Chaos and Natural Language Processing

Marius Crisan

Department of Computer and Software Engineering
Polytechnic University of Timisoara
V. Parvan 2, 300223 Timisoara, Romania
Tel.: +40256403254, E-mail: crisan@cs.upt.ro, http://www.cs.upt.ro/~crisan

Abstract: The article explores the possibility to construct a unified word feature out of the component features of letters. Each letter is modeled by a different attractor and finally embedded in a quadratic iterated map. The result is the word feature that can account for the meaning extraction process of language understanding. This is a new approach in natural language processing based on the deterministic chaotic behavior of dynamical systems.

1   Introduction

There is an increased interest in the modern era for developing techniques for both speech (or character/word sequences) recognition and synthesis. Natural language processing (NLP) provides those computational techniques that process spoken and written human language. An important class of methods for language recognition and generation is based on probabilistic models, such as N-grams model, Hidden Markov and Maximum Entropy model [1]. Given a sequence of units (words, letters, morphemes, sentences, etc.) these models try to compute a probability distribution over possible labels and choose the best label sequence. Another approach in NLP is to use neural networks, in particular self-organizing maps of symbol strings [2]-[4]. However, an important challenge for any NLP approach which may hinder its success is dealing with the nonlinear character of language phenomenon. Starting from the premise that natural language phenomena can be viewed as a dynamical system the purpose of this work is to investigate the possibility of modeling words/characters by a chaotic attractor.

2   Meaning as Wholeness in Dynamical Systems

We may consider consistently with other theories of language that the notion of *word* is the constituent element of a sentence (utterance). We might, at first, also

consider in a general formalization that a word is any sound-sequence that possesses the property of inflection. Normally, each word takes either a verbal, i.e., conjugational inflection, in which case it is called a verb, or a nominal, i.e., declensional inflection, in which case it is of a non-verbal category (substantives, adjectives, participles, etc.). All the other words which do not have declensional inflections, such as prepositions, may be considered to possess invariant inflection. However, to classify words only in terms of their inflection property is an incomplete task, and does not seem to help much in explaining how the meaning as structured information is conveyed by a sentence. Therefore, I suggest the employment of semantic criteria in defining the notion of word. According to such a view, a word is the meaning-bearing element of a sentence. The semantic criterion determines the minimum sequence length of the phonemes which convey a meaning. Thus, words may vary in complexity, from the shortest meaning-bearing ones to the more complex compound words. Based only on meaningful words, we may define, in general terms, a sentence as being a cluster of words capable to generate a cognitive meaning in an ideal receiver (hearer/reader). This cognition is a result of a reaction mechanism triggered by the series of words in the sentence.

An ideal receiver is qualified by the 'capacity' to extract meaning from a sentence. This capacity can be described by the cognition of four cognitive properties that have been assigned by the transmitter (speaker/writer) to a sentence: (1) semantic competency, (2) expectancy (syntactic/semantic), (3) contiguity in space and time, and (4) transmitter's intention [5]. These cognitive properties are the requirements for defining a grammatical and meaning-bearing sentence. A sentence is said to have semantic competency when the objects denoted by the respective words are compatible one to another. For instance, the sentence '*He sees the light.*' is grammatically acceptable, and has semantic competency, while the sentence '*He hears the color.*' even if it is grammatically acceptable, lacks semantic competency. Semantic expectancy refers to the capacity of an ideal receiver to infer the meaning of an incomplete sentence (utterance). Syntactic expectancy refers to the syntactic property x which has o be assigned to a sentence s when it is not grammatical, in order to make it suitable to transmit the meaning. This expectancy is measured by the predictor of the entropy of the entropic source. Contiguity is the property which imposes the absence of any unnecessary spatial (in written text) or temporal (in uttered) interval between the words of a sentence.

In a previous work [6], in defining meaning as something that must have a finite description, I introduced the concept of undivided meaning-whole (UMW). This is conceived as structured information which exists internally at the agent's information level. Even if UMW is a unitary information structure, it is describable rationally in terms of cognitive semantic units. These semantic units are the generating principle of producing the sequence of uttered words.

When an agent wants to communicate, it begins with the UMW existing internally in its knowledge base. When words are uttered producing different sounds in

sequence, it appears only to have differentiation. Ultimately, the sound sequence is perceived as a unity or UMW and only then the word meaning, which is also inherently present in the receiver's mind, is identified.

The above described capacity of the receiver to extract meaning from series of words led to another assumption, that the whole word/sentence meaning has to be inherently present in the mind of each agent. Thus, it can be explained how it is possible the UMW to be grasped by the hearer even before the whole sentence has been uttered. The sounds which differ from one another because of difference in pronouncement cause the cognition of the one changeless UMW without determining any change in it. Sometimes, reasoning may have to be applied to the components of the sentence so that the cognition is sufficiently clear to make possible the perception of the meaning-whole. It appears that the unitary word-meaning is an object of each person's own cognitive perception. When a word, such as 'tree' is pronounced or read there is the unitary perception or simultaneous cognition of trunk, branches, leaves, fruits, etc. in the receiver's mind. Communication (verbal or written) between people is only possible because of the existence of the UMW, which is potentially perceivable by everybody and revealed by words' sounds or symbols.

The concept of UMW is consistent with a more general view, suggested by Bohm in [7], regarding the possibilities for wholeness in the quantum theory to have an objective significance. This is in contrast with the classical view which must treat a whole as merely a convenient way of thinking about what is considered to be in reality nothing but a collection of independent parts in a mechanical kind of interaction. If wholeness and non-locality is an underlying reality then all the other natural phenomena must, one way or another, be consistent with such a model. Natural language generation and understanding is a phenomenon that might be modeled in such a way. UMW is like 'active information' in Bohm's language, and is the activity of form, rather than of substance. As Bohm puts it clearly [7], '…when we read a printed page, we do not assimilate the substance of the paper, but only the forms of the letters, and it is these forms which give rise to an information content in the reader which is manifested actively in his or her subsequent activities.' But, similar so called mind-like quality of matter reveals itself strongly at the quantum level. The form of the wave function manifests itself in the movements of the particles. From here, a new possibility of modeling the mind as a dynamical system is considered.

In line with Kantian thought, in [8] we find a similar insight, as above, regarding the linguistic apprehension. This is the interplay of two factors of different levels: (1) the empirical manifold of the separate letters or words and (2) the *a priori* synthesis of the manifold which imparts a unity to those elements which would otherwise have remained a mere manifold.

According to this kind of observations it appears motivated to use the concept of manifold for modeling the mind as the seat of language generation and

understanding. Manifolds are defined as topological spaces possessing families of local coordinate systems that are related to each other by coordinate transformations pertaining to a specific class. They may be seen also as the multidimensional analogue of a curved surface. This property seems suitable to represent both the natural language constraints and semantic content of linguistic objects.

Usually, a dynamical system is a smooth action of the reals or the integers on a manifold. The manifold is the state space or phase space of the system. Having a continuous function, F, the evolution of a variable x can then be given by the equation:

$$x_{t+1} = F(x_t). \tag{1}$$

The same system can behave either predictably or chaotically, depending on small changes in a single term of the equations that describe the system. Equation (1) can also be viewed as a difference equation ($x_{t+1} - x_t = F(x_t) - x_t$) and generates iterated maps. An important property of dynamical systems is that even very simple systems, described by simple equations, can have chaotic solutions. This doesn't mean that chaotic processes are random. They follow rules, but even the simple rules can produce amazing complexity. In this regard, another important concept is that of an attractor. An attractor is a region of state space invariant under the dynamics, towards which neighboring states in a given basin of attraction asymptotically approach in the course of dynamic evolution. The basin of attraction defines the set of points in the space of system variables such that initial conditions chosen in this set dynamically evolve to a particular attractor. It is important to note that a dynamical system may have multiple attractors that may coexist, each with its own basin of attraction [9]. This type of behavior is suitable for modeling self-organizing processes, and is thought to be a condition for a realistic representation of natural processes.

One example of such an approach is the topological feature map proposed by Kohonen [10], [11] for the projection of high dimensional pattern data into a low-dimensional feature space. The process of ordering an initial random map is called in this approach self-organization. The result is the topological ordering of pattern projections, or in other words the self-organizing map (SOM). Each input dimension is called a feature and is represented by an N-dimensional vector. Each node in the SOM is assigned an N-dimensional vector and is connected to every input dimension. The components or weights of this vector are adjusted following an unsupervised learning process. First, it is found the winning node, i.e, the node whose weight vector shows the best match with the input vector in the N-dimensional space. Next, all weight vectors in the neighborhood in the direction given by the input vector are adjusted. This process requires many iterations until it converges, i.e., all the adjustments approach zero. It begins with a large neighborhood and then gradually reduces it to a very small neighborhood. Consequently, the feature maps achieve both ordering and convergence properties,

and offer the advantages, of reducing dimensions and displaying similarities. However, SOM solutions (and neural networks in general) are yet in the need for improvement. For instance, in [12], an important problem for SOM is discussed. In order to obtain a realistic speech projection, the problem is to find a hypercubical SOM lattice where the sequences of projected speech feature vectors form continuous trajectories. In another work [13], both SOM and a supervised multilayer perceptron were used for bird sounds recognition. The conclusion was that although the tested algorithms proved to be quite robust recognition methods for a limited set of birds, the proposed method cannot beat a human expert listener.

On the other hand, the unexplored domain of dynamical systems and chaos theory may offer promising perspectives in modeling natural processes, and NLP might be one of them.

3   Attractor-based Word Modeling

In quantum experiments, when particles interact, it is as if they were all connected by indivisible links into a single whole. The same behavior is manifested by the chaotic solutions in an attractor, as we will see in this section. In spite of the apparent random behavior of these phenomena, there is an ordered pattern given by the form of the quantum wave (or potential) in the former case, and by the equations of the dynamic system in the latter.

Let's consider the simplest case of the quadratic iterated map described by the equation:

$$x_{t+1} = a_1 + a_2 x_t + a_3 x_t^2 \tag{2}$$

Even if it is so simple, it is nonlinearly stable and can manifest chaotic solutions. The initial conditions are drawn to a special type of attractor called a strange attractor. This may appear as a complicated geometrical object which gives the form of the dynamic behavior.

In nonlinear dynamics the problem is to predict if a given flow will pass through a given region of state space in finite time. One way to decide if the nonlinear system is stable is to actually simulate the dynamics of the equation. The primary method in the field of nonlinear dynamic systems is simply varying the coefficients of the nonlinear terms in a nonlinear equation and examining the behavior of the solutions. The initial values of the components of the model vector, $m_i(t)$, were selected at random in a process of finding a strange attractor. Strange attractors are bounded regions of phase space corresponding to positive Lyapunov exponents. We found more than 100 attractors. In Table I we presented a list of several coefficients along with the Lyapunov exponent for which the attractors were found by random search. The initial condition x_0 was selected in

the range 0.01 – 1 and lies within the basin in many cases. The Lyapunov exponent is computed in an iterated process according to the following equation [14], [15]:

$$LE = \Sigma \log_2 |a_2 + 2a_3 x_t| \, /N \qquad\qquad (3)$$

The sum is taken from a value of $t = 1$ to a value of $t = N$, where N is some large number.

Table 1

The coefficients values and the Lyapunov exponent for 25 attractors of (2)

Cur. No.	a_1	a_2	a_3	LE
1	1.2	-1	-1	0.4235
2	1.2	-0.2	-1.1	0.1198
3	1.1	-1.2	-0.8	0.3564
4	1.1	-1	-0.6	6.6073
5	1.1	-0.6	-1	0.1443
6	1	-0.7	-1.1	0.2512
7	0.9	-1.1	-1.1	0.3571
8	0.9	-1.1	-0.8	0.256
9	0.8	-1.2	-1.2	0.411
10	0.8	-0.9	-1	0.1383
11	0.7	-1.2	-0.8	0.2001
12	0.7	-1.1	-1.2	0.3029
13	-1.2	-1.2	0.7	0.2918
14	-1.2	-0.9	0.8	0.2793
15	-1.2	-0.6	1	0.2662
16	-1.1	-0.8	1	0.286
17	-1.1	-1	0.9	0.3054
18	-1	-1	0.7	0.1209
19	-0.8	-1.1	1.1	0.3047
20	-0.8	-1.1	0.7	6.9382
21	-0.7	-1	1	0.1248
22	-0.7	-1.2	1	0.285
23	-0.6	-1.2	1.2	0.2801
24	-0.5	-1.1	1.2	0.1375
25	-0.4	-1.2	1.2	0.1344

LE gives the rate of exponential divergence from perturbed initial conditions. If the value is positive (for instance, greater than 0.005) then there is sensitivity to initial conditions and a strange attractor can manifest. If the solution is chaotic, the successive iterates get farther apart, and the difference usually increases exponentially. The larger the *LE,* the greater is the rate of exponential divergence, and the wider the corresponding separatrix of the chaotic region. If *LE* is negative,

the solutions approach one another. If *LE* is 0 then the attractors are regular. They act as limit cycles, in which trajectories circle around a limiting trajectory which they asymptotically approach, but never reach.
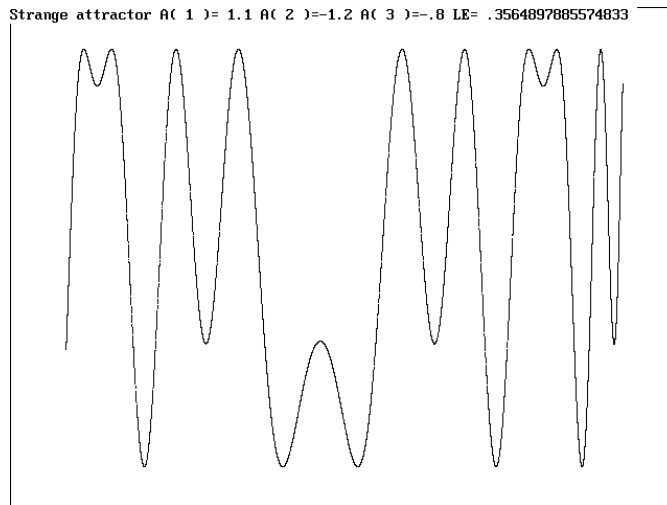

Figure 1
Quadratic iterated map of (2)

It's interesting to analyze in more details the behavior of an attractor. The idea of the self-organizing maps is to project the *N*-dimensional data into something that is better understood visually. A similar idea we follow in constructing iterated maps. It is convenient to plot the values in the iterated process versus their fifth previous iterate for a more suggestive aspect. In Fig. 1 it is presented the iterated map for the strange attractor No. 3. A remarkable property of the chaotic solutions, as noted above in connection with quantum, is the 'ballet-like' behavior as iterations progress. Each new dot on the map, representing the solution x_{t+1}, appears in a random position but orderly following the attractor's form.

In Fig. 2 it is shown the same attractor only after a few iterates (2000). It can be seen the sparse distribution of dots but along with the ordered path. This type of behavior is similar with the quantum phenomena, such as the distribution of photons along the interference pattern lines in the two slit interference experiment, when the photons are emitted in series one after the other. This is also akin to the quality of the perception act (word meaning). It is observed that a word meaning is at first perceived vaguely and then more and more clearly. Thus, through the process of repeated perception or iterations finally the meaning is revealed. We may suggest, therefore, that meaning can be mathematically modeled as a basin of attraction.
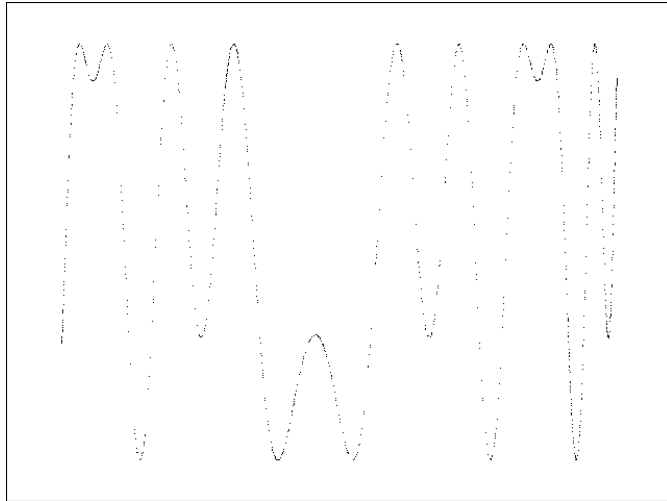
Figure 2

Quadratic iterated map of (2) after 2000 iterates. Note the sparse distribution of dots along the regular pattern of the strange attractor.

Another interesting property is the symmetry of a_1 and a_3 and the corresponding iterated map. Considering again the strange attractor $a_1 = 1.1$, $a_2 = -1.2$, $a_3 = -0.8$, a symmetric behavior can be obtain for the values $a_1 = -1.1$, $a_2 = -1.2$, $a_3 = 0.8$ as in Fig. 3.
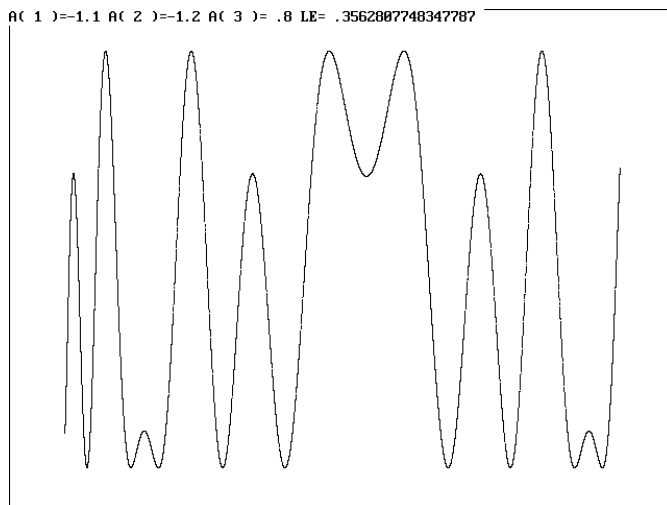

Figure 3

The symmetric quadratic iterated map of Fig 1, obtained by inversing the sign of a_1 and a_3

There is a huge possibility to obtain other attractors by tuning the values of the coefficients. The shape of the attractor changes smoothly with small variations of the coefficients. Even if the interval of variation is rather small, visible changes in the shape of the map can be obtained. For instance, if $a_1 = 1.02$ the value of LE is 0.09 and the limit cycles can be observed as the attractor becomes regular. If $a_1 = 1.3$ regular oscillations are manifested.

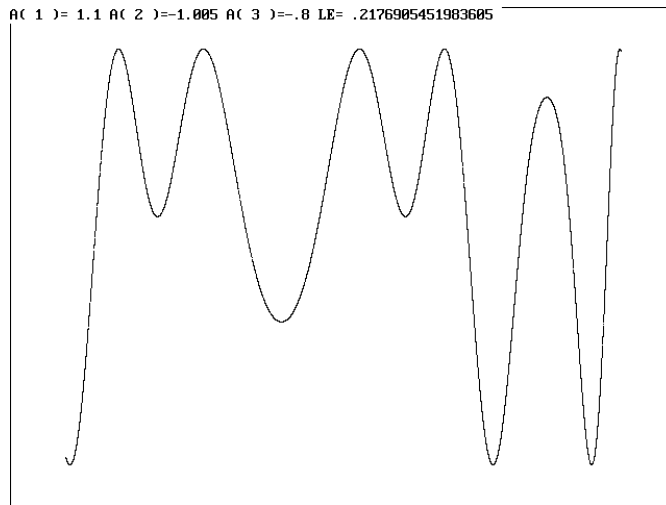A(1)= 1.1 A(2)=-1.005 A(3)=-.8 LE= .2176905451983605

Figure 4

The completely changed quadratic iterated map of (2), obtained for $a_2 = -1.005$

An important change in shape can be obtained by modifying a_2. The value for a_2 always has to be negative for a bounded behavior. For $a_2 = -1.005$ the shape of the map is drastically changed as shown in Fig. 4.

4   Language Recognition

One widely used method to classify an object is to measure its features (characteristic properties). In general the features that are to be observed depend on the specific problem one has to solve. In language recognition, we deal with several kinds of features such as graphological, phonological, statistical, syntactic, lexical, semantic, and pragmatic. Graphological features are for instance letter positions and word shape. Phonological features are considered as the distinctive features from which phonemes can be constructed [16]. The syntactic features are present in the construction of words and sentences, and are part of speech tags and various components from a parse tree. Statistical features exploit the fact that more frequently occurring words are more familiar and hence more easily recognized. These features may be the frequency of occurrence of letters, letter

pairs and triplets, the average word length, the ratio of certain characters, word endings, consonant congestion [17], etc. Lexical features are used to represent the context. They consist of unigrams, bigrams (a pair of words that occur close to each other in text and in a particular order), and the surface form of the target word which may restrict its possible senses [18]. Semantic features indicate the meaning of words, and are usable for disambiguation of words in context. Pragmatic features are based on how the words are used. In general, the above mentioned features are tried to be described by morphology using the concept of morphemes or the constituent parts of words.

Irrespective of the feature's nature, the result of feature extraction or measurement is a set described as an **n**-dimensional feature vector associated with the observed object, which can thus be represented as a point in the **n**-dimensional feature space. Next, a classifier will assign the vector to one of several categories. While the use of features has a central place in pattern classification [19], the design and detection of features in natural language remains a difficult task because of language high complexity and the lack of a unitary theory.

The analysis in the previous section revealed the fact that attractors offer dynamic properties that can map in a continuous manner the feature vectors according to some input patterns. Considering the assumption of UMW, the goal is to construct a unified word feature that might account for the word meaning. I propose a possible non-linear many-to-one mapping from a conventional feature space to a new space constructed so that each word has a unique feature vector. Let's consider the simpler case of a 3-dimensional feature vector characterizing a letter. The vector for a generic letter 'A' is defined by the values $\boldsymbol{a} = [a_1, a_2, a_3]$, and similarly for the generic letters 'B' and 'C' the vectors are $\boldsymbol{b} = [b_1, b_2, b_3]$, and $\boldsymbol{c} = [c_1, c_2, c_3]$ respectively. The letter feature of 'A' results in an iterated process as

$$A_{t+1} = a_1 + a_2 A_t + a_3 A_t^2,\tag{3}$$

starting from an initial condition A_0.

Similar equations result for the letter features of 'B' and 'C', with the initial conditions B_0 and C_0 respectively, as the following:

$$B_{t+1} = b_1 + b_2 B_t + b_3 B_t^2,\tag{4}$$

$$C_{t+1} = c_1 + c_2 C_t + c_3 C_t^2.\tag{5}$$

Based on letters features, for each letter in a word (for instance with length 3) a unified feature vector $W = [A, B, C]$ can be constructed and mapped to the three coefficients of an equation of type (2). The result is of the following form:

$$W_{t+1} = A_t + B_t W_t + C_t W_t^2.\tag{6}$$

Eq. (6) is computed starting from an initial condition W_0 and manifests a chaotic deterministic behavior for a proper combination of the coefficients A, B, and C. In Fig. 5 it is presented the iterated map of (6) for the input vectors $\boldsymbol{a} = [0.8, -1.2, -$

0.9], b = [–1, –0.9, 1.1], and c = [1.1, –1.2, –0.8] after 5000 iterations, and the initial condition for all parameters of value 0.01. In order to have a suggestive view of the unified feature space and observe its internal structure, the values were plotted versus their third previous iteration. Also, the values of W were bounded to 10^8 for a convenient screening. The same sparse distribution of dots along the regular pattern of the feature space, typical for deterministic chaos, can be observed as the iterations progress.
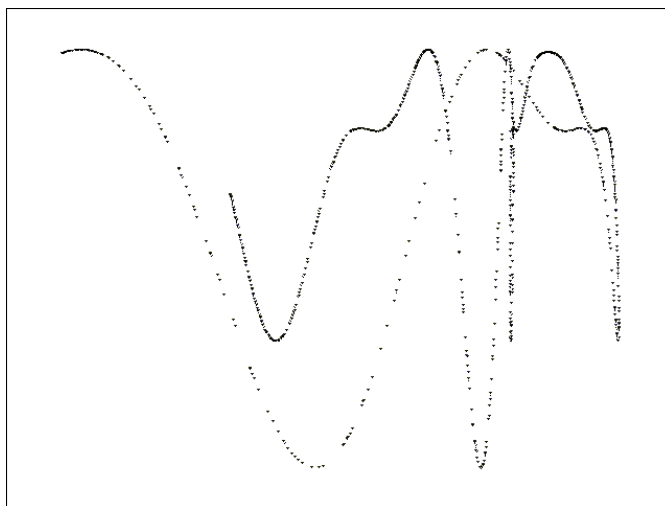

Figure 5
The chaotic deterministic behavior of (6) for the letter feature vectors $\mathbf{a}$ = [0.8, –1.2, –0.9], $\mathbf{b}$ = [–1, –0.9, 1.1], and $\mathbf{c}$ = [1.1, –1.2, –0.8]

Each valid word of length 3 will determine a corresponding iterated map. Small variations in the input will be tolerated and recognized with the same meaning but other illegal combinations will be rejected. For instance, in Fig. 6 we can see the feature space for a rather consistent deformation of the input vectors $\boldsymbol{b}$ = [–1.3, –0.6, 1.3] and $\boldsymbol{c}$ = [.9, –1.3, –1].

Comparing the feature spaces of Fig. 5 and Fig. 6 we can observe the vague resemblance between the two, and after a closer examination we can identify in fact a similar chaotic pattern. This means that the meaning was conserved even if some visible alterations affected two of the letter features. If the changes are more dramatic we expect a completely different pattern or even an unbound behavior. This indicates the lack of properties for a meaningful word. In Fig. 7 it is presented the case where the vectors $\mathbf{a}$ and $\mathbf{c}$ swapped their contents. This means another word where the first and last letters are interchanged.
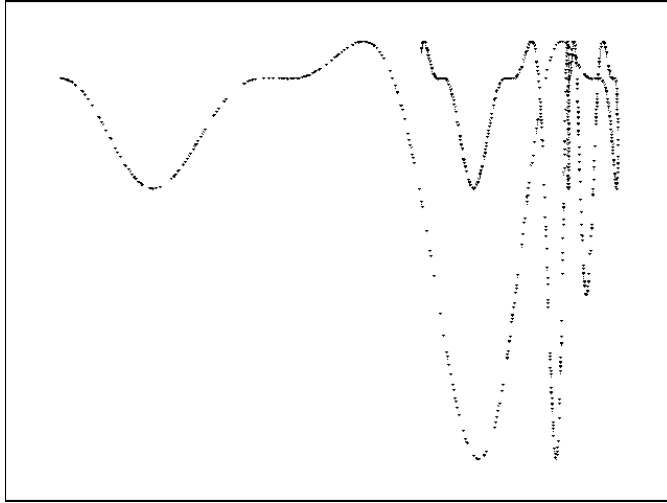
Figure 6

The feature space of (6) for **a** = [0.8, –1.2, –0.9], **b** = [–1.3, –0.6, 1.3], and **c** = [.9, –1.3, –1]. Note the vague resemblance with Fig. 5.
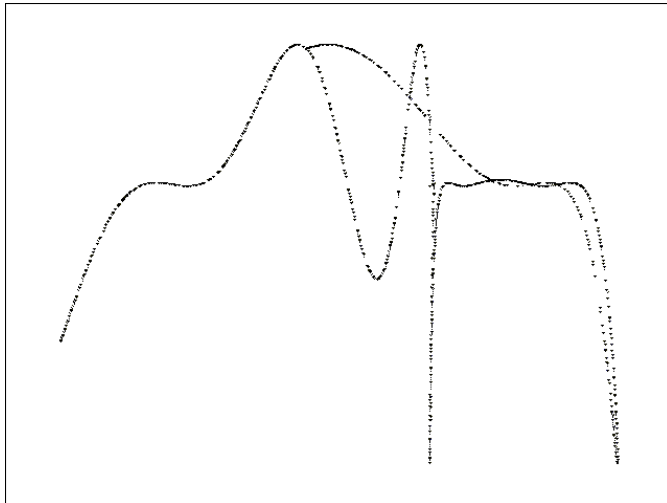

Figure 7

The chaotic deterministic behavior of (6) for the letter feature vectors **a** = [1.1, –1.2, –0.8], **b** = [–1, –0.9, 1.1], and **c** = [0.8, –1.2, –0.9]

A completely different pattern is obtained comparing to Fig. 5. Of course, depending on the classifier conventions, the pattern can be meaningful or not. In any case, it represents the unique feature vector for that word construction.

For words with higher length, higher-order iterated maps can be used. The proposed approach can be extended for a whole sentence. In this case, the unified feature vector of the sentence is constructed based on the features of the individual component words. This will be the UMW equivalent of the whole sentence.

Conclusions

Our purpose was to study the possibility of using dynamical systems in modeling natural language processing. We started from the premise of UMW and the observation facts of language apprehension and noted a similitude with the chaotic behavior of dynamical systems. The attractor behavior as was studied for the quadratic iterated maps seems to be robust enough to model the feature vectors formed for each word of length 3 in the dictionary. The unified word feature vector is obtained by a many-to-one mapping, starting from the component letters, and bears the unique information structure of the word meaning. Slight variations in the input feature vectors of the component letters are tolerated, without major changes of the pattern in feature space structure. This is an indication of meaning preservation in the case of noise. The chaotic deterministic behavior of the patterns in the feature space may account for meaning recognition process after a series of repeated perceptions. After enough iterations (or repeated perception) the attractor shape is recognized and consequently the corresponding meaning. The present work may be continued in the future by constructing the unified feature vector at the sentence level.

References

[1]   D. Jurafsky, J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Prentice-Hall, 2000

[2]   T. Kohonen, "Self-organizing maps of symbol strings", Report A42, Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, Finland, 1996

[3]   T. Kohonen, P. Somervuo, "Self-organizing maps of symbol strings with application to speech recognition", in Proc. of Workshop on Self-Organizing Maps (WSOM'97), pp. 2-7, Espoo, Finland, 1997

[4]   T. Honkela., "Self-Organizing Maps in Natural Language Processing", Espoo, Finland, 1997

[5]   B. K. Matilal, *Logic, Language and Reality*, Motilal Banarsidass Publ., Delhi, 1997

[6]   M. Crisan, "Meaning as Cognition," *Proceedings of the I International Conference on Multidisciplinary Information Sciences and Technologies-InSciT2006*, Merida, Spain, 2006, pp. 369-373

[7]   D. Bohm, "A new theory of the relationship of mind and matter," *Philosophical Psychology*, Vol. 3, No. 2, 1990, pp. 271-286

[8]   H. G. Coward, *The Sphota Theory of Language*, Motilal Banarsidass Delhi, 3rd ed. 1997

[9]   G. Pulin and X. Jianxue, "On the multiple-attractor coexisting system with parameter uncertainties using generalized cell mapping method," *Journal Applied Mathematics and Mechanics*, Vol. 19, No. 12 December, 1998, pp. 1179-1187

[10]   T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, 43:59-69, 1982

[11]   T. Kohonen, *Self-Organizing Maps*, Springer, Berlin (3rd extended ed. 2001), 1997

[12]   P. Somervuo, "Speech Dimensionality Analysis on Hypercubical Self-Organizing Maps," *Neural Processing Letters*, Vol. 17-2, April 2003, pp. 125-136

[13]   A. Selin, J. Turunen, and J. T. Tanttu, "Wavelets in Recognition of Bird Sounds," *EURASIP Journal on Advances in Signal Processing*, Vol. 2007, Article ID 51806, 9 pages, doi:10.1155/2007/51806

[14]   J. C. Sprott, *Strange Attractors: Creating Patterns in Chaos*, M & T Books, 1993-09

[15]   J. C. Sprott, *Chaos and Time-Series Analysis*, Oxford University Press, 2003

[16]   S. King and P. Taylor, "Detection of phonological features in continuous speech using neural networks," *Computer Speech & Language*, Volume 14, Number 4, October 2000 , pp. 333-353(21)

[17]   G. Windisch and L. Csink, "Language Identification Using Global Statistics of Natural Languages," *Proc. of 2[nd]Romanian-Hungarian Joint Symposium on Applied Computational Intelligence*, Timisoara, Romania, May 12-14, 2005, pp. 243-255

[18]   S. Mohammad and T. Pedersen, "Combining Lexical and Syntactic Features for Supervised Word Sense Disambiguation", *Proceedings of the Conference on Computational Natural Language Learning* (CoNLL-2004), May, 2004, Boston, MA

[19]   R. O. Duda, P. E. Hart and David G. Stork, *Pattern Classification*, John Wiley Interscience, 2001