On Representations

Bo Xu[1]
Department of Physiology and Biophysics
School of Medicine
Indiana University
itgt500@indycms.iupui.edu

Liqing Zheng
Department of Electrical Engineering
Purdue University, Indianapolis

Abstract ---- First, it was pointed out in this paper that there exists a hierarchical organization of entity to be represented by the neural networks. This hierarchical organization leads to four levels of correspondences between the entity, patterns, elements, items and the units. Then we classified representations further into more detailed categories from two viewpoints. From the correspondence viewpoint, representations can be classified into five categories: (1) local representation, (2) one-to-one DR, (3) one-to-many DR, (4) many-to-one DR, and (5) many-to-many DR. The second viewpoint of classifying representations is according to the locations where the representations exist. From the location viewpoint, representations fall into two classes: internal representation and external representation. Finally, it was pointed out that the origins of the confusion on representations may come from: (1) non-distinction of the hierarchical organization of entity, (2) non-distinction of the four levels of correspondences between the entity, patterns, elements, items and the units, (3) use of the too general or too broad term *distributed representation*, (4) non-distinction of the two levels of graceful degradation, and (5) non-distinction of similarities between different representations.

Keywords ---- Hierarchical organization of entity, Local representation (LR), Distributed representation (DR), One-to-one DR, One-to-many DR, Many-to-one DR, Many-to-many DR, Internal representation, External representation.

1. INTRODUCTION

According to David Marr, "a representation is a formal system for making explicit certain entities or types of information, together with a specification of how the system does this" (Marr, 1982). In the general sense, representation is not a new concept. It has long been used consciously or subconsciously in nearly all aspects of our activities, from everyday life to various special areas. In these areas, because a representation is only a description of a *static object* (a representation was simply a description of an object to be represented), the importance of representation was not noticed. Therefore, the representation was not given special attention in these areas.

However, the situation in the information processing systems (such as neural networks) is quite different. Because an information processing system is a *dynamic system* instead of a *static object,* the representation becomes a factor of crucial importance to the performance of the information processing system. (In this paper, we will consider the representations in neural networks mainly.)

For a neural net, the representation at the input layer not only describes the input information, but also serves as an input format for the neural net to process in the forward direction. So the

[1] Bo Xu thanks Dr. Ali Minai for a long but helpful dialogue at Connectionists in 1991. We are grateful to Dr. Scott Fahlman for valuable discussions and suggestions.

This paper was accepted as an oral presentation by the *SPIE Intelligent Information Systems OR'92--Science of Artificial Neural Networks,* Orlando, Florida.

input representation will influence the dynamic performance of a neural net in addtion to describing the input information.

At the output layer, the output representation not only describes the output information, but also affects the neural net's performance through the network topology and backpropagation. So the output representation also influence the dynamic performance of the network in addtion to describing the output information.

In sum, both the input representation and the output representation have two functions: describing information and influencing the dynamic performance of the network. It is this functional duality of representations in neural nets that makes the representation a very important factor in neural networks. Due to this reason, we suggest that the neural network architecture include the representation in addition to the learning algorithm, unit characteristics, and network topology (Lippmann, 1987).

Up to now, there are mainly two kinds of representations being proposed and studied in the neural network area: local representation (LR), also called grandmother cell representation, and distributed representation (DR). Originally, the LR is defined as the one in which each entity is represented by a single unit (Barlow, 1972; Feldman, 1986), and the DR is defined as the one in which each entity is represented by a pattern of activity distributed over many units, and each unit is involved in representing many different entities (Hinton, 1981; Hinton, McClelland and Rumelhart, 1986; Anderson and Rosenfeld, 1988).

However, there exists much confusion on the advantages and disadvantages of local vs. distributed representations. Realizing the necessity of separating the representation of an object into two parts: one for the type, and the other for the instance contained within the type (so-called type/instance hierarchy), Hinton, McClelland, and Rumelhart (1986) made a very good statement on the confused problems facing the distributed representation: "The question of how to represent the relationship between an item and the constituent items of which it is composed has been a major stumbling block for theories that postulate distributed representations". They further pointed out that for the distributed representations, two sets of units have to be used in order to represent many objects with the same property (Hinton, McClelland, and Rumelhart,1986). Realizing the difficulties of this distributed representation approach, they said: "In general, it is hard to specify in advance just what 'order' of conjunctions will be required"; "It is often unclear how structure is to be captured in a distributed representational scheme". In order to describe different situations contained in the distributed representation, terms such as semi-distributed representation and full-distributed representation (Sejnowski, Koch, and Churchland, 1990) were used. However, these terms have not provided a complete differentiation between different distributed representations.

There was a discussion on the advantages and disadvantages of local representation vs. distributed representation at Connectionists in 1991. Near the end of that discussion, Bo Xu pointed out in a long dialogue with Ali Minai that the confusion probably came from the vague and inaccurate definition of *distributed representation*. This paper is a further development of our presentation there. In this paper, we first pointed out that there exists a hierarchical organization of entity to be represented by the neural networks. This hierarchical organization leads to four levels of correspondences between the entity, patterns, elements, items and the units. Then we classified the representations from two different viewpoints: correspondence viewpoint and location viewpoint. The first viewpoint is according to the correspondences between the entity, patterns, elements and items to be represented by the units and the units on an appropriate layer of a neural net. From the correspondence viewpoint, representations can be classified into five categories: (1) LR, (2) one-to-one DR, (3) one-to-many DR, (4) many-to-one DR, and (5) many-to-many DR. The second viewpoint of classifying representations is according to the locations where the representations exist. From the location viewpoint, representations may fall into two classes: (1) external representation, and (2) internal representation. Finally, it was pointed out that the origins of the confusion on representations may come from: (1) non-distinction of the hierarchical organization of entity, (2) non-distinction of the four levels of correspondences between the entity, patterns, elements, items and the units, (3) use of the too general or too broad term *distributed representation*, (4) non-distinction of the two levels of graceful degradation, and (5) non-

distinction of similarities between different representations.

2. HIERARCHICAL RELATIONSHIP BETWEEN ENTITY, PATTERNS, ELEMENTS, AND ITEMS

In the original definition of local and distributed representations, the term "entity" was used to refer to the object to be represented by the neural net. The key point underlying the classifications of different kinds of representations is the correspondences between the entity and the units of the net.

However, since an entity may have many patterns, each pattern is usually composed of many elements, and each element may contain many items (i.e. there is a hierarchical structure within the entity), it is not clear to define representations simply as correspondences from the entity to the units. It is necessary to differentiate the entity from its constituents (patterns, elements, and items) and make clear at what level the correspondence is.

In order to make these things clear, we first define the terms within the hierarchical organization of entity as follows:

Entity ---- a collective concept representing a group of objects to be represented by the neural net.

Pattern ---- a concrete individual object belonging to the entity to be represented by the neural net. Usually, an entity may contain many patterns.

Element ---- a structural component or property of a pattern. Usually, a pattern can be decomposed into many elements based on properties the pattern may have. Each property contains a group of constituents (items) *belonging to the same property. Each element corresponds to one of these property groups.*

Item ---- a possible instance or occurrence an element may have. An element usually contains two (binary) or more (analog) items. The items are therefore represented by binary or analog values. Each value is a measurement of an item.

Following, for simplicity, we will use interface layer to refer to either input layer and/or output layer on which the units may reside. As will be pointed out later, we will mainly consider the external reperesentation. Therefore, unless otherwise pointed out, the layer will be an interface layer. For clarity, we will adopt the following notations:

P ---- the total number of *patterns* an entity may have;

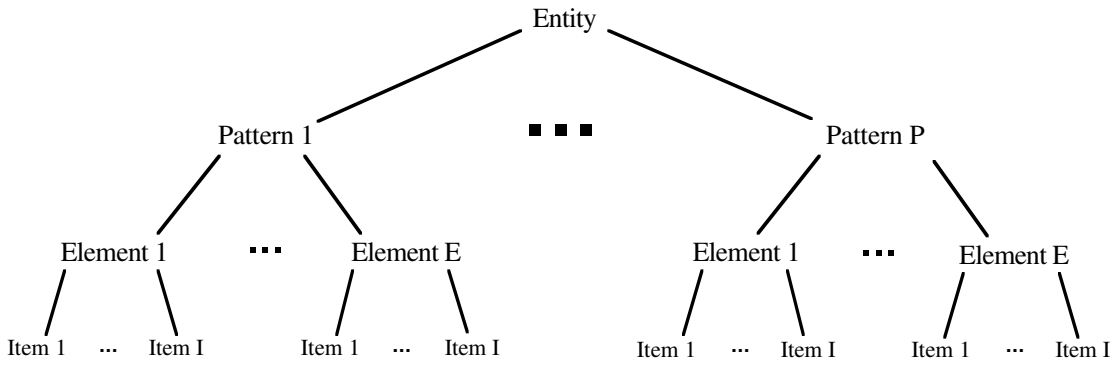
E ---- the total number of *elements* a pattern may contain (for simplicity, we assume that the E is the same for all patterns);

I ---- the total number of *items* an element may contain (for simplicity, we assume that the I is the same for all elements);

I_u ---- the total number of *items* a *unit* may have (it is also the total number of analog values a unit may take);

U ---- the total number of *units* on an appropriate layer of the neural net considered.

Thus, the hierarchical relationship can be illustrated as follows:



Usually, an entity may have many patterns, a pattern can be decomposed into elements in many different ways, and an element may contain many items. The hierarchical relationship between the entity, patterns, elements, and items is very important to representations in all neural network applications. Following we will use examples in several important application fields to illustrate the hierarchical relationship between the entity, patterns, elements, and items.

2.1. Picture (Pattern, Image) Representation

Suppose we want to represent a group of pictures (In order to avoid confusion between the *pattern* defined above and the term pattern used in pattern recognition, we will use picture instead of pattern to stand for the latter). The collective concept "picture" now is the *entity* defined above. All the concrete individual pictures in this group of pictures considered are the *patterns* defined above. Each picture (*pattern*) can be decomposed into many pixels (m x n or whatsoever). These pixels are the *elements* defined above. The different intensities each pixel may have are *items* of the piexls (elements) defined above.

2.2. Word Reading Representation

Reading texts is a very important field of neural network applications. In order for a neural net to read a text, one should first decide how to represent the words in the text by means of the input units of the network. Now the collective concept "word" is the *entity* defined above. All the individual words in the text to be read are the *patterns* defined above. Each word is composed of many characters. Therefore a word can be decomposed into many positions of characters (one characher is placed at each of these positions). These positions of characters are the *elements* defined above. All the possible characters which may be put into the positions (e.g. the 26 letters and the punctuation) are *items* defined above.

2.3. Spatial Point Representation

Many problems can be solved in a n-dimensional (n-dim) space. Because a curve in a n-dim space can be regarded (sampled) as a set of spatial points in the same space, the representation of a set of spatial points is the first thing to be solved for these problems. We discuss two situations here: static curve and dynamic curve.

(1). Static Curve

If there is only one curve in the n-dim space and it is not a function of time, the curve can be viewed as a set of I static points in the space. In order to represent this curve or a set of I static points in the n-dim space, the collective concept "point" is the *entity* defined above. All the I points contained within this set are the *patterns* defined above. Each point can be decomposed into n coordinates in the n-dim space. These n coordinates are the *elements* defined above. Each coordinate (element) may take I different analog values corresponding to the coordinate's values of the I points. The I analog values each coordinate (element) may take are the *items* defined above.

(2). Dynamic Curve

If there are a group of I curves related to each other according to time (now these curves can be viewed as a series of time-varying curves) or some other serial order, each curve may be viewed as a set of N points in the n-dim space (sampling each curve by N points), and the I curves can be viewed as a series of I N-point-sets. In order to represent the I curves (N-point-sets) in a n-dim space, the collective concept "curve" ("N-point-set") is the *entity* defined above. All the individual curves in the group of I curves are the *patterns* defined above. Each N-point-set (a sampled curve) can be decomposed into n*N coordinates in the n-dim space. These n*N coordinates are the *elements* defined above. Each coordinate (element) may take I different analog values corresponding to the coordinate's values of the I curves. These I analog values each coordinate (element) may take are the *items* defined above.

2.4. Jets and Sharks Model

Jets and Sharks model is a model McClelland developed to illustrate how a system with multi-properties would act as a content-addressable memory (McClelland, 1981). In order to represent the peoples listed in this model, the collective concept "people" is the *entity* defined above. All the individual persons of the two gangs (Jets and Sharks) are the *patterns* defined above. All the properties (Name, Gang, Age, Edu, Mar, Occupation) each person (pattern) may have are the *elements* defined above. All the possible entries for the properties of these persons are the *items* defined above.

2.5. Apple Representation

This is an example Ali Minai raised during our dialogue on Connectionists. Apple is taken only as an example. The conclusion here applys to orange, pear, or any similar objects of interest. (In fact, this example is very similar to the Jets and Sharks model discussed above.)

If we have a bunch of apples, and we want to use neural nets to represent them, the collective concept "apple" is the *entity* defined above. All the concrete individual apples under consideration are the *patterns* defined above. Each apple may have many properties: size, weight, color, taste, shape, or any other properties of interests. We can decompose a pattern (an apple) into many elements. Each element corresponds to one of these properties. These properties are the *elements* defined above. Each property can take many values according to its own measurement. All the possible values a property may take are the *items* defined above.

Above we illustrate the hierarchical relationship between entity, patterns, elements, and items in several important applications. Here we only pointed out one possible hierarchical structure. In fact, in all applications, there may be many ways to decompose a pattern into different elements. Generally speaking, there are many kinds of relationships between a pattern and its elements. Which relationship is chosen to decompose a pattern into elements in an application depends on the pattern's properties people are interested in (in applications, people usually will be interested in only some of the properties, not all of them). However, in spite of these variations, the hierarchical organization of an entity always exists.

3. CORRESPONDENCES

According to Marr's definition, the key point in the representation of an entity is the description of the entity by a formal system (e.g. the external layers of neural networks). The description is in fact a mapping or correspondence process from the entity space to the space of the formal system. In other words, the mapping or correspondence between the entity and the formal system forms the core of representations. (Following we will use *correspondence* instead of *mapping* to describe the relationships between an entity and the formal system in order to avoid confusion with the *mapping*

between the input space and output space of the neural network.) Therefore, by analyzing the correspondences between an entity and a formal system, we will be able to get a clear look at the similarities and differences between various representations.

3.1. Four Levels of Correspondences

Because of the hierarchical organization of the entity, there are four levels of correspondences between the entity, patterns, elements, or items and the units:

(1). Entity-to-unit correspondence ---- correspondence between the entity to be represented by the neural net and the units of the neural net.
(2). Pattern-to-unit correspondence ---- correspondence between the patterns of the entity and the units of the neural net;
(3). Element-to-unit correspondence ---- correspondence between the elements of a pattern and the units of the neural net.
(4). Item-to-unit correspondence ---- correspondence between the items of an element and the units of the neural net.

Theoretically speaking, representations can be viewed from all four levels. However, due to the hierarchical relationship between the entity, patterns, elements, and items, there are overlappings between correspondences at the four levels. Therefore, we need to decide which level can best characterize a representation.

Usually, a lower level constituent contains more detailed information than a higher level constituent. From this consideration, the level priority of information detailedness should be:

Entity < Pattern < Element < Item

However, besides the information detailedness, there is another consideration should be taken into account of: the difference between the structural component and non-structural component. To show this difference, we first define the *structural* component as follows:

Structural component ---- a constituent of the pattern that will be present in all patterns. It is like a structural frame that is necessary to all patterns.

According to this definition, we can see the difference between the element and the item. An *element* is a structural component because it appears in all patterns. However, an *item* is not a structural component of the pattern because it is only a possible instance or occurrence of the corresponding element. An item will not occur in all patterns. It will be present in only one or some patterns. Therefore, it is the element (not the item) that is the smallest structural constituent of the entity. A structural component contains more information than a non-structural component because the structural component has grouped many items together based on the property which is common to these items.

Combining above two considerations, it is the element-to-unit level of correspondence that can best characterize representations. We will classify various representations at this level. In the special case (see LR below) where a pattern contain only one element, the element level degenerates to the pattern level. In order to contrast with the LR people usually refer to, we discuss LR at pattern-to-unit level for this special case.

One problem associated with the item level correspondence can be seen from the binding problem (Hinton, McClelland, and Rumelhart, 1986). The representation used in the binding problem is in fact an item-to-unit one-to-one correspondence (i.e. an item local representation). This can be seen more clearly if we notice that the problem here is a special case of the n-dim spatial point problem of the dynamic curve discussed above with n=2 and N=2 . The binding problem can be overcome if we use the higher structural level correspondence (element-to-unit correspondence) for the representation. So the binding problem also favors the element level correspondence to the item level correspondece.

3.2. Four Types of Correspondences

At each level of correspondence, there are generally four different types of correspondences between the entity, patterns, elements, or items and the units:

(1). one-to-one correspondence ---- the correspondence pairing each entity, pattern, element, or item uniquely with each unit;
(2). one-to-many correspondence ---- the correspondence pairing one entity, pattern, element or item with many units;
(3). many-to-one correspondence ---- the correspondence pairing many entities, patterns, elements, or items with one unit;
(4). many-to-many correspondence ---- the correspondence pairing many entities, patterns, elements, or items with many units.

Theoretically speaking, there are sixteen correspondences all together at the four levels. As we pointed out above, due to the overlapping between the four levels, we will only concentrate on the element-to-unit (for LR, it is the pattern-to-unit) level mainly.

With these correspondences, we will be able to discuss various representations more specifically.

4. CORRESPONDENCE CLASSIFICATION

4.1. Local Representation (LR)

The original LR defined by (Barlow, 1972; Feldman, 1986) is the pattern-to-unit one-to-one correspondence defined above.

LR is best characterized by the pattern-to-unit correspondence because in LR the pattern is the smallest constituent of the entity. A pattern is not further divided into more elements (each pattern contains only one element). Therefore, only one unit taking binary values is needed to represent a pattern. For a total of P patterns, U=P binary units are needed (Fig. 1).

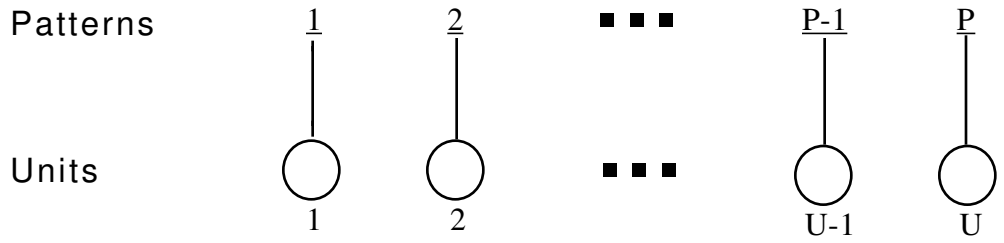


Fig. 1. Local representation (LR). P ---- the total number of patterns. U ---- the total number of units. There is no element level in between the patterns and the units in LR. Because now the pattern is the smallest constituent of the entity, the LR can be best described by the pattern-to-unit correspondence. Each pattern corresponds directly to each unit. Therefore U=P. Each unit takes either 0 or 1 binary values.

So LR is characterized by the pattern-to-unit one-to-one correspondence. Each time when a pattern is presented, only one of the U units is 1, and all the others are 0. Which unit should be 1 is determined by the corresponding pattern. Hence, we can define the LR as follows:

Local representation ---- the pattern-to-unit one-to-one correspondence in which each pattern is represented by one unit, and each unit represents only one pattern. Each unit here always takes

binary values only because it only represents whether the pattern it corresponds is present or not.

If we change our viewpoint from the pattern-to-unit correspondence to the higher level entity-to-unit correspondence, LR is an entity-to-unit one-to-many correspondence. However, as we pointed out above, because entity usually is not the smallest constituent, we will not discuss representations at the entity level.

The output representation of the binary-to-local problem (Jacobs, 1988) is an example of LR.

4.2. Distributed Representations (DRs)

In distributed representations, a pattern usually can be decomposed into many elements. The element now is the smallest structural constituent of the entity. Therefore, it is convenient to characterize the distributed representations at the element-to-unit level. So we will discuss the four element-to-unit correspondences for DRs.

Because all the four correspondences at the element-to-unit level belong to distributed representations, the original term *distributed representation (DR)* is probably a too broad or too general concept. It is not clear to use a single term *distributed representation* to refer to the four situations.

Following we will give specific definitions for all the four distributed representations (DRs). Because all of the four DRs are at the element-to-unit level, we will omit the "element-to-unit" in the names of these DRs.

4.2.1. One-to-one DR

In one-to-one DR, a pattern is decomposed into E elements. Each element is represented by one unit. So there is a one-to-one correspondence at the element level between the E elements in the pattern and the U units in the network. Therefore, E=U (Fig. 2).

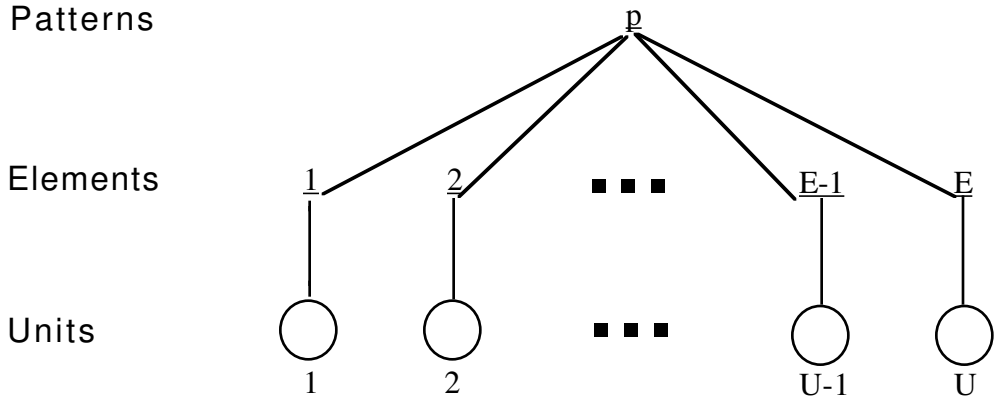


Fig. 2. One-to-one DR. P ---- the total number of patterns (p = 1, 2, ..., P); E ---- the total number of elements contained within a pattern; I ---- the total number of items contained within an element; U ---- the total number of units. Because now the element is the smallest structural constituent of the entity, the representation can be best described by the element-to-unit correspondence. There is a one-to-one correspondence between each element of a pattern and each unit. Therefore E=U. All the P patterns use the same set of U units. Each unit takes I_u=I analog values.

All the P patterns use the same set of U units. Each unit takes I analog values to represent the I different items ($I \leq P$). Since different patterns are represented by the combinations of different items of their elements, the number of the items I can be infinite when P tends to infinity (Xu and

Zheng, 1992).

So from the correspondence viewpoint, we can define the one-to-one DR as follows:

One-to-one DR ---- the element-to-unit one-to-one correspondence in which each of the E elements of the pattern is represented by one unit, and each unit represents only this element. The units take I_u analog values (including binary values) which represent the measurements of the I items contained in the corresponding element.

During the discussion at Connectionists, we called this representation "analog-distributed representation". In order to show the correspondence relationship more clearly, we change its name to "one-to-one DR" here.

If we shift our viewpoint from the element level to the higher pattern level, one-to-one DR is a pattern-to-unit many-to-many correspondence because (1) each pattern is represented by many units; (2) each unit participates in representing many patterns.

The output representation of the XOR problem can be regarded as the simplest example of the one-to-one DR with E=1, U=1, and binary units (I_u=I=2).

The input representation of the XOR problem can be regarded as an example of the one-to-one DR with E=2, U=2, and binary units (I_u=I=2). Therefore $P=I^U=4$ (Xu and Zheng, 1992).

The 8-bit byte representation for 256 characters (Anderson and Rosenfeld, 1988) is a familiar example of the one-to-one DR with E=U=8, binary units ($I=I_u=2$), and $P_{max}=I^E=256$ (Xu and Zheng, 1992).

The input representation in the binary-to-local problem (Jacobs, 1988) is another example of the one-to-one DR with E=U=3, binary units ($I=I_u=2$), and $P_{max}=I^E=8$.

The output representation of NETtalk (Sejnowski and Rosenberg, 1986) can be regarded as an example of the one-to-one DR with E=26, U=26, binary units (I_u=I=2), and $P_{max}=I^E\approx 6.7\times 10^7$. This P_{max} is large enough to guarantee that the NETtalk can generate most of the phonemes and punctuations that may encounter in reading texts.

The input representation and output representation in the parabolic problem (Xu and Zheng, 1991) are analog examples (units take analog values instead of binary ones) of the one-to-one DR with I_u=I=101 (for training) or 201 (for test), E=1, and U=1. Therefore, $P=I^U=101$ (for training) or 201 (for test). There is one input unit representing x values and one output unit representing y values of the parabola. Both the input unit and the output unit may take infinite analog values corresponding to the infinite real x and y values.

Because the analog one-to-one DR does not work well on BPNN (it cannot converge for many problems on BPNN), it was not widely used in the current neural network applications. However, we will show that this representation is one of the best representations for many practical applications (Zheng and Xu, 1992).

4.2.2. One-to-many DR

In one-to-many DR, the pattern is divided into E elements. Each element is represented by a subgroup of I units. So there are E subgroups of units. Each subgroup contains I units. Therefore there are a total of U=E*I units in the network for this representation (Fig. 3).

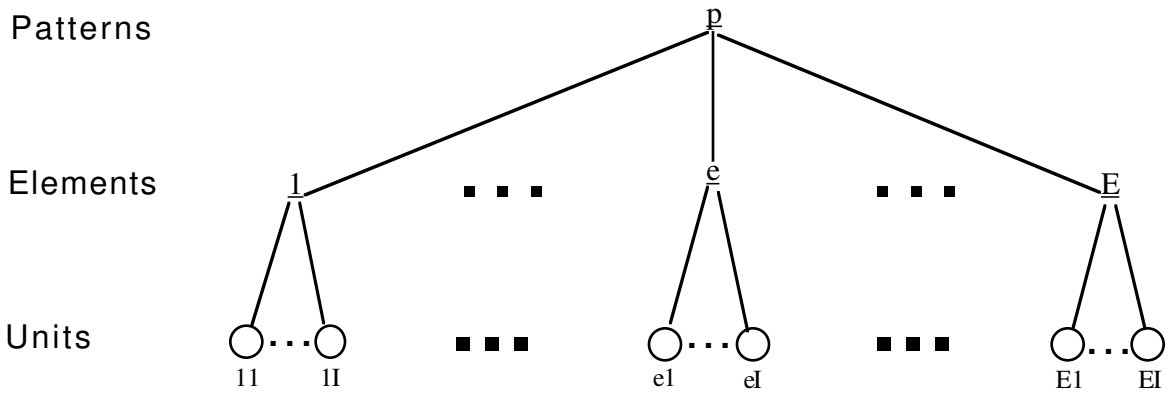


Fig.3. One-to-many DR. P ---- the total number of patterns (p = 1, 2, ..., P); E ---- the total number of elements contained within a pattern (e = 1, 2, ..., E); I ---- the total number of items an element may have; U ---- the total number of units (U=E*I). Because now the element is the smallest structural constituent of the entity, the representation can be best described by the element-to-unit correspondence. All the P patterns use the same set of U units. Each unit takes either 0 or 1 binary values. Therefore I_u=2. Only one unit in each of the E subgroups of units takes the value 1; all other units take the value 0.

All the P patterns use the same set of U units. Each unit takes only binary values because each unit only represents the presence of one of the I items of the element. Only one unit in each of the E subgroups of units takes the value 1. Which unit in each subgroup should take 1 is determined by the corresponding item existed.

In this sense, the one-to-many DR is a kind of item-local representation (each item is locally represented by one binary unit). To avoid confusion with the pattern local representation (which is the local representation people usually refer to), we prefer the (element-to-unit) one-to-many DR to the (item-) local representation,.

So according to the correspondence between the elements of the pattern and the units, we define the one-to-many DR as follows:

One-to-many DR ---- the element-to-unit one-to-many correspondence in which each element is represented by many (a subgroup of I) units and each unit represents only one of the I items of the element. Units here take only binary values because each unit only represents the existence of one item. Each subgroup contains I units for the I items in the corresponding element. There are E subgroups of units for E element.

During the discussion at Connectionists, we called this representation "binary-distributed representation". In order to show the correspondence relationship more clearly, we change its name to "one-to-many DR" here.

From the pattern-to-unit correspondence viewpoint, one-to-many DR is a many-to-many correspondence because each pattern is represented by many units, and each unit participates in representing many patterns.

The input representation in NETtalk (Sejnowski and Rosenberg, 1986) is an example of the one-to-many DR with E=7, I=29, U=E*I=203, and binary units. The entity in NETtalk is the seven-character window of text. The patterns are all possible strings of seven characters. E=7 because each pattern is decomposed into seven elements (positions of characters). I=29 because each position of character has 29 possible items (26 letters and 3 punctuation and word boundaries) and is represented by 29 units (Fig. 3). The maximum number of words (strings of seven characters, or patterns defined in this paper) which can be represented and read by NETtalk is: $P_{max} = I^E = 29^7 \approx 1.7 \times 10^{10}$ (Xu and Zheng, 1992). This is large enough for NETtalk to read all existing text strings of seven characters. So one-to-many DR is a safe representation (i.e. with

enough capacity to represent all possible words) for text reading applications of neural networks. However, we will show that the one-to-many DR is not a good representation for problems with continuous and infinite patterns.

4.2.3. Many-to-one DR

In many-to-one DR, a pattern is decomposed into E elements. A group of g elements are represented by one unit, and each unit represents these g elements only. So there is a many-to-one correspondence at the element level (Fig. 4).

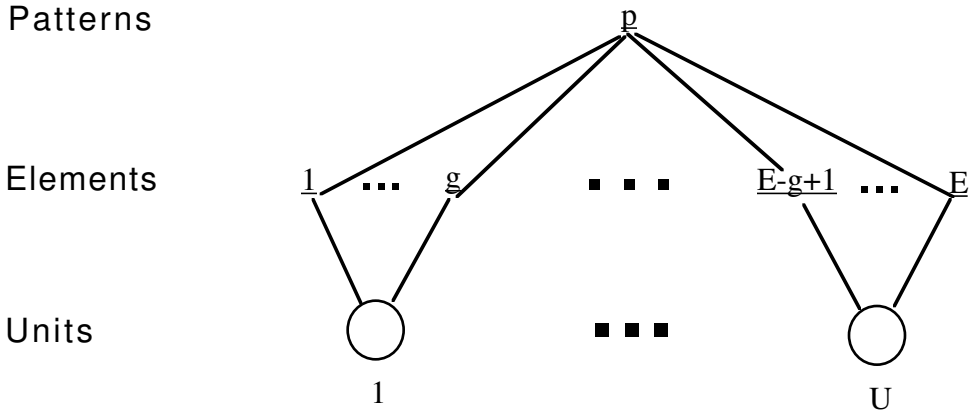


Fig. 4. Many-to-one DR. P ---- the total number of patterns (p = 1, 2, ..., P); E ---- the total number of elements contained within a pattern; g ---- the number of elements in one group corresponding to a unit (for simplicity, we assume that the g is the same for all groups); U ---- the total number of units. So E = U*g. Because now the element is the smallest structural constituent of the entity, the representation can be best described by the element-to-unit correspondence. All the P patterns use the same set of U units. Each unit takes I_u analog values representing different combinations of items of the g corresponding elements. Therefore, $I_u = I^g$

All the P patterns use the same set of U units. Each unit takes I_u analog values to represent the I_u different combinations of items the g corresponding elements may have. As the I in the one-to-one DR, I_u here can be infinite too because $I_u=I^g$.

Thus, from the correspondence viewpoint, we can define the many-to-one DR as follows:

Many-to-one DR ---- the element-to-unit many-to-one correspondence in which g elements are represented by one unit and each unit represents these g elements only. Each unit here takes I_u analog values (including binary values) to represent different combinations of items in the g corresponding elements.

If we regard the g elements corresponding to the same unit as one big element, the many-to-one DR can be reduced to the one-to-one DR. It is probably due to this reason, the many-to-one DR has not been used in current neural network applications. However, this representation may be useful to some applications because of its simplicity (it uses fewer units) or higher capacity (Xu and Zheng, 1992).

4.2.4. Many-to-many DR

In many-to-many DR, the pattern is divided into E elements. Each element is represented by U units and each unit participates in representing all E elements. There is no definite relationship between E and U here (Fig. 5).

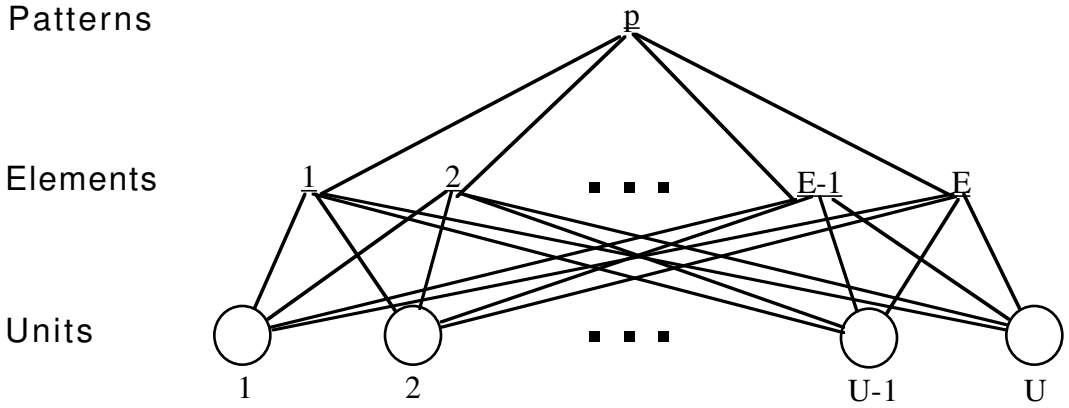


Fig. 5. Many-to-many DR. P ---- the total number of patterns (p = 1, 2, ..., P); E ---- the total number of elements contained within a pattern; U ---- the total number of units. Because now the element is the smallest structural constituent of the entity, the representation can be best described by the element-to-unit correspondence. There is a many-to-many correspondence between the elements and the units. The figure shows the maximum connections (full connections) between elements and units. There may exist intermediate connections (non-full connections) between the elements and units. All the P patterns use the same set of U units.

All the P patterns use the same set of U units. Each of the E elements is represented by U units, and each of the U units represents the E elements. This is an extreme situation (full connections between the elements and units) of the many-to-many DR. By varying the degree of mixing, one may obtain many variations of many-to-many DR with intermediate (non-full) connections. However, all of them share the same characteristics as many-to-many DR.

So from the element-to-unit correspondence, we can define the many-to-many DR as follows:

Many-to-many DR ---- the element-to-unit many-to-many correspondence in which each element is represented by many units and each unit represents many elements. The units may take I_u *analog (including binary) values which represent* I_u *possible combinations of items of the many corresponding elements.*

During the discussion at Connectionists, we called this representation "mixed-distributed representation". In order to show the correspondence relationship more clearly, we change its name to "many-to-many DR" here.

From the pattern-to-unit correspondence viewpoint, the many-to-many DR is a pattern-to-unit many-to-many correspondence. However, because there are overlappings between pattern-to-unit and element-to-unit correspondence, and the pattern now is not the smallest structural constituent of the entity, we will keep on the viewpoint of element-to-unit correspondence.

In fact, all the internal representations (see below) in most current neural networks are examples of many-to-many DR. The units here usually take analog values. However, as we will discuss later, these internal many-to-many DRs cannot be arbitrarily chosen by us. They are determined by the topology of the network. The working mechanism of the internal representations is still a black-box. Therefore, we will not discuss internal representations here and will concentrate on external representations mainly.

Despite the degeneration of representations (Xu and Zheng, 1992), the four DRs are different representations in essence. Because the four distributed representations discussed above are very different in nature, it will be unclear if we call all of them simply *distributed representation*. In other words, the term *distributed representation* is a vague concept. In order to compare the advantages and disadvantages of local and distributed representations, it is needed to point out

specifically which distributed representation is considered because the four distributed representations have different advantages and disadvantages. A general term *distributed representation* is unclear and not enough.

5. LOCATION CLASSIFICATION

Correspondence classification is from the viewpoint of how the patterns or elements of an entity are represented by the units of the neural net. Because the units can be on the interface layers (input and/or output layers) as well as on hidden layers, it is needed to differentiate representations from the location viewpoint.

The key point underlying the classification of representations from the location viewpoint is where the representations exist. From this location viewpoint, representations fall into two classes:

(1). Internal representations ---- representations existed at the hidden layers. The internal representations are responsible for encoding the mappings from the input space to the output space of a neural network. For internal representations, correspondences are between the units on an interface layer and the hidden layer nearest to the interface layer, or between the units on two adjacent hidden layers.

The internal representation is not a new concept. It had been mentioned in (Marr, 1982; Rumelhart, Hinton, and Williams, 1986) and other places.

(2). External representations ---- representations existed at the interface layers (input and/or output layers). The external representations are responsible for the information transmission between a neural network and the outside world (coding the input information at the input layer and decoding the output information at the output layer). For external representations, correspondences are between the external entity, patterns, elements, or items and the units on the interface layers.

Because the correspondence classification and the location classification are based on different viewpoints, there are overlappings between these two kinds of classifications.

One difference between the internal representation and the external representation is that there exists a signal transformation (or modification) by weights in the internal representation: each signal in the internal representation is multiplied by the weight in the connection (correspondence route) it passes. However, in the external representation, each signal passes its correspondence route unchanged because there is no weight in the connection of the external correspondences.

The internal representations are very important representations because the mapping from the input space to the output space is the core of neural nets. However, the internal representations are determined by the topology of the network, and we cannot change them unless we change the network topology. Most of the current networks' topology ensure that the internal representations are many-to-many DRs. The detailed working mechanism of the internal representation is still a black-box. Without changing the topology of the network, what we can choose and select are different external representations. Therefore, unless pointed out separately, we only consider external representations here.

With the correspondence and the location classifications, we will be able to talk about representations more specifically as to “what kind of representational scheme it is” and “where it locates”.

6. DISCUSSIONS

Above we classified and defined various representations. Following we try to analyze the possible causes or sources leading to current confusion on representations.

6.1. The Hierarchical Organization of Entity

The first source leading to the confusion on representations may come from the hierarchical relationship between the entity, patterns, elements, and items. In current literature, terms such as concept, object, entity, pattern, item and others are used to stand for the *entity* or *pattern* defined in this paper. The hierarchical organization of the *entity* was not noticed. These terms sometime refer to the entity defined in this paper, some other time refer to patterns, elements, or even items defined in this paper. There is no clear distinction between the four constituents of entity. The mixing of the four levels of constituents (entity, pattern, element, and item) can easily cause confusion (people may be talking about different constituents at different levels while using the same term). Therefore, the non-distinction of the hierarchical organization of entity is probably one of the origins causing the current confusion on representations.

6.2. The Four Levels of Correspondences

The non-distinction of the hierarchical relationship between the entity, patterns, elements, and items will lead to the non-distinction of the four levels of correspondences: entity-to-unit, pattern-to-unit, element-to-unit, and item-to-unit. As we pointed out earlier, the four levels of correspondences are different from each other, and there are overlappings among them. Without sticking to a given level of correspondence, one may overlap and mix these correspondences and get confused. The non-distinction of the four levels of correspondences is probably another origin of the confusion on representations.

6.3. The Vague Term *Distributed Representation*

As we pointed out above, there are at least four kinds of DRs. These four DRs are very different from each other in nature. However, in current literature, they are all called *distributed representation.* It is not clear to use one term to refer to four different situations. In other words, the too general or too broad term *distributed representation* is probably another source leading to the confusion on representations.

6.4. The Two Levels Of Graceful Degradation

The graceful degradation was one of the criteria proposed in the discussion on the advantages vs. disadvantages of representations. It was pointed out that there are two levels of graceful degradation (Zheng and Xu, 1992).

The non-distinction between the two levels of graceful degradation may lead to an illusion that some DRs (like the LR) do not have graceful degradation. However, we should be more specific as to which graceful degradation we are referring to. Otherwise, it is easy to get confused by a too general conclusion. It was shown that (1) the LR has neither pattern- nor element-graceful degradation; (2) although some DRs do not have element-graceful degradation, all DRs have pattern-graceful degradation.

So one possible source leading to the confusion may come from the non-distinction of the two levels of graceful degradation.

6.5. Similarities Between Representations

6.5.1. NETtalk Input Representation

One souce of confusion may arise from the similarity between the LR (which has a pattern-to-

unit one-to-one correspondence) and the one-to-many DR (which has an element-to-unit one-to-many but item-to-unit one-to-one corresondence). In both cases, the unit takes binary values: 0 or 1. They both have an appearance of local representation. However, they are different representations.

The pattern-to-unit one-to-one correspondence is the local representation people usually refer to. Each binary unit in it corresponds to each pattern. So it is in fact a pattern-local representation.

While the element-to-unit one-to-many correspondence is in fact a distributed representation because each pattern is represented by many units here. Each item contained within an element is represented by one unit, and each unit represent only that item contained within the element. The binary values each unit may take represent only the existence of the item corresponding to this unit. So it is in fact an item-local (but pattern-distributed) representation.

For example, the input representation of NETtalk was regarded as a LR (Sejnowski and Rosenberg, 1986; Anderson and Rosenfeld, 1988). However, according to above discussion, we see that the input representation of NETtalk is in fact a one-to-many DR (i.e. item-local but pattern-distributed representation) instead of a LR (i.e. pattern-local representation).

6.5.2. Binding Problem Representation

Another example of confusion due to the similarity between different representations is the binding problem (Hinton, McClelland, and Rumelhart, 1986). As we pointed out before, the representation used in the binding problem is in fact an item-to-unit one-to-one correspondence (i.e. an item-local representation). However, it was regarded as LR. Similar to the situation in above example, the two representations here, LR and item-to-unit one-to-one correspondence, are very similar from the one-to-one correspondence point of view. But they are at different levels. LR is at the pattern level, and the item-to-unit one-to-one correspondence is at the item level (The latter is one of the sixteen correspondences theoretically existed at the four levels). As we mentioned before, because the item is not a structural component of the entity, it is not suggested to view representations at the item level (It is suggested to view representations at the element level). The binding problem may be avoided if we change our viewpoint from the item level to the element level.

6.5.3. Many-to-many DR and Distributed Representation

It was believed during the Connectionists discussion that the term *distributed representation* most people currently refer to is the many-to-many DR (it was called mixed-distributed representation during that discussion). However, as we analyzed in this paper, this viewpoint may be misled. In fact, except the internal representations, most distributed representations currently referred to fall into the one-to-one DR (analog-distributed representation) or one-to-many DR (binary-distributed representation). This mixing of the distributed representation with the many-to-many DR may be another cause leading to the confusion on representations.

In sum, the similarity between different representations is probably another source leading to the confusion on the advantages vs. disadvantages of different representations (people may be talking about different representations while using the same term).

7. CONCLUSIONS

In this paper, we first pointed out that there exists a hierarchical organization within an entity to be represented by the neural networks. This hierarchical organization leads to four levels of correspondences between the entity, patterns, elements, items and the units. Then we classified representations further into more detailed and more specific categories from two viewpoints:

- From the correspondence viewpoint, representations were classified into five categories:

(1) LR, (2) one-to-one DR, (3) one-to-many DR, (4) many-to-one DR, and (5) many-to-many DR;
- From the location viewpoint, representations were classified into two types: (1) internal representation, and (2) external representation.

Finally, we pointed out that the sources of the confusion on representations may come from: (1) non-distinction of the hierarchical organization of entity, (2) non-distinction of the four levels of correspondences between the entity, patterns, elements, items and the units, (3) use of the too general or too broad term *distributed representation*, (4) non-distinction of the two levels of graceful degradation, and (5) non-distinction of the similarity between different representations.

REFERENCES

Anderson, J. A., and Rosenfeld, E. (1988). *Neurocomputing: Foundations of Research.* Cambridge, MA: MIT Press.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception,* **1,** 371-394.

Feldman, J. A. (1986). Neural representation of conceptual knowledge. Technical Report TR 189, Department of Computer Science, University of Rochester, Rochester, NY.

Hinton, G.E. (1981). Implementing semantic networks in parallel hardware. In Hinton, G.E., and Anderson, J.A. (Eds.), *Parallel Models of Associative Memory.* Hillsdale, NJ: Erlbaum.

Hinton, G. E., McClelland, J. L., and Rumelhart, D. E. (1986). Distributed representations. In Rumelhart, D. E., McClelland, J. L., and the PDP Research Group, Editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* Vol. 1, Foundations, Cambridge, MA: MIT Press.

Jacobs, R. A. (1988). Increased rates of convergence through learning rate adaptation. *Neural Networks,* **1,** 295-307.

Lippmann, R. P. (1987). An introduction to computing with neural nets. *IEEE Acoustics Speech Signal Processing Magazine,* **4,** 4-22.

McClelland, J.L. (1981). Retrieving general and specific knowledge from stored knowledge of specifics. *Proceedings of the Third Annual Conference of the Cognitive Science Society,* Berkeley, CA.

Marr, D. (1982). *Vision.* San Francisco: W.H.Freeman.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning internal representations by error propagation. In Rumelhart, D. E., McClelland, J. L., and the PDP Research Group, (Editors), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* Vol. 1, Foundations, Cambridge, MA: MIT Press.

Sejnowski, T. J., and Koch, C., and Churchland, P. S. (1990). Computational neuroscience. In Hanson, S. J. and Olson, C. R. (Eds). *Connectionist modeling and brain function: the developing interface.* A Bradford Book, Cambridge, Massachusetts: The MIT Press.

Sejnowski, T. J., and Rosenberg, C. R. (1986). NETtalk: a parallel network that learns to read aloud. The Johns Hopkins University Electrical Engineering and Computer Science Technical Report JHU/EECS-86/01, 32pp.

Xu, B., and Zheng, L. (1991). PPNN: A faster learning and better generalizing neural net. *Proceedings of IJCNN,* Singapore.

Xu, B., and Zheng, L. (1992). Capacity of representations. Submitted to *The Fourteenth Annual Conference of the Cognitive Science Society,* Bloomington, Indiana.

Zheng, L., and Xu, B. (1992). Evaluation of representations. Submitted to *The Fourteenth Annual Conference of the Cognitive Science Society*, Bloomington, Indiana.