Dina Vishnyakova

CONFIDENT ADAPTIVE RESONANCE THEORY

Abstract. This article is representing the idea of transparency in Adaptive Resonance Theory. Nowadays, there are different kinds of Neural Networks researches. They are different, with their own advantages and disadvantages, but all of them have one common goal. The goal is to achieve the ability ofreal nervous system to process information. One of artificial Neural Network realization came closer to this goal. It is Adaptive Resonance Theory (ART), the whole family of various neural networks. They realized plasticity-stability principle of perception. The attractive feature of the ART Neural Networks is that they preserve plasticity with the memorizing of the new images and at the same time they prevent the modification of the old images in memory. This Neural Network has a test on the comparison of the produced image with a content of the memory. One of the complicated tasks of building perfect Neural Networks besides solving of plasticity-stability dilemma - is a making the Neural Network more transparent and confident.

Neural Networks were traditionally connected with statistical data processing and were focused on classification accuracy. “An algorithm that tests with a higher percentage of correctly classified instances suits better than another to the bias of the application domain. The use of this measure is appropriate in some application domains e.g. pattern recognition where predictive accuracy is the main focus” [2]. In data mining, as well as in Neural Networks it is quite important to understand underlying decisions which are made by artificial intelligence model. Hence, appears such term as transparency. The problem is to realize and to measure this transparency in ART models. Transparency also helps to built trust between human beings and machine, in my case I raise problem of understanding of underlying processes in ART model and building trust.According to the Sociology science Trust is the belief in the good character of one party, presumed to seek to fulfil policies, ethical codes, law and their previous promises. A failure in trust will be forgiven more easily if it is interpreted as a failure of competence rather than a lack of benevolence. From this perspective, trust is an internal state, and cannot be measured. Only confidence, which is expressed as behavior, can be measured. Trust may be considered a moral choice. In this case, machine-human trust is meaningless, because computers have no moral sense.

There are different ways of making transparency. We can trust a bit more to Artificial Intelligence if we can see what is going on inside the ‘black box’. “An integrated approach for the realization of trust in IA by presenting crucial building blocks for trust namely - transparency of the embodied CI models, human interactivity in the learning process of the IA, stability of the acquired capabilities and security in interactions with humans and other agents. Previous attempts to make CI models transparent have focused on generation of comprehensible rules as explanation of the underlying reasoning process. Although this is an important factor in the
development of trust, it is just one of other factors needed to make a CI model more transparent and to provide justifications for its decision making process”.[2] Though, except this method of extracting rules there are more methods which help to make transparency, for example confidence parameter, or graphs for the visualization of knowledge hierarchy.

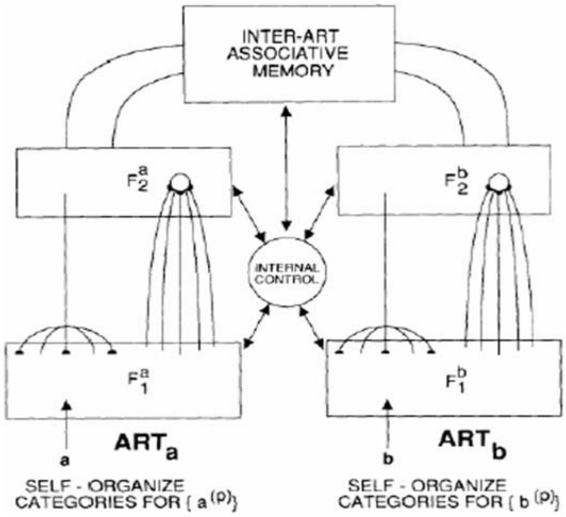

Fig.1. Abstract scheme of ARTMAP [1]

During my project time I worked on the topic of "Transparent Adaptive Resonance Theory". This was an attempt to improve on the ART work of Grossberg and Carpenter and develop a model with better utility and comprehensibility. The object of my research was ARTMAP (see Figure1) model, one of the models from the ART family. ARTMAP model, also known as Predictive ART, learns to associate arbitrary sequences of input/output pattern pairs. Consists of two modified ART1 models
into a supervised learning structure where the first unit takes the input data and the second unit takes the correct output data, then with matching parameter used to make the minimum possible adjustment of the vigilance parameter in the first unit in order to make the correct classification. "ARTMAP autonomously learns to classify arbitrarily many, arbitrarily ordered vectors into recognition categories based on predictive success. This supervised learning system is built up from a pair of Adaptive Resonance Theory modules that are capable of self-organizing stable recognition categories in response to arbitrary sequences of input patterns. These ART modules are linked by an associative learning network and an internal controller that ensures autonomous system operation in real time" [1].

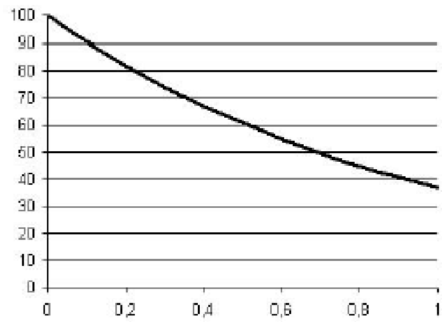

Fig.2. Graphic of exponential dependency

According to explanations of measuring Trust it is already known that confidence can be measured, because it express the behavior in current example it can express behavior of the ARTMAP during data classification. While analyzing the ARTMAP the main accent was emphasized on the question: “Which parameter plays role in modification system behavior?” hence, I retrieved algorithm and revealed that during training, ARTMAP make all decisions according to vigilance parameter and match tracking mechanism. Now, the question is how to connect confidence and match tracking. The dependency of confidence from vigilance parameter assume as a curve on the Figure 2, which represents exponential dependency. According to this graphic: axis Y is confidence from 0 to 100 in %, and axis X is a scale of the vigilance parameter ρ from 0 to 1. According to the fact that the variations of the vigilance parameter cause the result of modification of synapsis weights and result of the prediction. It means the greater parameter _ the less is confidence K, and vice verse, small value of parameter _ gives greater confidence. If the new vector doesn’t pass the match tracking check, than ARTMAP corrects vigilance parameter [3]. Confidence is applied in the training phase, and results will represent on how much percents ARTMAP was assure with data and modification its weights. For the testing of the ARTMAP classification and confidence parameter, I used 3 data sets: Iris (150 instances, 3 classes, 4 attributes), Cicada (248 instances, 3 classes, 5 attributes), Weather (14 instances, 2 classes, 4 attributes). The result without confidence:

DATA	**ARTMAP**	**distribut. ARTMAP**	**Multilayer Perceptron**	**Bayes Neural Net**
Iris	100%	100%	98.66%	94.66%
Cicada	91.34%	91.34%	70.19%	65.38%
Weather	100%	100%	100%	78.57%

According to experiments ARTMAP gives correct decisions between 51% - 100% with baseline vigilance equal to 0.001. If to analyze the results with confidence parameter than it is obvious according to output data that ARTMAP is giving correct predictions if confidence parameter greater than 51%. Hence, applying so called threshold to all input vectors which have confidence less than 51% should go again on processing I got following results:

DATA	**ARTMAP**	**distribut. ARTMAP**	**Confident ARTMAP**
Iris	100%	100%	100%
Cicada	91.34%	91.34%	94.23%
Weather	100%	100%	100%

Implementing confidence parameter to make system more transparent. Results of applying confidence function to the ARTMAP algorithm are represented in xml structure; where by simulator in % represents the degree of confidence which show on how much ARTMAP is assured in comparability of current input vector with Class. Results are represented in following structure:

<data>
<x1>71</x1>
<x2>91</x2>
<x3>1</x3>
<x4>0 </x4>
<class>2</class>
<conf>75.1%</conf>
</data>

Where <conf>75.1%</conf> represents confidence. After testing structure of the data is:

<data>
<x1>71</x1>
<x2>91</x2>
<x3>1</x3>
<x4>0 </x4>
<class>2,2</class>
<conf>75.1%</conf>
</data>

It means that originally data belong to Class 2 and ARTMAP put this data into Class 2 but was assure about this decision only on 75.1% The results produced by ARTMAP proves again all advantages of ARTMAP and also proves that it is possible to achieve confidence and to built trust, to improve results at least rely on the confidence parameter which is provided by ART.

REFERENCES

1. *Gail A. Carpenter, Stephen Grossberg, and J. H. Reynolds,* ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network, Neural Networks 4 (1991), 565–588.
2. *P.Owotoki,* Trustable agents and transparent computational intelligence models, Artificial Intelligence,2006.
3. *Dina Vishnyakova,* Transparent adaptive resonance theory, Tech. report, Hamburg University of Science and Technology, October 2006.

Н.В. Гладкова

ОБ ОДНОМ ПОДХОДЕ К СОЗДАНИЮ ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ*

Интеллектуальные информационные системы, такие как экспертные системы, системы обработки естественного языка и нейронные вычислительные системы увеличивают производительность и облегчают выполнение сложных задач. Они также обеспечивают поддержку, когда информационный поток неполный или «нечеткий» [1]. Интеллектуальные системы могут использоваться индивидуально, но во многих случаях они интегрированы между собой и с другими информационными системами.

Главной характеристикой интеллектуальных информационных систем является наличие в них, по крайней мере, одной модели. Основной задачей является осуществление анализа системы на заданной математической модели.

Также весьма эффективными на практике являются имитационные модели. Имитация в течение долгого времени является полезным методом, привносящим понимание в сложные ситуационные задачи оптимизации и принятия решений. Однако, техника имитации обычно не позволяет лицам, принимающим решения видеть, как решение сложной задачи развивается во времени. Также она не дает им способность взаимодействовать с этим решением. Имитационная техника дает только статистические ответы в конце множества специальных экспериментов. Как результат, ЛПР не являются неотъемлемой частью развития имитационного про-

* Работа выполнена при поддержке РФФИ, грант № 06 – 01 –00272 и программы развития научного потенциала высшей школы 2006-2008 гг. (РНП.2.1.2. 3193).