words in the main body of a hypertext, words in its title, heading, hyperlinks, keywords, abstract, annotations, etc.). In particular, the power of the method increases by combining it with annotating the contents of Web pages supported by such languages as SHOE and XML.

A peculiarity of information structured in the Web is that "hyperlinks encode a considerable amount of latent human judgement" [50]. So, the creator of a Web page $p$, by including a hyperlink to a page $q$, has a certain authority on $q$, whereas a hub page points to multiple relevant authorities. It seems perfectly reasonable to associate each of authorities with a vector included in a frame of the semantic space. Then the user is provided with facilities to search for Web pages that are close, to a certain prescribed degree, to particular chosen authorities.

Summarizing one can suppose that application of frames might provide a new means to advance computer systems for retrieval of information and extraction of knowledge from the Internet.

References

[54] Krot A.M., Minervina H.B. Minimal attractor embedding estimation based on matrix decomposition for analysis of nonlinear dynamical systems. - Submitted to ”Nonlinear Phenomena in Complex Systems”.


Malik M., Camm A.J. (Eds.) Heart Rate Variability. (Futura, Armonk, 1995).


