

Title: Neural modeling fields

URL: https://en.wikipedia.org/wiki/Neural_modeling_fields

PageID: 19208664

Categories: Category:Machine learning

Source: Wikipedia (CC BY-SA 4.0).

Neural modeling field (NMF) is a mathematical framework for machine learning which combines ideas from neural networks , fuzzy logic , and model based recognition . It has also been referred to as modeling fields , modeling fields theory (MFT), Maximum likelihood artificial neural networks (MLANS). [1] [2] [3] [4] [5] [6] This framework has been developed by Leonid Perlovsky at the AFRL . NMF is interpreted as a mathematical description of the mind's mechanisms , including concepts , emotions , instincts , imagination , thinking , and understanding . NMF is a multi-level, hetero-hierarchical system. At each level in NMF there are concept-models encapsulating the knowledge; they generate so-called top-down signals, interacting with input, bottom-up signals . These interactions are governed by dynamic equations , which drive concept-model learning, adaptation, and formation of new concept-models for better correspondence to the input, bottom-up signals.

Concept models and similarity measures

In the general case, NMF system consists of multiple processing levels. At each level, output signals are the concepts recognized in (or formed from) input, bottom-up signals. Input signals are associated with (or recognized, or grouped into) concepts according to the models and at this level. In the process of learning the concept-models are adapted for better representation of the input signals so that similarity between the concept-models and signals increases. This increase in similarity can be interpreted as satisfaction of an instinct for knowledge, and is felt as aesthetic emotions .

Each hierarchical level consists of N "neurons" enumerated by index $n=1,2,..N$. These neurons receive input, bottom-up signals, $X(n)$, from lower levels in the processing hierarchy. $X(n)$ is a field of bottom-up neuronal synaptic activations, coming from neurons at a lower level. Each neuron has a number of synapses; for generality, each neuron activation is described as a set of numbers,

, where D is the number or dimensions necessary to describe individual neuron's activation.

Top-down, or priming signals to these neurons are sent by concept-models, $M_m(S_m,n)$

, where M is the number of models. Each model is characterized by its parameters, S_m ; in the neuron structure of the brain they are encoded by strength of synaptic connections, mathematically, they are given by a set of numbers,

, where A is the number of dimensions necessary to describe individual model.

Models represent signals in the following way. Suppose that signal $X(n)$ is coming from sensory neurons n activated by object m , which is characterized by parameters S_m . These parameters may include position, orientation, or lighting of an object m . Model $M_m(S_m,n)$ predicts a value $X(n)$ of a signal at neuron n . For example, during visual perception, a neuron n in the visual cortex receives a signal $X(n)$ from retina and a priming signal $M_m(S_m,n)$ from an object-concept-model m . Neuron n is activated if both the bottom-up signal from lower-level-input and the top-down priming signal are strong. Various models compete for evidence in the bottom-up signals, while adapting their parameters for better match as described below. This is a simplified description of perception. The most benign everyday visual perception uses many levels from retina to object perception. The NMF premise is that the same laws describe the basic interaction dynamics at each level. Perception of minute features, or everyday objects, or cognition of complex abstract concepts is due to the same mechanism described below. Perception and cognition involve concept-models and learning. In perception, concept-models correspond to objects; in cognition models correspond to relationships and situations.

Learning is an essential part of perception and cognition, and in NMF theory it is driven by the dynamics that increase a similarity measure between the sets of models and signals, $L(\{X\}, \{M\})$. The similarity measure is a function of model parameters and associations between the input bottom-up signals and top-down, concept-model signals. In constructing a mathematical description of the similarity measure, it is important to acknowledge two principles:

Therefore, the similarity measure is constructed so that it accounts for all bottom-up signals, $X(n)$,

This expression contains a product of partial similarities, $l(X(n))$, over all bottom-up signals; therefore it forces the NMF system to account for every signal (even if one term in the product is zero, the product is zero, the similarity is low and the knowledge instinct is not satisfied); this is a reflection of the first principle. Second, before perception occurs, the mind does not know which object gave rise to a signal from a particular retinal neuron. Therefore, a partial similarity measure is constructed so that it treats each model as an alternative (a sum over concept-models) for each input neuron signal. Its constituent elements are conditional partial similarities between signal $X(n)$ and model M_m , $l(X(n)|m)$. This measure is "conditional" on object m being present, therefore, when combining these quantities into the overall similarity measure, L , they are multiplied by $r(m)$, which represent a probabilistic measure of object m actually being present. Combining these elements with the two principles noted above, a similarity measure is constructed as follows:

The structure of the expression above follows standard principles of the probability theory: a summation is taken over alternatives, m , and various pieces of evidence, n , are multiplied. This expression is not necessarily a probability, but it has a probabilistic structure. If learning is successful, it approximates probabilistic description and leads to near-optimal Bayesian decisions. The name "conditional partial similarity" for $l(X(n)|m)$ (or simply $l(n|m)$) follows the probabilistic terminology. If learning is successful, $l(n|m)$ becomes a conditional probability density function, a probabilistic measure that signal in neuron n originated from object m . Then L is a total likelihood of observing signals $\{X(n)\}$ coming from objects described by concept-model $\{M_m\}$. Coefficients $r(m)$, called priors in probability theory, contain preliminary biases or expectations, expected objects m have relatively high $r(m)$ values; their true values are usually unknown and should be learned, like other parameters S_m .

Note that in probability theory, a product of probabilities usually assumes that evidence is independent. Expression for L contains a product over n , but it does not assume independence among various signals $X(n)$. There is a dependence among signals due to concept-models: each model $M_m(S_m, n)$ predicts expected signal values in many neurons n .

During the learning process, concept-models are constantly modified. Usually, the functional forms of models, $M_m(S_m, n)$, are all fixed and learning-adaptation involves only model parameters, S_m . From time to time a system forms a new concept, while retaining an old one as well; alternatively, old concepts are sometimes merged or eliminated. This requires a modification of the similarity measure L ; The reason is that more models always result in a better fit between the models and data. This is a well known problem, it is addressed by reducing similarity L using a "skeptical penalty function," (Penalty method) $p(N, M)$ that grows with the number of models M , and this growth is steeper for a smaller amount of data N . For example, an asymptotically unbiased maximum likelihood estimation leads to multiplicative $p(N, M) = \exp(-N \text{ par} / 2)$, where $N \text{ par}$ is a total number of adaptive parameters in all models (this penalty function is known as Akaike information criterion, see (Perlovsky 2001) for further discussion and references).

Learning in NMF using dynamic logic algorithm

The learning process consists of estimating model parameters S and associating signals with concepts by maximizing the similarity L . Note that all possible combinations of signals and models are accounted for in expression (2) for L . This can be seen by expanding a sum and multiplying all the terms resulting in MN items, a huge number. This is the number of combinations between all signals (N) and all models (M). This is the source of Combinatorial Complexity, which is solved in NMF by utilizing the idea of dynamic logic. [7] [8] An important aspect of dynamic logic is matching vagueness or fuzziness of similarity measures to the uncertainty of models. Initially, parameter values are not known, and uncertainty of models is high; so is the fuzziness of the similarity measures. In the process of learning, models become more accurate, and the similarity

measure more crisp, the value of the similarity increases.

The maximization of similarity L is done as follows. First, the unknown parameters $\{S_m\}$ are randomly initialized. Then the association variables $f(m|n)$ are computed,

Equation for $f(m|n)$ looks like the Bayes formula for a posteriori probabilities; if $l(n|m)$ in the result of learning become conditional likelihoods, $f(m|n)$ become Bayesian probabilities for signal n originating from object m . The dynamic logic of the NMF is defined as follows:

The following theorem has been proved (Perlovsky 2001):

Theorem . Equations (3), (4), and (5) define a convergent dynamic NMF system with stationary states defined by $\max\{S_m\}L$.

It follows that the stationary states of an MF system are the maximum similarity states. When partial similarities are specified as probability density functions (pdf), or likelihoods, the stationary values of parameters $\{S_m\}$ are asymptotically unbiased and efficient estimates of these parameters. [9]
The computational complexity of dynamic logic is linear in N .

Practically, when solving the equations through successive iterations, $f(m|n)$ can be recomputed at every iteration using (3), as opposed to incremental formula (5).

The proof of the above theorem contains a proof that similarity L increases at each iteration. This has a psychological interpretation that the instinct for increasing knowledge is satisfied at each step, resulting in the positive emotions: NMF-dynamic logic system emotionally enjoys learning.

Example of dynamic logic operations

Finding patterns below noise can be an exceedingly complex problem. If an exact pattern shape is not known and depends on unknown parameters, these parameters should be found by fitting the pattern model to the data. However, when the locations and orientations of patterns are not known, it is not clear which subset of the data points should be selected for fitting. A standard approach for solving this kind of problem is multiple hypothesis testing (Singer et al. 1974). Since all combinations of subsets and models are exhaustively searched, this method faces the problem of combinatorial complexity. In the current example, noisy 'smile' and 'frown' patterns are sought. They are shown in Fig.1a without noise, and in Fig.1b with the noise, as actually measured. The true number of patterns is 3, which is not known. Therefore, at least 4 patterns should be fit to the data, to decide that 3 patterns fit best. The image size in this example is $100 \times 100 = 10,000$ points. If one attempts to fit 4 models to all subsets of 10,000 data points, computation of complexity, $MN \sim 10^6$. An alternative computation by searching through the parameter space, yields lower complexity: each pattern is characterized by a 3-parameter parabolic shape. Fitting $4 \times 3 = 12$ parameters to 100×100 grid by a brute-force testing would take about 10^{32} to 10^{40} operations, still a prohibitive computational complexity.

To apply NMF and dynamic logic to this problem one needs to develop parametric adaptive models of expected patterns. The models and conditional partial similarities for this case are described in details in: [10] a uniform model for noise, Gaussian blobs for highly-fuzzy, poorly resolved patterns, and parabolic models for 'smiles' and 'frowns'. The number of computer operations in this example was about 10^{10} . Thus, a problem that was not solvable due to combinatorial complexity becomes solvable using dynamic logic.

During an adaptation process, initially fuzzy and uncertain models are associated with structures in the input signals, and fuzzy models become more definite and crisp with successive iterations. The type, shape, and number, of models are selected so that the internal representation within the system is similar to input signals: the NMF concept-models represent structure-objects in the signals. The figure below illustrates operations of dynamic logic. In Fig. 1(a) true 'smile' and 'frown' patterns are shown without noise; (b) actual image available for recognition (signal is below noise, signal-to-noise ratio is between -2 dB and -0.7 dB); (c) an initial fuzzy model, a large fuzziness corresponds to uncertainty of knowledge; (d) through (m) show improved models at various iteration stages (total of 22 iterations). Every five iterations the algorithm tried to increase or decrease the number of models. Between iterations (d) and (e) the algorithm decided, that it needs three Gaussian models for the 'best' fit.

There are several types of models: one uniform model describing noise (it is not shown) and a variable number of blob models and parabolic models; their number, location, and curvature are estimated from the data. Until about stage (g) the algorithm used simple blob models, at (g) and beyond, the algorithm decided that it needs more complex parabolic models to describe the data. Iterations stopped at (h), when similarity stopped increasing.

Neural modeling fields hierarchical organization

Above, a single processing level in a hierarchical NMF system was described. At each level of hierarchy there are input signals from lower levels, models, similarity measures (L), emotions, which are defined as changes in similarity, and actions; actions include adaptation, behavior satisfying the knowledge instinct – maximization of similarity. An input to each level is a set of signals $X(n)$, or in neural terminology, an input field of neuronal activations. The result of signal processing at a given level are activated models, or concepts m recognized in the input signals n ; these models along with the corresponding instinctual signals and emotions may activate behavioral models and generate behavior at this level.

The activated models initiate other actions. They serve as input signals to the next processing level, where more general concept-models are recognized or created. Output signals from a given level, serving as input to the next level, are the model activation signals, a_m , defined as

$$a_m = \sum_{n=1..N} f(m|n).$$

The hierarchical NMF system is illustrated in Fig. 2. Within the hierarchy of the mind, each concept-model finds its "mental" meaning and purpose at a higher level (in addition to other purposes). For example, consider a concept-model "chair." It has a "behavioral" purpose of initiating sitting behavior (if sitting is required by the body), this is the "bodily" purpose at the same hierarchical level. In addition, it has a "purely mental" purpose at a higher level in the hierarchy, a purpose of helping to recognize a more general concept, say of a "concert hall," a model of which contains rows of chairs.

From time to time a system forms a new concept or eliminates an old one. At every level, the NMF system always keeps a reserve of vague (fuzzy) inactive concept-models. They are inactive in that their parameters are not adapted to the data; therefore their similarities to signals are low. Yet, because of a large vagueness (covariance) the similarities are not exactly zero. When a new signal does not fit well into any of the active models, its similarities to inactive models automatically increase (because first, every piece of data is accounted for, and second, inactive models are vague-fuzzy and potentially can "grab" every signal that does not fit into more specific, less fuzzy, active models. When the activation signal a_m for an inactive model, m , exceeds a certain threshold, the model is activated. Similarly, when an activation signal for a particular model falls below a threshold, the model is deactivated. Thresholds for activation and deactivation are set usually based on information existing at a higher hierarchical level (prior information, system resources, numbers of activated models of various types, etc.). Activation signals for active models at a particular level $\{a_m\}$ form a "neuronal field," which serve as input signals to the next level, where more abstract and more general concepts are formed.

References

Related

Leonid Perlovsky