

ASIGNING RELEVANCIES TO INDIVIDUAL FEATURES FOR LARGE PATTERNS IN ARTMAP NETWORKS

COSOI, Alexandru Catalin; VLAD, Madalin Stefan & SGARCIU, Valentin

Abstract: Spam has become a global problem. Latest studies estimate that as much as 9 out of 10 emails are spam (Routers, 2006). Many solutions have been published so far, but every time a suitable solution is found, spam mutates into something new, so new ways to fight it must be found. A good method to fight spam at a proactive level would be the use of neural networks (Cosoi, 2006), but, as you will see in this paper, applying neural network theory per se is not enough.

Key words: ART, ARTMAP, spam, AntiSpam, heuristics

1. INTRODUCTION

Although sending billions of email messages advertising ridiculous products that most of us would never in our lives consider buying, what makes spamming profitable is its large volume. According to the New York Times, people click and buy products advertised in pharmaceutical spam emails. Other articles suggest that it costs about 300\$ to send 1 million emails. Assuming that a spammer makes just 25\$ from each sale (which is the lowest profit he can make), it's easy to see that it makes only slightly more than 2 million messages to make an immediate 10 000\$ profit (Beckman, 2007).

Over time, several techniques have been proposed to address this problem, like Bayesian Filtering Techniques (Graham, 2002), URL filtering, heuristic filtering, spam image filtering (Cosoi, 2006) and so on, but each time an acceptable solution was found, spam quickly mutated to something new and harder to catch.

Due to the fact that all the techniques enumerated above are all reactive, the need for a proactive solution is obvious. Heuristic filters look for patterns in the content of an email and match them against a database of known spam characteristics. These characteristics can be in the form of certain words, phrases, punctuation and altered dates. These are strong patterns and they match a single type of spam, offering zero false positives (a legitimate email mistakenly classified as spam), but the process of creating strong patterns is usually insidious and time consuming.

A good way to create strong patterns would be to use a neural network that combines short weaker patterns (if the email has words like "Viagra", "Valium", or if the date of the message is in the future and so on, which individually have a high false positive rate) and to use a neural network in order to combine these into stronger and longer patterns.

2. PROPOSED METHOD

A good neural network type up for this task would be ARTMAP networks (Cosoi, 2006). ARTMAP architectures are neural networks that develop stable recognition codes in real time into response to arbitrary sequences of input patterns. They were designed to solve the stability-plasticity dilemma that every intelligent machine learning system is facing: how to keep learning from new events without forgetting previously learned information. ARTMAP networks were designed to

accept binary or fuzzy input patterns (Carpenter & Grossberg, 1991). ARTMAP networks consist of two ART1 networks, ARTa and ARTb, bridged via an inter-ART module, as shown on Fig 1. An ART module has three layers: the input layer (F0), the comparison layer (F1), and the recognition layer (F2).

The neurons, or nodes, in the F2 layer represent input categories. The F1 and F2 layers interact with each other through weighted bottom-up and top-down connections, which are modified when the network learns. There are additional gain control signals in the network that regulate its operation.

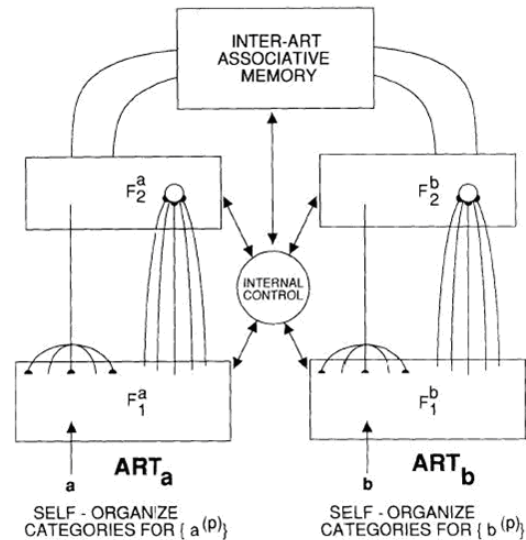


Fig. 1. ARTMAP system diagram (Carpenter & Grossberg, 1991)

In the training phase, the system has to receive a list of features extracted from the email messages and an output category. For example, ARTa will receive an input vector where each field indicates the existence of a certain spam or legitimate characteristic. Also, each input vector will be associated to a label which indicates if the current pattern was extracted from a spam or a legitimate email message, which will be fed to the ARTb module. When the training phase starts, the system will quickly associate inputs and outputs by creating strong patterns for each category.

The results are very good (Cosoi, 2006), with a false positive rate of almost 1% (which is not exactly the best yet obtained, but it can be rated among the top 3 AntiSpam filters) and a false positives rate (spam messages mistakenly misclassified as legitimate messages) under 10%. The problem that appears is that since the training phase is performed on a few million legitimate and spam messages samples, and since the individual heuristics are generally weak, the extracted patterns can be quite confusing for the neural network

algorithm. For example we can have a situation where important legitimate features and standard weak spam features can determine a mistakenly “this is spam” answer, and vice-versa.

These situations are generally determined by the large corpus of messages on which the neural network has to train in order to achieve an acceptable accuracy. In many situations, in our experiments, the training phase stopped after a fixed number of training iterations was achieved, and not when reaching a pre-established accuracy.

The solution we found to address this problem is to a priori offer a numerical relevance to each individual feature, and also the category (spam or legitimate) for which this feature was created. Our purpose was to create an inhibited connection, in order to stop the neural network to give an answer if the relevance of the pattern was smaller than a pre-established threshold T . Of course, this means that good hits would be eliminated to, but common-sense would say that we can't actually say an email is a spam message only because it contains the word “Viagra”.

If we consider I and S the relevance for the legitimate heuristics within a subset of a pattern and respectively S the relevance for the spam heuristics, we can combine them in a total relevance for a pattern by using the following simple rule:

$$R = \frac{1 - I + S}{2} \quad (1)$$

Where, I and S are computed as percents of the total sum of the relevancies within a pattern.

By using this result, the neural network can determine if this is an important pattern for the decision process or not. Of course, now this approach is more of a heuristic filter than a neural network. In order to keep all the facilities that a neural network would offer, (and we also chose this type of neural network in order to solve the stability-plasticity dilemma), we had to add a punishment-reward system in the control subsystem of the ARTa module (see Fig. 2). The process we developed is quite simple to explain. Each time the prediction matched the expectation we increased by a small amount the relevance of that pattern. If the prediction and the expectation were different, we decreased the relevance with a small amount. The process can be defined using the following formula.

$$R_{i+1} = (1 - w)R_i + w(R + (-1)^c \cdot \frac{w}{100}) \quad (2)$$

Where $(-1)^c$ has a negative value when the expectation and the prediction are different, and a positive one when the two are the same.

3. RESULTS

Our tests showed that by applying the improvements presented in this paper, the false positives rates dropped radically from an initial 1% to 1 in a million, while the false negative rate reached 13%, compared to an initial value of 10%. Although this method provides a slightly increase of the false negatives rate, it is far more important to prevent tagging as spam a legitimate email message than overlooking a few spam messages.

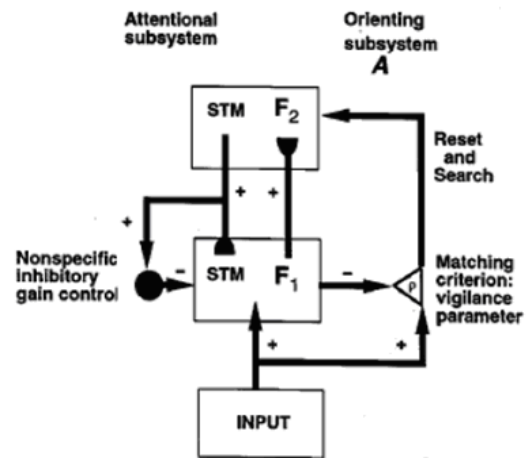


Fig. 2. ARTa system diagram (Carpenter & Grossberg, 1991)

The conditions in which the experiments took place are the following:

- 2.5 million spam messages
- Almost 1 million legitimate email messages
- 75% of the message corpus were used for training the neural network and,
- 25% were used in testing the neural network.

4. CONCLUSIONS

From the end-consumer point of view, not being able to receive crucial information (a salary increase or getting fired), represents a greater loss than having it's inbox bombed with another 3% of spam messages.

By using our approach, we minimize the risk of false positives which is an important problem in the AntiSpam community, although this solution has its side effects.

We are confident that this theory can also be applied to other neural network types, providing us a base for future research projects.

5. REFERENCES

- Beckman S. (2007). High-Performance Asynchronous IO for SMTP Multiplexing
Available from: <http://www.spamconference.org>
Accessed: 2007-03-31
- Cosoi, A. C. (2006). The medium or the message? Dealing with image spam.
Available from: <http://www.virusbtn.com>
Accessed: 2006-12-3
- Cosoi A. C. (2006). An AntiSpam filter based on adaptive neural networks
Available from: <http://www.spamconference.org>
Accessed: 2006-04-15
- Graham P. (2002). A plan for spam
Available from: <http://www.paulgraham.com/spam.html>
Accessed: 2007-05-27
- Carpenter, G. & Grossberg, S. (1991). Supervised real-time learning and classification of nonstationary data by a self-organizing neural network, In: *Pattern recognition by self organizing neural networks*, Carpenter, G. & Grossberg, S., (Ed. MIT press), 501-544, Publisher MIT press, ISBN 0-262-03176-0, Cambridge Massachusetts

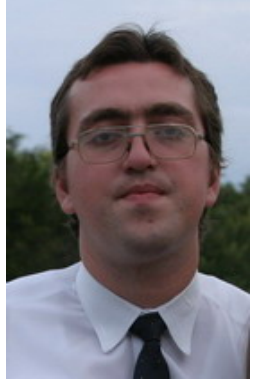
DAAAM AUTHOR QUESTIONNAIRE

PAPER DATA


Name and email address of corresponding author: Cosoi Alexandru Catalin, me@catalincosoi.com
This paper will be presented (oral presentation / poster): oral presentation
This paper will be NOT presented at Symposium please send the Proceedings to following address:
Please send PDF Offprints of Paper to following e-mail address:

AUTHORS DATA

1. Digital Photo (not to small):	
2. First / Middle / Family Name (Full names not initials only!): Catalin Alexandru Cosoi	
3. Academic Titles: student	
4. Position / Since: student, 2002	
5. Institution: University "Politehnica" Bucharest	
6. Place, Date and Country of Birth (yyyy-mm-dd): Buzau, 1983-02-11, Romania	
7. Nationality / Citizenship: Romanian	
8. Field of interests (key words): Brain models, neural networks, AntiSpam filters, AntiSpam technologies, pattern recognition and pattern extraction, cryptography, fractal (and chaos) theory, cognitive information processing, mathematics of neural systems, learning and memory, sensory-motor control and robotics, cognitive-emotional interactions, competitive learning, computational neuroscience, self-organizing maps, and decision making under risk	
9. Hobbies: tennis	
10. E-mail address: me@catalincosoi.com	
11. Site: www.catalincosoi.com	
12. Phone & Fax #: +40 742 586 994	
13. Postal address: 13'th Garleni street, C44, ap 24	
14. In wich DAAAM activities are you interested (We have many of possibilities such as: publishing of paper, to be active member of one of our international committees, official photograph of daaam international, reviewer of papers and manuscripts, supporter, sponsor, organizer. others). Please write your choice: publishing, reviewer	
15. Place & Date: Bucuresti, 2007-31-05	
16. 16. Additional CV data (optional): see www.catalincosoi.com	

17. Digital Photo (not to small):	
18. First / Middle / Family Name (Full names not initials only!): Madalin Stefan VLAD	
19. Academic Titles: Dr	

20. Position / Since: Researcher / Since 2006
21. Institution: University POLITEHNICA of Bucharest, Faculty of Automatic Control and Computer Science, ROMANIA
22. Place, Date and Country of Birth (yyyy-mm-dd): Bucharest, 1979-02-17, Romania
23. Nationality / Citizenship: Romanian/Romanian
24. Field of interests (key words): Smart card, RFID, Security
25. Hobbies: IT, Music
26. E-mail address: madalinv@ac.pub.ro
27. Site: www.ac.pub.ro
28. Phone & Fax #: 004 021 4029310 / 004 01 3181014
29. Postal address: Splaiul Independentei 313, Sector 6, Bucharest, zip code 060032
30. In wich DAAAM activities are you interested (We have many of possibilities such as: publishing of paper, to be active member of one of our international committees, official photograph of daaam international, reviewer of papers and manuscripts, supporter, sponsor, organizer. others). Please write your choice: publishing, reviewer
31. Place & Date: Bucharest 26 May 2007
32. 16. Additional CV data (optional):

	
33. Digital Photo (not to small):	
34. First / Middle / Surname (Full names not initials only!): Valentin SGARCIU	
35. Academic Titles: Prof.dr.eng.	
36. Position / Since: Professor of Department of Control and Industrial Informatics / 1995	
37. Institution: University POLITEHNICA of Bucharest, Faculty of Automatic Control and Computer Science, ROMANIA	
38. Place, Date and Country of Birth (yyyy-mm-dd): Clipicesti – Vrancea, Romania, 1949-06-01	
39. Nationality / Citizenship: Romanian	
40. Field of interests (key words): Reliability and diagnosis, Data processing, Sensors and transducers, Smart card applications, Intelligent buildings	
41. Hobbies: music (classic)	
42. E-mail address: vsgarcu@aii.pub.ro	
43. Site: www.acs.pub.ro	
44. Phone & Fax #: +40213181014	
45. Postal address: Splaiul Independentei 313, Sector 6, Bucharest, zip code 060032	
46. In wich DAAAM activities are you interested (We have many of possibilities such as: publishing of paper, to be active member of one of our international committees, official photograph of daaam international, reviewer of papers and manuscripts, supporter, sponsor, organizer. others). Please write your choice: publishing of paper	
47. Place & Date: Bucharest 26 May 2007	
48. Additional CV data (optional):	