Perception in NARS

Pei Wang

January 15, 2018

[This Technical Report summarizes my ideas on this topic beyond the previous Open-NARS Wiki page.]

1. Knowledge Representation

In NARS, knowledge involved in sensation and perception is also represented in Narsese like other types of knowledge, though there are special compound terms introduced for temporal-spatial patterns.

NARS supports multiple input/output channels. Besides the *primary* channels that directly recognize Narsese tasks, there can also be multiple *sensorimotor* channels, each dedicated to a special type of sensor/actuator or a few types of related sensor/actuator. Within the system, there is also an overall channel (the previous *new task buffer*) that integrates significant events from all other channels.

NARS treats sensation as a special type of operation, so each sensor is often also an actuator or related to one. Each sensor can be triggered from the inside (via operation execution) or outside of NARS to receive a type of signal (which can be physical, chemical, biological, etc.), and the result is an event "{T} → [V]", where T is a **sensory term** representing the sensation, V the type of the sensation (which can just be the identifier of the channel), and the truth-value the strength of the sensation (with two factors for *frequency* and *confidence*, respectively).

Perception is the process where relations are derived among the sensory terms, as well as between them and the other (generic) terms. Beside the semantic relations provided by the copulas and the syntactic relations by the term connectors, there are also *temporal-spatial* relations directly coming from the input channels.

The representation of temporal knowledge has been specified in NAL-7, and mostly implemented in the code. Temporal perception happens in every channel, and the results are temporal judgments and compound events. Temporal experience in each channel is a one-dimensional stream with direction and without bound in length. It can be represented by the term connector **list** (#), as specified in the Tech Report on NLP (PAGI-TR-6).

Spatial experience is initiated by multiple sensors of the same type. A new type of compound term '**array**' is introduced specially for spatial experience. Like in many programming languages, an array has a fixed number of components in fixed dimensions. In the near future, in NARS an array can be 1-dimentional ('vector'), 2-dimentional ('matrix'), or 3-dimentional ('space'). The familiar format A[i, j, k] will be used to indicate a component in array A, and all indices are positive integers. When there are multiple dimensions, their order is purely conventional, so there is no 'primary' dimension, and all sub-arrays in the same dimension will have the same length.

When implemented, each spatial sensation is an array of NAL truth-values with the maximum resolution of the corresponding channel. For example, after every observation a channel for brightness may produce a 1024-by-1024 matrix T, and each ‘pixel’ at location [i, j] is a truth-value of the event {T[i, j]}$\rightarrow$ [bright] (similar to “{T} $\rightarrow$ [V]” specified previously), and the array (matrix) T as a whole represents the co-occurrence (*parallel conjunction*) of these events. At each pixel, the *frequency* value indicates the level of brightness at that sensor (which can be approximately the quantile among all signals produced by that sensor), and the *confidence* indicates the reliability of the sensation, so as to summarize factors like noise, accuracy, attention, etc. The *confidence* at different indices of the same matrix can be different, and some of them can even be 0, if the perceptive field is not a rectangle coinciding perfectly with the matrix. For example, a visual field can be circular, with higher confidence near the center than near the boundary. In this way, arrays can represent various finite spatial patterns in sensation, each with a focus on a certain part in the perceptive field.

2. Inference Rules，Term Constructors, and Mental Operators

The sensory terms (arrays and array elements) are treated by the inference rules in the same way as the other types of terms.

There are special variants of rules that are dedicated to the sensory terms. For example, **temporal induction/comparison** do not require shared term in the premises, but their closeness in time. Similarly, **spatial induction/comparison** can be carried out among array elements that are close spatially to each other, so as to achieve functions like auto-filling, associative memory, and perceptual set (A perceptual bias or predisposition or readiness to perceive particular features of a stimulus).

There will be several constructors and operators for arrays:

- The *append* type constructor of list will be there in array to concatenate two arrays with compatible dimensions, such as from a p-by-n matrix and a q-by-n matrix to get a (p+q)-by-n matrix. As a special case, a matrix can be part of another matrix, if the above q is a negative number.

- An existing array can be *compressed* into a smaller array, and in the process the elements of the former will be summarized into those of the latter through a many-to-one mapping followed by truth-value revision. For example, if the compression ratio is 2:1 when a new matrix is constructed from an old one, the number of indices in both dimensions will be reduced by half, but it does not necessarily mean that every 2-by-2 sub-matrix in the original will be mapped into a single element in the compressed one, as the elements may come from a 4-by-4 sub-matrix in the original, where the ‘boundary elements’ get a confidence discount when merged with the central elements.

- A *shift* operation can change the center of the perceptive field, like eye movement (saccade) and “active vision” that directly adjust the sensors. Mental representation must include the operations to find invariants in perception, and the operations are associated with spatial relations (*above/below, left/right, front/back* …) without explicitly mentioning the numerical indices.

- When there are multiple objects in the scene, NARS will not segment a static image to get the objects and their relations, but shift its focus using the operations involved to code the relations of the objects recognized, such as "The triangle is above the circle''.

- Set theoretic constructors *union/intersection/difference* can be naturally defined on arrays, though their applicable situations on whole arrays remain to be identified. Within an array, they will be used to identify sameness (intersection), changes (difference), and alternatives (union). For instance, borders in an image is formed and enhanced as subjective contrast perception.

Though arrays are used with numerical indices, their usage in sensation and perception in NARS should not require numerical terms to be explicitly used. That is, the system never explicitly consider "the third pixel in the second row" when processing an visual image.

Inheritance and *similarity* judgments can be built between arrays of the same dimensionality, and evidence is collected by comparing elements at the corresponding locations in the two. An additional compression step will allow arrays of different sizes to be compared for this purpose.

Inheritance and *similarity* judgments can also be built between arrays and other terms, as in categorizing sensory and perceptive patterns.

Mental operators on arrays may be needed for *rotation*, *overlapping, dimensional reduction*, *masking*, and other manipulations.

3-D perception starts at the three degrees of freedom of body movements, combined with the feedbacks in the related sensorimotor channels (visual, auditory, kinesthetic, tactile …).

Movements are sequence of events. Object movements are perceived with compensation of movements of sensor and perceptive field.

3. Inference Tasks

With the coming of sensory experience, spontaneous forward inference happens as far as the significance of the signal is above the threshold of the sensory channel, which can be adjusted by factors including the system's anticipation, extent of busyness, emotional status, etc. This spontaneous inference can be seen as the return values of the system's *observation* operations.

The perceptive process will summarize the sensory experience into descriptions at multiple levels of generalization and abstraction in parallel. Also, the array-based 'sensory' representation and the concept-based 'symbolic' representation will co-exist at these levels, that is, the system represents the situation both as a mental image and as a judgment like "A cat is on a mat", where the latter is formed by matching the parts of the image with concepts in the system and noticing their relations. These two types of representation interweave at all levels. For instance, the image corresponds to an existing concept will be remembered better and accessed easier than an arbitrary image. This feature should allow the model to explain phenomena like Gestalt shapes, visual illusions, Bongard figures, and so on. The conceptual representation also provides the figure-ground differentiation.

During perception, the bottom-up signal-compression and the top-down anticipation form a mutual confirmation process (generative methods for "top-down" and discriminative methods for "bottom-up"). The sensory input first suggests some patterns with associated concepts, and anticipation and inference then increased the confidence of the guess, which in turn lead the fill-in of details.

As the system changes its internal states, it is normal for the same situation to be perceived differently, with different objects and events recognized. The perception is under constant revision with the coming of new experience, as well as with the continuous consideration of the system. It is not a function that maps the input signals into a unique 'correct' representation.

As the "current situation" is partially provided by previous perception, there will be *after effects*, as observed in human perception.

Beside the automatic self-organizing process in perception, the most common deliberative tasks are 'recognition' and 'imagination'. Roughly speaking, the former is to find a concept for an image, while the latter is to find an image for a concept, and the relation between an image and the corresponding concept is a special type of *inheritance* which is similar to the *represent* relation in NLP. In NARS, both processes are carried out by inference, with all types of uncertainty involved, and the final answer is chosen among the available candidates by balancing truthfulness, simplicity, and usefulness.

4. Features

Psychologists reached the consensus long ago that perception is multi-level abstraction, and deep learning just realizes this in special-purpose systems. Compare to them, the approach of NARS for perception will also achieve that, with the following characteristics:

- Using meaningful term connectors to carry out abstraction from level to level. It is assumed that the existing term connectors are sufficient for all necessary patterns — convolution and neuron models are basically weighted average functions followed by a non-linear step, which should be roughly equivalent to the set-theoretic operators.

- Carrying out multiple tasks in learning, so the intermediate results are not bounded to a single task, but have independent meaning. Therefore, learning results are naturally transferable. As there is no distinction between "hidden layer" and "input/output layer", results at any layer are understandable (to various degrees), and are adaptive with experience-grounded meaning.

- Using multi-level abstraction to solve "over-fitting" and "inductive bias", and to keep multiple hypotheses for a given problem. For the same observation, more abstract results are less confident, though they are simpler and can be supported by other observations later, so to become preferred than the more specific results.

- Using dynamic resource allocation to carry out local and incremental adjustments to provide real-time responses.

- Having stronger top-down influences, in the form of anticipation, familiarity, emotion, etc. The existing conceptual hierarchy plays a significant role in deciding what is perceived, while being adjusted in the process, as Piaget's assimilation-accommodation with stable perception as their equilibrium.

- Integrating perception with action, as (1) perception is carried out by operation, (2) perception and operation have unified representation, and (3) perceptive patterns are identified as invariants during certain operations.

Like the other processes, perception in NARS will not attempt to simulate human perception in all details, but its general principles. Consequently, it will still be closer to human than the existing AI techniques, like the perception-decision-action cycle used in many AI systems.