Principles of Regulated Activation Networks

Alexandre Miguel Pinto and Leandro Barroso

CISUC - University of Coimbra, Portugal
ampinto@dei.uc.pt, lafb@student.dei.uc.pt

Abstract. In a context of growing awareness and prevalence of mental disorders, cognitive modeling has emerged as an important contribution to the study of the mind and its processes. Computational models have proved to be indispensable tools for precise and systematic simulations of cognitive processes, and have a potential application in the diagnosis and treatment of such pathologies.

We propose a connectionist cognitive model that incorporates regulatory mechanisms, called Regulated Activation Networks (RANs), that will be applied to the modeling of psychological phenomena. This paper summarizes the current early stages of the development of the RANs model. The objectives, principles and approach taken are described, as well as the architecture of the RANs model, some preliminary results and plans for future work.

Keywords: Cognitive model, connectionism, conceptual spaces, psychological phenomena.

1 Introduction

Context: Every year, a third of the European Union population suffers from mental disorders [8]. In a global scale the numbers are also worrying: the World Health Organization predicts that depression will be the leading cause of disease burden by 2030 [16]. This alarming trend is one of the reasons for the growing importance of the cognitive science research field, which focuses on the study of the mind and its processes, especially information representation, processing and transformation. Particularly, the development of computational models which provide algorithmic specificity, conceptual clarity and precision, has allowed the realization of simulations that can either be useful to test and validate psychological theories or to generate new hypotheses about how the mind works. This has turned them into indispensable tools in the study of the human mind.

Motivation: We aim at modeling mental processes with our RANs tool and, hopefully, use it to aid in the diagnosis and treatment of some mental pathologies such as depression, schizophrenia and autism. To achieve this the RANs model will need to include the parameters that are enough to mimic the neurological features responsible for simulating and regulating both healthy and pathological cognitive processes. We envisage a scenario where, given a particular instance of the RANs model, we allow it to learn the values of those regulatory parameters from the inputs (e.g., words) given to, and outputs (e.g., other words) obtained from, a specific human user – with different users we train different instances of the RANs. As some users may be previously diagnosed as

N. Hernandez et al. (Eds.): ICCS 2014, LNAI 8577, pp. 231–244, 2014.
DOI: 10.1007/978-3-319-08389-6_19, © Springer International Publishing Switzerland 2014

clinical patients with some disorders, while others may be classified as healthy, we can then use this collection of data to train a classifier to allow the diagnosis of new human subjects. Some results in the literature [3,15,21] have already been achieved by other researchers, which shows that this ambitious goal is not out of reach of computational cognitive modeling. This type of computational tools with the ability to capture cognitive phenomena has also the potential to simulate and help study some mental states and processes such as those linked to creativity [12].

We aim to develop of a computational cognitive model with two main very long-term goals. The first is to help understand and simulate cognitive phenomena such as perception, emotion, learning and reasoning, creativity and, ultimately, different kinds of psychological features and personality traits. The second is to provide a tool that can aid in the simulation, diagnosis and, eventually, treatment, of said pathologies.

Challenges: Taking an incremental development and validation approach, in these first steps we focus only on the following features: 1) the model must be able to receive (sensory-like) input information from the outside, 2) it must allow for the representation and simulation of the dynamic time-variant cognitive state (initially set by the input data), 3) it must be able to learn and abstract the patterns in its cognitive state, and 4) it must be able to exhibit creative behavior by creating new concepts and by generating new patterns of cognitive state. While developing mechanisms to represent the input data may not be hard, simulating a dynamic cognitive state and corresponding learning, and creative processes are not trivial. We do so by modeling certain psychological phenomena concerned with the reception and response to the received stimulus; and also two different kinds of learning. Specifically, we aim, at this stage, to model the effect of Priming, the occurrence of False Memories and the Habituation and Sensitization processes — we describe these phenomena in detail in subsection 2.1 – as these will allow for the simulation of a simple cognitive state that changes through time. The learning mechanism we implemented at this stage progressively identifies correlations between elements of the cognitive state. At a later stage, left for now for future work, we will address more complex kinds of learning (including the ones capable of creating new concepts) and reasoning, and emotion representation, elicitation, and processing.

Paper Structure: Section 2 gives an overview of the basic psychological phenomena we model in this early stage of our work. We also describe the cognitive modeling approaches in the literature that are most relevant to our model, as it draws inspiration and features from those approaches. In section 3 we present the overarching principles behind the design our Regulated Activation Networks (RANs) cognitive model, and also discuss the desiderata properties for the future full version of the model, as well as its global characteristics, some of which are already implemented in the preliminary version detailed in section 4.

In section 5 we show our first preliminary results; and in section 6 we provide conclusions and description of future work, which includes the design of a validation plan we will use to assess the reliability regarding the ability to simulate the identified psychological phenomena.

2 Background

First, we review the main psychological phenomena our current simplified RANs model is intended to capture, and why these phenomena matter for the future full version. Then, we show the cognitive modeling approaches in the literature that we found to have features that are important for our purposes – the RANs model includes an innovative combination of these features, plus other new ones we describe later.

2.1 Psychological Phenomena

Priming: Priming is a phenomenon in which a response to a stimulus is influenced by a previous exposure to the same or a similar stimulus. It happens in a non-conscious way, making it an implicit memory effect [11]. Priming effects can be divided in two main types, according to the relation between the stimulus. Perceptual priming stands for stimulus with a similar form and conceptual/semantic priming for stimulus with similar meaning. E.g., if someone is shown a list of words containing the word "mystery", and then the subject is asked to do a word completion task in which there is the incomplete pattern "_ys_e_y", the probability that the person will select the word "mystery" is higher than if the person had not been primed [19]. This is a form of perceptual priming for the stimulus relate in its form. As an example of conceptual priming, if we consider the concept "fruit", that stimulus will have positive effects on the response to semantically related concepts, such as "apple", leading to a faster response to that stimulus [13]. This can occur even when the first concept is consciously forgotten. Other priming types have been suggested (e.g. associative priming) in which stimulus are not related semantically but are frequently associated or have a high probability of occurring together. Another similar effect is context priming, in which context is used to deal in a faster way with stimulus more likely to occur in the context.

The simulation of the priming effect is central to the modeling of psychological phenomena related to implicit memory such as time-varying cognitive and emotional context, bias, predisposition, prejudice, and many others, such as learning associations and recalling related concepts.

Deese-Roediger-McDermott (DRM) Paradigm: The DRM paradigm is a procedure initialized by J. Deese in the 1950s, and extended by H.L. Roediger and K. McDermott in the 1990s with the aim to study false memory and false recall phenomena [20]. The process typically involves reading a list of words to a subject, being all the words semantically related to a non-present word. E.g., the words "bed", "rest", "awake" and the non-present word "sleep". After hearing the list, the subject is asked to recall the words or to select those words from a new list. In both cases, the non-present target word is recalled with the same frequency as the other words. Also, a high percentage of subjects assure remembering hearing the word, suggesting the occurring of a false memory. This phenomenon is quite similar to the conceptual priming described above, suggesting that the underlying mechanisms in both effects are the same. However, the false memory in the DRM paradigm implies that, in the retrieval phase, the target concept reaches an activation level similar to the other concepts making the subject believe to have heard the word and not only facilitating the response to it.

We will use the modeling of the false memory mechanism to simulate thought-drifting, dreams, delirium, and other divergent cognitive processes that may be necessary to capture a variety of healthy mind processes and psychological pathologies.

Habituation: Habituation is a behavioral response decrease, common in humans and animals, that occurs after repeated exposure to the same stimulus. This process is distinct from sensory adaptation, in which sensory receptors change their sensitivity to the stimulus, and that distinction is demonstrated by the inverse process (dishabituation) and also by stimulus discrimination and spontaneous recovery of the habituation process [7]. The following are some of the characteristics of the habituation process we intend to model. Repeated exposure to a stimulus results in a gradual change of the response to an asymptotic level. In most cases this results in an exponential decrease, but linear habituation can also occur. Before the habituation, a response may show facilitation due to a simultaneous process of sensitization (response amplification to the stimulus). The decrease in habituation can be observed in response frequency, magnitude or both. Habituation is a recoverable process. When given enough time without exposure to the stimulus, the response recovers in a partial or total way (spontaneous recovery). After repeated series of habituation and spontaneous recovery, the habituation is potentiated, i.e. the decrease in response occurs progressively faster and more intensely. Potentiation can also occur by increasing the frequency of the stimulus, that will also result in faster and more intense decrease of response. The frequency of stimulation, after the response reaches an habituated level, has been suggested to determine the rate of recovery [17]. Repeated stimulation in this phase may delay the onset of spontaneous recovery. Associated with this process are also the concepts of stimulus generalization and discrimination. Once a response to a stimulus is habituated, a similar novel stimulus will also have a certain degree of response decrease, according to the rate of similarity between the novel and previous stimulus (generalization). Discrimination is observed when a different stimulus does not have its response altered by the habituation of a previous stimulus. Dishabituation occurs when the presentation of another stimulus results in the recovery of the habituated response of the original stimulus.

Modeling these processes is central for simulating learning processes, surprise, response to, and recovery from, traumatic stress, among others.

2.2 Cognitive Modeling Approaches

Spreading Activation: Spreading Activation is a theory of memory [1] based on Collins and Quillian's computer model [4] which has been widely used for the cognitive modeling of human associative memory and in other domains such as information retrieval [5]. It intends to capture both the way information is represented and how it is processed. According to the theory, long-term memory is represented by nodes and associative links between them, forming a semantic network of concepts. The links are characterized by a weight denoting the associative or semantic relation between the concepts. The model assumes activating one concept implies the spreading of activation to related nodes, making those memory areas more available for further cognitive processing. This activation decays over time, and the further it spreads, which can occur through multiple levels [14], the weaker it is. That is usually modeled using a decaying factor for activation. The method of spreading activation has been central in many cognitive models due to its tractability and resemblance of interrelated groups of neurons in the human brain [18]. The connection between spreading activation theories and priming effects is clear: when an activated node propagates activation to a semantically related node semantic priming effects can be observed. I.e., automatic spreading of activation between concepts are the underlying mechanisms for conceptual priming. As for the creation of false memories it has been discussed whether or not the automatic spreading of activation would be sufficient to originate the high rates of false recall observed in the DRM paradigm. Evidence from [22] leads to the conclusion that the target concept can be activated with such mechanisms.

Hopfield Networks: Invented by John Hopfield in 1982, Hopfield Networks are recurrent neural networks. Each node is a binary threshold unit, i.e. it only assumes two possible values, normally 1 and -1, determined by whether the node's input is above or under its threshold. Each pair of nodes has a connection characterized by a weight *w*, being the connections symmetrical: $w_{ij} = w_{ji}, \forall i, j$, and a node is not allowed to connect to itself. Hence, the input of a node is the sum of other nodes' states multiplied by the weight of the connection between them.

One of the most interesting properties of Hopfield Networks is their ability to store and retrieve patterns, working as an associative memory. This model uses an energy function to determine the current state of the network, and remembering a learned pattern is achieved by descending a gradient of energy toward a local minimum corresponding to the pattern. Learning patterns results from training the network by lowering the energy of the state that the network should remember. A common way to do this is using Hebbian learning, strengthening the weights of connections between simultaneously activated nodes, and reducing the weight of the connection otherwise. This allows the network to recover a "stored pattern" when given an incomplete version of that pattern. With this training process, Hopfield Networks can store, approximately, 0.14 * *n* patterns, *n* being the number of nodes [23].

Conceptual Spaces: Traditionally there are two main approaches to the problem of modeling representations in artificial intelligence and cognitive science. One is the symbolic approach, in which information is represented by symbols that when combined give rise to expressions that relate to each other in a logical way. Processing information in the symbolic paradigm corresponds to manipulating symbols, not regarding their semantic content. The other is the connectionist approach, from which artificial neural networks are the prime example. In this approach, cognitive processes are represented by the dynamic activity of patterns of several interconnected units. Peter Gärdenfors argues that none of these approaches can model some aspects of cognitive phenomena and proposes a new way to represent information based on the use of geometrical structures, rather than symbols or connections between neurons. However, this approach is not a substitute for previous approaches, but an intermediate level explaining how symbolist representations can arise from connectionist ones [9]. This level of representation is called conceptual for it provides a way to describe concept formation [10]. The thesis focuses on the existence of conceptual spaces as a way to locate concepts in a domain, being the conceptual space formed by a set of quality dimensions which represent object properties and are used to specify relations between them. Examples for quality dimensions can be height, width, depth, temperature, color, etc. These dimensions are endowed with the appropriate geometrical structure for its representation. For instance, Henning's tetrahedron[6] representation for the human gustatory space could be used for the quality dimension "taste". The spatial location of a point in the conceptual space allows for the calculation of distance between points and the measure of similarity between concepts, which would be impossible to do in a symbolic approach. However, the conceptual spaces theory imposes some constraints on what kinds of subspaces can be considered concepts, namely requiring them to be convex, which may compromise its applicability in general. We need a more general geometric notion of concept and that requires a more powerful way of extracting the features, from particular examples, that define the shape of the concept the examples belong to.

Deep Learning: "Deep learning" is a recent family of machine learning methods that attempt to model high-level abstractions in data by using architectures composed of multiple layers [2]. They usually resort to (restricted) Boltzmann Machines in each layer using a feed forward input layer with no lateral connections. Although deep learning techniques are very powerful indeed for extracting features from complex data and creating new representations for those more abstract concepts, they are not fit for representing the direct relationships between same level concepts, vis-à-vis their lack of "lateral" connections between same layer nodes, which is crucial for the simulation of priming and DRM.

3 Principles of Regulated Activation Networks

The RANs model must be capable of representing and simulating the dynamic cognitive state of an agent, its learning and recall processes, the association of ideas, and the creation of new, more abstract, concepts. We assume these can be broken down into, and emerge from, more basic cognitive phenomena simulated by simpler computational processes. Particularly, in order to simulate the dynamic cognitive state, we need at least 1) a notion of a time-variant activation state of a given concept in the agent's mind; and 2) an adaptive notion of relation and influence between two concepts dependent on their respective activations in a given instant. These two constructs are in principle enough to simulate the Priming phenomenon and, along with a sufficiently time-condensed sequence of activation of concepts, enough to simulate as well the False Memory phenomenon – when several concepts (e.g., representing words) are activated by input, the concepts positively related to them should become more active as well, and if their received combined activation is strong enough, the concept should be sufficiently active to be considered "remembered" by the agent. For the Habituation and Sensitization phenomena we need also our model to 3) be able to dynamically change the parameters controlling its behavior. Finally, we also need 4) a learning mechanism that can create new more abstract concepts as patterns of activation are detected among the existing ones.

The first step of the RANs model development is the creation of a connectionist layer of nodes, each one representing a dimension of a concept – for very simple concepts, a single node might suffice to represent the whole concept. A possible extension of the RANs model consists in establishing a correspondence between individual nodes and concepts in a user-defined ontology. Under this setting, a high activation level of a given node may be interpreted as the detection of an instance of the related ontological concept, and the spreading of activation may afford a kind of inference.

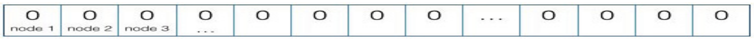

Fig. 1. Initialized single-layer RAN model (connections not shown)

In this way it resembles a semantic network, in which each node has an activation state representing its importance at the moment, and related nodes are associated through a link with a weight representing the strength of the relation. This interpretation of the nodes also allows the representation of a conceptual space where each node stands for a dimension of the space and its activation level corresponds to a particular value along that dimension. This layer of nodes receives input information, responding to the stimulus and learning an internal representation of that stimulus through Hebbian-like weight changing.

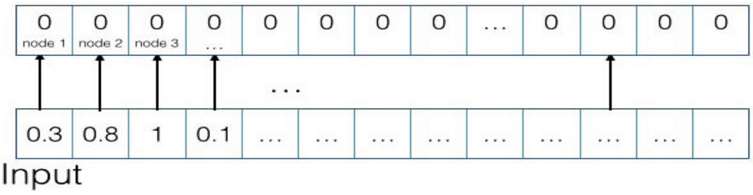

Fig. 2. "Sensory" activation input to nodes in a single-layer RAN model (connections not shown)

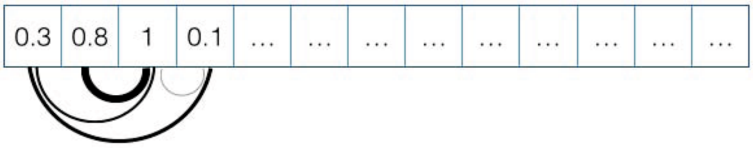

Fig. 3. Hebbian learning in a RAN

This Hebbian learning means, from a conceptual spaces perspective, the identification of a correlation between the dimensions of the concept. Also, since each node is connected to every other node in the same layer, it resembles a Hopfield Network, allowing the learning, and the emerging, of certain patterns of nodes' activations which then function as attractors. These attractors, points in the conceptual space, may be seen as prototypical examples of the concept represented by all the points that converge, via the spreading of activation, to the attractor. This representation has the advantage, over

Gärdenfors specification, of not imposing the restriction that all concepts must be convex subspaces. The particular geometric shape of the concept will emerge from the Hebbian learning on the layer.

Within a given cognitive level of abstraction, mechanisms such as spreading activation, and Hebbian correlation identification may be enough for some of our cognitive modeling goals. In particular, a single-layer RANs model (as the one we have implemented and describe herein) only allows us to model the learning of simple correlations between concepts. However, if we wish our RANs model to capture the generalization and abstraction processes involved in higher-level cognitive processes, it must contain some mechanism to progressively build new more abstract concepts into the model. The RANs model will have to combine the "lateral" connectivity typical of models where spreading activation can be applied, with deep learning capabilities for producing new, more abstract layers of concepts. In the future full version of the RANs model a higher learning mechanism will be triggered when, within one layer, the nodes' activation states stabilize: at that moment the RANs model will capture the network state, and represent it by creating a new corresponding node in a new layer of higher abstraction. This capturing can be done via a (restricted) Boltzman Machine, as it is usually the case in deep learning architectures, or any other process that affords the extraction of the relevant features in the pattern. This way, an instance of the RANs model evolves into a deep structured set of layers, each one with a superior level of semantic abstraction, reducing the number of dimensions (a pattern of activation in lower nodes is mapped into a single higher-layer one), potentially drifting apart from connectionism, into a gradually more symbolic representation. With these features, we intend the RANs model to allow for the incremental learning of bottom-up deep representations of progressively more abstract concepts in conceptual spaces. While the semantics of the input "sensory" nodes might be user specified, the new nodes that are eventually created for representing detected patterns, and/or features, of activation may have no immediately obvious semantics. However, it may be easier to recognize complex features/concepts represented in the top level most abstract nodes.

Recalling concepts in the RANs model can be elicited at any desired level of abstraction. The user just has to input activation into any set of nodes and let the activation spread across the RAN until it stabilizes in a fixed-point pattern. When a node N at level n is activated, it spreads its activation, not only to its companion nodes in the same n layer, but also to the nodes in layer $n-1$ below which correspond to the pattern the node N represents. These in turn repeat the process spreading activation "laterally" to nodes in layer $n-1$ and also downwards. Whenever the intra-layer spreading of activation causes a stable fixed-point state to emerge, the nodes in the layer above get also activated according to the similarity between the pattern of activation in the layer below they represent and the current pattern active in the layer below. This mechanism thus allows for activation to be spread 1) "upwards" whenever a stable state emerges in a lower layer, 2) "downwards" whenever a node is activated for recall, and 3) "laterally" to nodes in the same layer in all cases. Naturally, all these dynamics depend strongly on the parameters like the decay factor (which controls how much the activation decays inside a node before it is spread), the learning rates (which control how strongly to update the weights when learning), and others.

4 Architecture of the Single-layer RANs model

We now overview the basic characteristics of the single-layer RANs model.

Network Topology: The RANs' topology, at this point, consists of one layer of nodes. We have only tested the model with a full connectivity, but different configurations will be tested in the future, namely: each node linked only to a fixed percentage of the total of nodes; and layers with a small-world connectivity pattern.

Node Properties: Each node has an internal state, represented by its activation level, which ranges from some minimum value to some maximum value and varies continuously in time (the range $[-1, 1]$ was the one implemented, but we are currently experimenting with the range $[0, 1]$). The mean value in the domain is called the rest state. The activation of node at time t is denoted by $A_{ni}(t)$. When a node has a positive activation state (above rest state) we consider it to be active, when it has negative activation (below rest state) we consider it repressed, and if it is equal to the rest state it is considered indifferent. The semantics of the activation values may need to be redefined for the $[0, 1]$ interval.

In the absence of input activation from other nodes or input injected in the network, the activation level of each node gradually converges to the rest state, accordingly with a reposition rate (previously called decay factor), denoted by R_{n_i}, which ranges in $[0, 1]$.

Each node n_i has a threshold variable through time between the minimum and the maximum activation level. Other configurations not including a threshold are currently under consideration.

Activation Propagation: Links between nodes have a weight associated, which represents the importance of the activation from the source node to the next activation level of the target node. The weight of the link from node n_i to node n_j is denoted by w_{ij} and at the beginning all weights are set to zero: $\forall i, j, w_{ij} = 0$. This initialization to zero, drastically different from what happens in traditional feedforward networks where weights are initialized to random values, is inspired by the synaptogenesis process in the human brain where neurons have initially very few connections, and grow new ones as the child grows up.

At each processing step, each node will propagate activation and update its threshold if

$$|A_{n_i}(t)| > |T_{n_i}(t)| \quad (1)$$

In that case, the threshold will be updated according to the formula

$$T_{n_i}(t+1) = T_{n_i}(t) - \delta * \Delta TA \quad (2)$$

where $0 \leq \delta \leq 1$ is the threshold's learning rate, and $\Delta TA = |A_{n_i}(t) - T_{n_i}(t)|$.

The activation propagated to each node n_j linked with n_i is

$$A_{n_i}(t) * w_{ij} \tag{3}$$

When some node receives activation from its neighbors it is combined with the activation of the node itself. The activation level of node at time $t + 1$ is:

$$A_{n_i}(t+1) = \lambda_{n_i} * ((1 - R_{n_i}) * A_{n_i}(t)) + (1 - \lambda_{n_i}) * f(\sum_j A_{n_j}(t) * w_{ij}) \tag{4}$$

where $0 \leq \lambda_{n_i} \leq 1$ is the relative importance the node gives to its own activation versus the activation received from other nodes, and it is called the *solipsism factor*, and $f : \Re \rightarrow$ [minimum activation level, maximum activation level] is a sigmoid function, e.g. $f(x) = \frac{1}{1+e^{-\beta x}} * 2 - 1$.

In a similar way, when an input is injected in the network, that component is pondered with the node activation level, therefore being the activation level of node n_i at time $t + 1$ given by the formula:

$$A_{n_i}(t+1) = \lambda_{n_i} * ((1 - R_{n_i}) * A_{n_i}(t)) + (1 - \lambda_{n_i}) * I(i) \tag{5}$$

where $I(i)$ is the input to node n_i.

Learning: Learning in the RANs model consists of two different processes. The first one is the activation correlation process, in which weights are updated accordingly to the correlation between the nodes' activation level, in a Hebbian influenced learning. The goal is to strengthen the links between nodes with similar levels of activation and to weaken the links otherwise. This way, each time an input is received, for each node n_i, connections with other nodes are updated according to the level of similarity between its activation states pondered by a learning rate. In this case, as activations vary between -1 and 1, the similarity rate is transformed to that same interval. The variation of weights is thus

$$\Delta W_{ij} = \mu_{ac} * (2 * similarityRate - 1) \tag{6}$$

where $0 \leq similarityRate = 1 - \dfrac{|A_{ni}(t) - A_{nj}(t)|}{maxDif} \leq 1$; $0 \leq \mu_{ac} \leq 1$ is the learning rate for activation correlation; maxDif = actMax - actMin; actMax and actMin are, respectively, the maximum and minimum level of activation. After that process, the network is given one time instant to spread activation. The inclusion of a second error-driven learning process, similar to a backpropagation, is currently under consideration.

Simulation Procedure: The general procedure for running a simulation with the RANs model comprehends the following steps: Considering N the number of nodes, initialize NxN weight matrix and 1xN activation and threshold vectors to the same value as the rest state. Initialize 1xN reposition rate and solipsism factor vectors and threshold learning rate, activation correlation learning rate and simulation time with its respective values (these are currently being subject to grid-search experimentation). Schedule a set of 1xN input patterns and the time steps for their injection. Initialize time variable to 1.

Until time reaches the established simulation time, run the following execution cycle: In case of a pattern scheduled to that time instant, apply equation 5 to calculate the new level of activation for the nodes and equation 6 to update the weights according to the current activation level. Otherwise, apply equation 4 to update the nodes' activation level according to the output generated in the previous time step by the same nodes. After that, for each node to which the condition 1 applies, calculate its activation output using equation 3 and update its threshold using equation 2. In the future, these outputs may be subject to some transformation (e.g. by regulatory mechanisms) before being used as inputs. Finally, collect any desired data (current activation levels, weights, etc.) and update the time variable.

The execution cycle is summarized in the following pseudo code:

Algorithm 1. RANs cycle

```
while time < simulation time do
    if isPatternScheduled(time) then
        inject input pattern (5)
        do activation correlation learning process (6)
    else
        activation input (4)
    end if
    for each node do
        if node activation above threshold then (1)
            calculate node output (3)
            update threshold (2)
        end if
    end for
    collect data
    $time \leftarrow time + 1$
end while
```

5 Preliminary Results

The results herein shown regard a specific parametrization of the model. However, the architecture of the RANs model will be subject to detailed exploration and experimentation concerning its topological properties and the influence of different types of connectivity, learning, regulation processes and parameters on the network's behavior and utility to the modeling of the intended phenomena. Still, these preliminary results serve as an appetizer for the model's capabilities while providing some insight on how we are dealing with the input patterns.

The simulations used the following parameters: Num. of nodes: 50; Connectivity: Total; Activations in $[-1, 1]$; Weights in $[-\infty, \infty]$; $\forall i, R_{ni} = 0.05$; $\forall i, \lambda_{ni} =$ Random value (uniform distribution over $[0, 1]$); $\forall i, A_{ni}(0) = 0$; $\forall i, \delta_{ni} = 0.1$.

The simulation process involves generating a random pattern, and injecting it periodically (in this case, each 100 time steps) in the network. Fig. 4 shows how nodes' activation evolve through time (each line represents a single node activation state). From

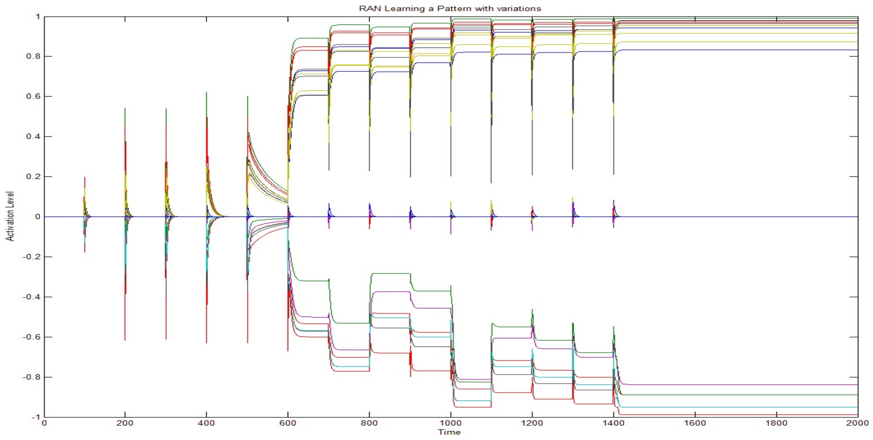

Fig. 4. RANs preliminary experiments with learning a pattern

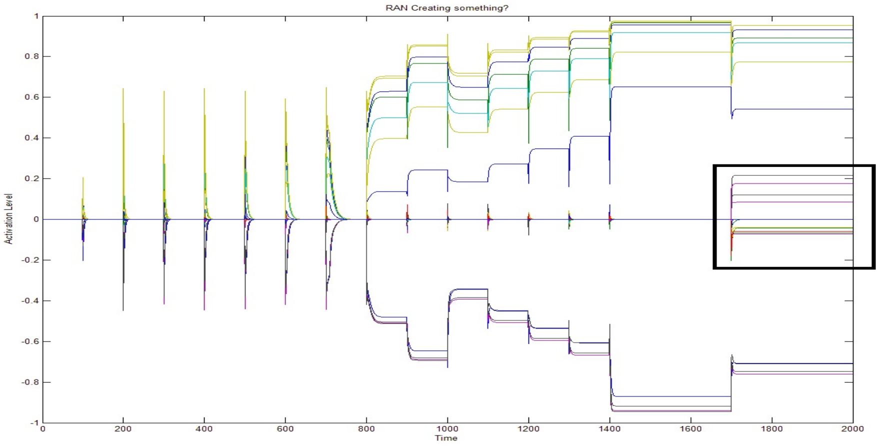

Fig. 5. RANs preliminary experiments with creative processes – “new zone” boxed

a global perspective, we observe that when an input is injected the network has an immediate response (for that input is directly injected in the nodes) and the activation is reposed to the rest state value (in this case, 0). Each time the pattern is injected, a round of Hebbian updating of weights takes place, and as a consequence of that learning, the reposition phase becomes progressively longer, i.e, the network takes longer to return to a neutral cognitive state. When an input is given in a non-neutral state ($t = 600$) the network alters its behavior and activations start to converge to a stable fixed-point state. Subsequent pattern injections only reveal minor changes in the value for each node activation, and can be considered as slight adjustments to the representation of the pattern previously learned.

A first attempt at simulating creativity was experimented as well. The process was very similar to the previous one, with the exception that at a time step where the network has already converged to a stable state ($t = 1400$, fig. 5) we stopped injecting the input patterns, and at $t = 1700$ a new random pattern was inserted. The interest in fig. 5 comes from the difference between the activations before and after that moment. Besides some minor changes to some nodes' activations, we observe a totally new "zone of activation" that was previously neutral. It is premature to assume that the production of the new RANs' state can be described as creative, but the fact that our model can integrate two different states in a new representation can be a good starting point for modeling creative processes.

6 Conclusions and Future Work

Prevalence of psychological and psychiatric diseases, such as depression, schizophrenia and others, is a growing concern in the industrialized world. Computational cognitive models can help in understanding and simulating mental processes, both healthy and pathological ones, hopefully contributing to the diagnosis and treatment of the latter. Also, there is a recent growing interest in understanding and potentiating creative processes, both in humans and in computers. For these reasons, we put forward the desiderata, and first simplified version, of our Regulated Activation Networks model, a new computational cognitive model capable of simulating a variety of psychological phenomena including, among others, Priming, False Memory, Habituation, different kinds of Learning, and Memory. The current preliminary version of the RANs model has only one fully connected layer of nodes representing concepts, uses a Hebbian learning rule, and resorts to spreading activation as means for inference and recall.

We are also currently experimenting with a probabilistic approach at the Hebbian learning of weights and the spreading of activation. Future work includes implementing the full version of the model with a deep learning mechanism that will afford the dynamic creation of new nodes representing more abstract concepts, as well as the implementation of regulatory parameters and mechanisms. We will develop and execute a validation plan in collaboration with psychologists, in order to assess how realistic and reliable are the simulations with the RANs model.

Acknowledgements: This is work is part of our contribution to the project ConCreTe. The project ConCreTe acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 611733.

References

1. Anderson, J.R.: A spreading activation theory of memory. J. of Verbal Learning and Verbal Behavior 22(3), 261–295 (1983)
2. Bengio, Y., Courville, A.C., Vincent, P.: Representation learning: A review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. 35(8), 1798–1828 (2013)
3. Braver, T.S., Barch, D.M., Cohen, J.D.: Cognition and control in schizophrenia: A computational model of dopamine and prefrontal function. Bio. Psychiatry 46, 312–328 (1999)

4. Collins, A.M., Quillian, M.R.: Retrieval time from semantic memory. J. of Verbal Learning and Verbal Behavior 8(2), 240–247 (1969)
5. Crestani, F.: Application of spreading activation techniques in information retrieval. Artificial Intelligence Review 11, 453–482 (1997)
6. Erickson, R.P., Ohrwall, H., von Skramlik, E., Henning, H.: Ohrwall, Henning and von Skramlik; the foundations of the four primary positions in taste. Neurosci. Biobehav. Rev. 8(1), 105–127 (1984)
7. Rankin, C.H., et al.: Habituation revisited: an updated and revised description of the behavioral characteristics of habituation. Neurobiol. Learn. Mem. 92(2), 135–138 (2009)
8. Wittchen, H.U., et al.: The size and burden of mental disorders and other disorders of the brain in Europe 2010. European Neuropsychopharmacology 21(9), 655–679 (2011)
9. Gärdenfors, P.: Symbolic, conceptual and subconceptual representations. In: Human and Machine Perception: Information Fusion, pp. 255–270 (1997)
10. Gärdenfors, P.: Conceptual spaces as a framework for knowledge representation. Mind and Matter 2, 9–27 (2004)
11. Jacoby, L.L.: Perceptual enhancement: persistent effects of an experience. J. Exp. Psychol. Learn. Mem. Cogn. 9(1), 21–38 (1983)
12. Kyaga, S., Landén, M., Boman, M., Hultman, C.M., Långström, N., Lichtenstein, P.: Mental illness, suicide and creativity: 40-year prospective total population study. J. of Psychiatric Research 47(1), 83–90 (2013)
13. Matsukawa, J., Snodgrass, J.G., Doniger, G.M.: Conceptual versus perceptual priming in incomplete picture identification. J. Psycholinguist. Res. 34(6), 515–540 (2005)
14. McNamara, T.P., Altarriba, J.: Depth of spreading activation revisited: Semantic mediated priming occurs in lexical decisions. J. of Memory and Language 27(5), 545–559 (1988)
15. O'Reilly, R.C.: Biologically based computational models of high-level cognition. Science 314(5796), 91–94 (2006)
16. World Health Organization. Global burden of mental disorders and the need for a comprehensive, coordinated response from health and social sectors at the country level (December 2011)
17. Rankin, C.H., Broster, B.S.: Factors affecting habituation and recovery from habituation in the nematode Caenorhabditis elegans. Behav. Neurosci. 106(2), 239–249 (1992)
18. Roediger, H.L., Balota, D.A., Watson, J.M.: Spreading activation and arousal of false memories. In: The Nature of Remembering: Essays in Honor of Robert G. Crowder, pp. 95–115 (2001)
19. Roediger, H.L., Blaxton, T.A.: Effects of varying modality, surface features, and retention interval on priming in word-fragment completion. Mem. & Cognition 15(5), 379–388 (1987)
20. Roediger, H.L., Mcdermott, K.B.: Creating false memories: Remembering words not presented in lists. Journal of Experimental Psychology: Learning, Memory, and Cognition 21(4), 803–814 (1995)
21. Rolls, E., Loh, M., Deco, G., Winterer, G.: Computational models of schizophrenia and dopamine modulation in the prefrontal cortex. Nat. Rev. Neurosci. 9(9), 696–709 (2008)
22. Seamon, J.G., Luo, C.R., Gallo, D.A.: Creating false memories of words with or without recognition of list items: Evidence for nonconscious processes. Psychological Science 9(1), 20–26 (1998)
23. Wei, G., Yu, Z.: Storage capacity of letter recognition in hopfield networks