On the biological plausibility of artificial metaplasticity learning algorithm

Diego Andina a,⁎, Francisco J. Ropero-Peláez b

a Group for Automation in Signal and Communications, Technical University of Madrid, Spain
b Center for Mathematics Computation and Cognition, Universidade Federal do ABC, Brazil

1. Introduction

Artificial Metaplasticity (AMP) term was first introduced by Andina et al. [1] for an Artificial Neural Network (ANN) of the Multilayer Perceptron type (MLP), referred as AMMLP. During the AMMLP training phase, the matrix weight W that models the synaptic strength of its artificial neurons is updated according to the probability of the input patterns and therefore of the corresponding synaptic activations. The concept of biological metaplasticity was defined in 1996 by Abraham [2] and now is widely applied in the fields of biology, neuroscience, physiology, neurology and others [3–5]. The prefix “meta” comes from Greek and means “beyond” or “above”. In neuroscience and other fields “metaplasticity” indicates a higher level of plasticity, expressed as a change or transformation in the way synaptic efficacy is modified. Metaplasticity is defined as the induction of synaptic changes, that depends on prior synaptic activity [3,5]. Metaplasticity is due, at least in part, to variations in the level of postsynaptic depolarisation that induce synaptic changes. These variations facilitate synaptic potentiation and inhibit synaptic depression in depressed synapses (and vice versa in potentiated synapses). The direction and the degree of the synaptic alteration are functions of postsynaptic depolarisation during synaptic activation. Upregulation, incrementing, reinforcement of synaptic efficacy, is termed long-term potentiation (LTP), whereas downregulation, decrementing inhibiting, is known as long-term depression (LTD). LTP and LTD are believed to be fundamental to storage of memory in the brain and hence learning.

The induction of synaptic changes in the levels of neural activity is explained [6] in Fig. 1. Metaplasticity can be represented as variations in curve elongation with respect to the level of activity and implies a rightward shift of the LTP threshold $T_M$, $M = 0, 1, 2, \ldots$, (where curves cut to horizontal axis), according to the time-averaged level of postsynaptic firing $\pi_M$. More recent studies [22] showed that there are also LTD thresholds that diminish in the same circumstances. In summary, once synapses are positively primed (i.e. there is an increment in weight), the interval between thresholds broadens, thereby favouring subsequent synaptic depression. Understanding metaplasticity may yield new insights into how the modification of synapses is regulated and how information is stored by synapses in the brain [7–9].

Synaptic plasticity refers to the efficient modulation of information transmission between neurons, being related to the regulation of the number of ionic channels in synapses. Synaptic plasticity mechanisms involve both molecular and structural modifications that affect synaptic functioning, either enhancing or depressing neuronal transmission. They include redistribution of postsynaptic receptors, activation of intracellular signaling cascades, and formation/retraction of the dendrites [10]. The first model of synaptic plasticity was postulated by Hebb and it is known as the Hebb rule [11].

In Fig. 1, the effect of metaplasticity is illustrated. Each curve indicates the biological variation of synaptic weight, $ΔO$, respective of the neurons activation frequency or postsynaptic activity. If postsynaptic activity is high, by metaplasticity property, the curve will move, reinforcing the LTP. In Fig. 1, for different values of the time-averaged level of postsynaptic firing $\pi_M$, a family of
function to be minimised

\[ \sum_{i,j} w_{ij}(t) \]

The value of postsynaptic activation is crucial for determining the amount of variation in the synaptic weight, and therefore in the associated learning. In the BCM model, the LTP threshold is the same for all neuron synapses, so that metaplasticity would affect even non-active synapses (heterosynaptic plasticity).

curves is obtained. Curves corresponding to higher values of \( \pi_M \) have a higher LTP, so LTP metaplasticity is conventionally referred to as the rightward shift of the LTP threshold for higher synaptic weights.

The value of postsynaptic activation is crucial for determining the amount of variation in the synaptic weight and therefore in the associated learning [12,13]. Andina postulated [1] that high postsynaptic activity has to be related to high frequent excitations -frequent input classes in an artificial model-. In the way, the left-hand side curves of the family corresponding to low previous synaptic activity correspond to low frequent excitations produced by patterns of unfrequent classes.

2. Methods

The AMP implementation applied tries to improve results in learning convergence and performance by capturing information associated with significant rare events. It is based on the idea of modifying the ANN learning procedure such that un-frequent patterns which can contribute heavily to the performance, are considered with greater relevance during learning without changing the convergence of the error minimisation algorithm. It has been proposed on the hypothesis that biological metaplasticity property maybe significantly due to an adaptation of nature to extract more information from un-frequent patterns (low synaptic activity) that, according to Shannon’s Theorem [15], implicitly carry more information.

A way to implement and apply Artificial Metaplasticity (AMP) for a Multilayer Perceptron (MLP) trained by Backpropagation Algorithm (BP) was presented in [1] and successfully applied in many multidisciplinary applications [16,24,25]. In summary, a weighting function, \( f_{\alpha}^s(x) \), related to the distribution of the input vectors, \( x \in R^n \) (where \( R^n \) is the n-dimensional space, i.e. \( x = (x_1, x_2, \ldots, x_n) \), \( x_i \in R \), \( i = 1, 2, \ldots, n \)) is applied to the error function to be minimised

\[
E^\alpha(W(t)) = \frac{E[W(t)]}{f^s_{\alpha}(X)}
\]

Being \( W(t) \) the weight matrix in each training iteration, \( t \), and \( E[W(t)] \) the error function. \( f^s_{\alpha}(X) \) was estimated in MLP by taking advantage of the inherent distributions estimated by the MLP, or by using a gaussian function as approximation of the training pattern distribution. The weighted error \( E^\alpha(W) \) is then minimised for weights reinforcement in each iteration \( t \in N \). That is, if \( s, j, i \in N \) are the MLP layer, node and input counters, respectively, for each \( W(t) \) component, \( w^s_{ij}(t) \in R \), and being \( \eta \in R^+ \) the learning rate, then the weight reinforcement in each iteration is given by

\[
w^s_{ij}(t+1) = w^s_{ij}(t) - \eta \frac{\partial E^\alpha(W(t))}{\partial w^s_{ij}}
\]

\[
= w^s_{ij}(t) - \eta \frac{1}{\int_X} \frac{\partial E[W(t)]}{\partial w^s_{ij}}
\]

(2)

In general, if the learning strategy of an Artificial Neural Network (ANN) training is to minimise an expected error, the method can be generalised as illustrated in Fig. 2.

As the weighting function proposed depends only on the patterns distribution and not on the network parameters, the AMP algorithm can then be summarised as a weighting operation for updating each weight in each learning iteration as

\[
\Delta^\alpha w = w^\alpha(x) \Delta w
\]

(3)

being \( \Delta w = w(t+1) - w(t) \) the weight updating value obtained without AMP and \( w^\alpha(x) \) the realisation of the described weighting function \( w^\alpha(x) \) for each input training pattern \( x \). During the training phase, the AMP algorithm assigns higher values for updating the weights in the less probable activations than in the ones with higher probability and so it can be considered a new probabilistic version of the presynaptic rule [6].

3. Results

With the purpose of illustrating the potential of the AMMLP algorithm, two selected results in two very different applications are presented: Credit Risk Evaluation and Mammography Classification. The same databases as the compared algorithms and authors have been used. For credit risk evaluation (Table 1) the well-known Australian and German datasets were used [24]. For Mammography Classification, the also well-known Wisconsin Database [16] (Table 2).

Note that, presently, there is no point in comparing the performance of the AMMLP on multidisciplinary applications with other metaplasticity proposals, as those presented in [16]. They have been proposed to feedback to neurology models and...
metaplasticity would affect even non-active synapses (heterosynaptic plasticity). For its heterosynaptic nature, the proposed AMMLP follows closer to the BCM model. This fact is forced by the Backpropagation Algorithm, as weighting the delta rule does not change the direction of the gradient vector in the gradient descent algorithm, whereas a homosynaptic implementation would change its direction, resulting in artificial learning instability. Nevertheless, the AMP algorithm does not require the heterosynapticity of weight updating, only the relation between weight updating and probability of training patterns, so AMP homosynaptic models can also be proposed.

Note that biological models are in fact postulations, and despite BCM is widely accepted at the present as the closer to biological reality, it cannot be regarded to be the ultimate model of synaptic plasticity. According to Mockett and colleagues [22] metaplasticity is inherently a homosynaptic phenomenon in contrast to the heterosynaptic nature of the BCM rule. And Artola, Bröcher and Singer’s [23] extended model (ABS model), LTP and LTD thresholds shift to lower values for higher levels of the activation of neighbouring synapses (this last model is not analytical, as those just discussed, but is based on the empirical experimental data). We do not intend to determine the superiority of BCM model, but neurobiology inspires computer science and vice versa, and we report that the empirical results of AMP show a great potential, in terms of improving learning and therefore performance in most cases, no matter in what multidisciplinary application is applied, as illustrated in the results section.

The outstanding multidisciplinary application results may relay on the significant concept that the less frequent patterns induce more adaptation. Note that a repetition of similar patterns can bias the network towards one particular class. Placing greater emphasis on the diversity of patterns, rather than treating each pattern independently, may be responsible of improving the results. Specifically in the case of MLP, standard backpropagation can lead to gross errors at low probability, so this can partially explain why results are improved by the proposed AMMLP, as it can compensate Least Mean Squares low probability patterns penalisation. Based on this characteristic, other authors propose different error measures that result in a better estimate of the posterior probability, as the entropic ones [26], although their performance application results are not so universal as the AMMLP, failing in many applications.

Finally, note that this paper is focalised in the biological plausibility of the AMP algorithm, and it discuss it in terms of the activation regime as a parameter (7), instead of the synaptic weight w, what is preferred by many bio-researchers, as it makes it less questionable [27]. If the reader is interested in implementing and testing the AMMLP algorithm please refer to [16] that is focalised in the implementation of the algorithm, but does not enter to discuss many questions that may arise about its biological plausibility, hopefully solved in the present contribution. On the other hand, for the discussion of other metaplasticy implementations more close to biological reality but not efficient yet for multidisciplinary applications (as the incremental equation of the presynaptic rule or its probabilistic version) please refer to [27,28]. The Artificial Metaplasticity Multilayer Perceptron could be considered a new probabilistic version of the presynaptic rule [27], as it assigns higher values for updating the weights in the less probable activations than in the ones with higher probability.

5. Conclusion

We describe and discuss the biological plausibility of an artificial model of metaplasticity, a relevant property of biological neurons. The model name coined by the authors is “Artificial Metaplasticity”. Tested on different multidisciplinary applications, it achieves a more

### Table 1
Comparison classification accuracies obtained with our method and other classifiers from literature.

<table>
<thead>
<tr>
<th>Author (year, algorithm)</th>
<th>Australia</th>
<th>German</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ong et al. (2005, Genetic Programming (GP))</td>
<td>88.27</td>
<td>77.34</td>
</tr>
<tr>
<td>Huang et al. (2006, GP two-stage (2SGP))</td>
<td>89.17</td>
<td>79.49</td>
</tr>
<tr>
<td>Martens et al. (2007, Improved 2SGP)</td>
<td>85.70</td>
<td>NA</td>
</tr>
<tr>
<td>Martens et al. (2007, SVM)</td>
<td>86.70 ± 1.50</td>
<td>76.30 ± 1.60</td>
</tr>
<tr>
<td>Hoffman et al. (2007, Evolutionary Fuzzy Prog. (MCQVP))</td>
<td>85.60 ± 2.42</td>
<td>77.92 ± 3.97</td>
</tr>
<tr>
<td>Huang et al. (2007, GA-SVM)</td>
<td>86.70 ± 1.50</td>
<td>76.30 ± 1.60</td>
</tr>
<tr>
<td>Peng et al. (2008, Multicrit. Convex Quadratic</td>
<td>86.38</td>
<td>94.00</td>
</tr>
</tbody>
</table>

**Note:**
- NA, not applied.
- * Average obtained for 100 simulations.

### Table 2
Classification accuracies obtained with our method and other classifiers from literature.

<table>
<thead>
<tr>
<th>Author (year)</th>
<th>Method</th>
<th>Classification accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setiono (2000)</td>
<td>Neuro-rule 2a</td>
<td>98.10</td>
</tr>
<tr>
<td>Albrecht et al. (2002)</td>
<td>LSA machine</td>
<td>98.80</td>
</tr>
<tr>
<td>Abonyi and Szefert (2003)</td>
<td>SFC</td>
<td>95.57</td>
</tr>
<tr>
<td>Ubeyli (2007)</td>
<td>SVM</td>
<td>99.54</td>
</tr>
<tr>
<td>Polat and Güney (2007)</td>
<td>LS-SVM</td>
<td>98.53</td>
</tr>
<tr>
<td>Gújarro et al. (2007)</td>
<td>LLS</td>
<td>96.00</td>
</tr>
<tr>
<td>Akay (2009)</td>
<td>SVM-CFS</td>
<td>99.51</td>
</tr>
<tr>
<td>Karabatak and Cevdet (2009)</td>
<td>AR + NN</td>
<td>97.40</td>
</tr>
<tr>
<td>Peng et al. (2010)</td>
<td>CFV</td>
<td>0.997*</td>
</tr>
<tr>
<td>Conforti and Guido (2010)</td>
<td>SVM-SDP</td>
<td>96.79</td>
</tr>
<tr>
<td>This study (2010)</td>
<td>AMMLP</td>
<td>99.63b</td>
</tr>
<tr>
<td>This study (2010)</td>
<td>AMMLP</td>
<td>99.58b</td>
</tr>
</tbody>
</table>

* Result obtained in the AUC of ROC.
* Best result obtained in one simulation.
* Average obtained for 100 simulations.

...not for competitive performance results, so no relevant performance has been reported.

4. Discussion

Plasticity and Metaplasticity still pose a considerable challenge for research in terms of experimental design and interpretation [17]. Along the years, different mathematical models of synaptic computation have been proposed. In the classical Hebb model [14], the curve relating the increment of synaptic weight to postsynaptic activation is a straight line without synaptic depression. In Sejnowski’s covariance model [18,19], regions of potentiation and depression are separated by a LTP threshold. In [20] Abraham and Bear consider it as a homosynaptic property (i.e. involving only the synapse under study, without the need of considering the influence of nearby synapses), whereas the model by Bienenstock, Cooper and Munro [21] yields a curve that is closer to reality, with LTP threshold determined by postsynaptic activation and without LTD threshold. In the BCM model, the LTP threshold is the same for all neuron synapses, so that...
efficient training and improves Artificial Neural Network Performance. The model is closer to BCM heterosynaptic biological model than any other. During the training phase, the Artificial Metaplasticity Multilayer Perceptron could be considered a new probabilistic version of the presynaptic rule, as it assigns higher values for updating the weights in the less probable activations than in the ones with higher probability.

References


