BitDefender Antispam NeuNet

Livre blanc

Cosoi Alexandru Catalin
Researcher BitDefender AntiSpam Laboratory

Table des matières

1. Présentation du problème du spam

2. A propos de Neural Networks

3. Nouvelle structure articulée sur des réseaux neuronaux

4. Efficacité

Présentation du problème du spam

Ces dernières années, les utilisateurs du courrier électronique du monde entier ont constaté qu'un nombre croissant de messages non sollicités parvenaient dans leurs boîtes aux lettres. A ce jour, on a proposé plusieurs méthodes de filtrage destinées à traiter ce problème, telles que l'approche bayésienne, les listes noires/ listes blanches, l'image, le filtrage des URL, l'heuristique, etc. Le raisonnement sous-tendant toute technique de filtrage du spam (heuristique, probabiliste ou basée sur des mots clés) est le même : étant donné que les messages de spam présentent le plus souvent un aspect différent de celui des messages légitimes, la détection de ces différences constitue une bonne façon de les identifier et de les arrêter. A en juger par les résultats de ces méthodes de filtrage et du fait que le spam évolue de jour en jour, la meilleure manière de résoudre ce problème serait d'utiliser toutes ces fonctionnalités pour obtenir un effet combiné et plus précis.

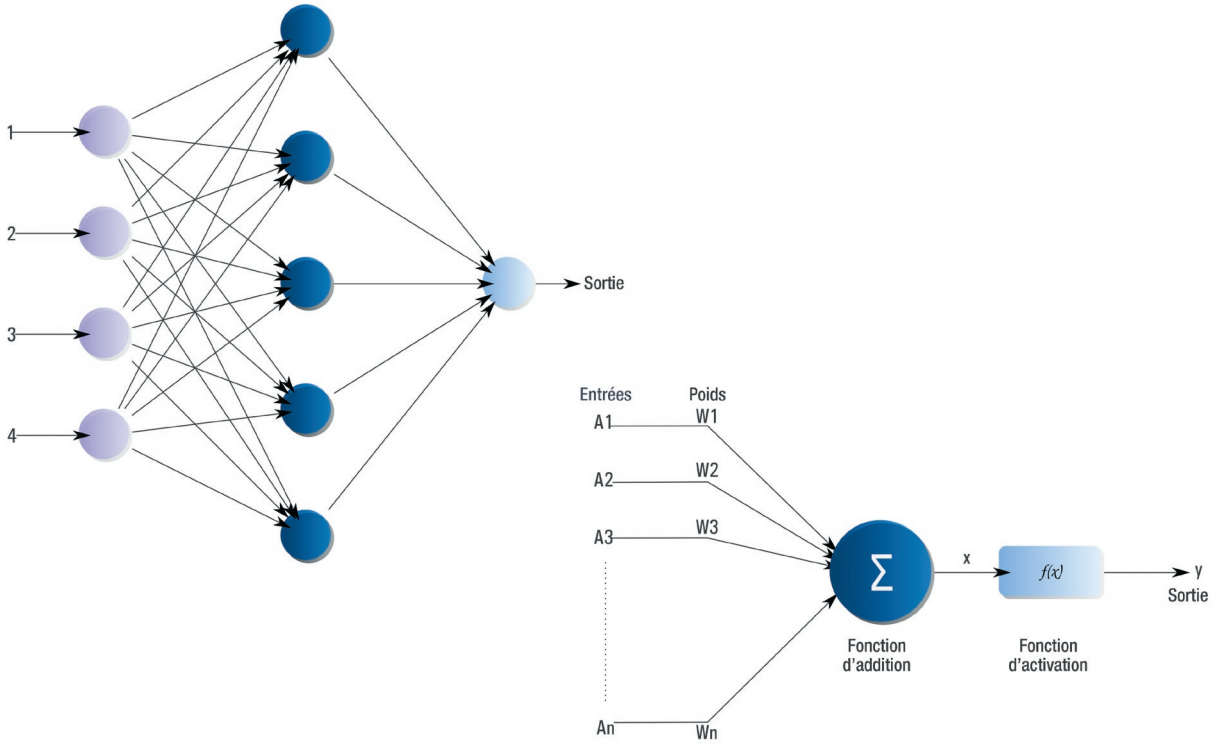
Plus facile à dire qu'à faire ! Depuis l'émergence de ces technologies, les spammeurs ont amélioré leurs techniques, pour permettre au spam de continuer à atteindre sa destination. Ils ont utilisé l'obscurcissement, en masquant les mots de manière que seul un humain les comprenne, tiré parti des failles des analyseurs html ou même masqué le contenu de manière qu'il soit pratiquement impossible à un ordinateur de faire la différence. Les solutions anti-spam ont dû accroître la fréquence des mises à jour et développer davantage d'heuristique en moins de temps. Le besoin de disposer d'un processus automatique apprenant rapidement les caractéristiques du nouveau spam sans nuire à la précision de la détection du spam moins récent est devenu vital. Pour nous attaquer à ce problème, nous nous sommes tournés vers les réseaux neuronaux artificiels.

A propos de Neural Networks

Un réseau neuronal artificiel peut également être considéré comme un paradigme de traitement de l'information s'inspirant de la façon dont des systèmes nerveux biologiques, tels que le cerveau, traitent des informations. Les réseaux neuronaux adoptent une approche différente de celle des ordinateurs conventionnels vis-à-vis de la résolution des problèmes.

Les ordinateurs conventionnels utilisent une approche algorithmique, dans laquelle l'ordinateur suit un ensemble d'instructions pour résoudre un problème. Si les étapes spécifiques nécessaires à l'ordinateur sont inconnues, celui-ci ne peut pas résoudre le problème. La capacité de résolution des problèmes des ordinateurs conventionnels se limite donc aux problèmes que nous comprenons déjà et savons résoudre. En outre, les ordinateurs conventionnels utilisent une approche cognitive de la résolution des problèmes : le mode de résolution du problème doit être connu et indiqué dans de petites instructions non ambiguës. Ces instructions sont ensuite converties en un langage de programmation de haut niveau, puis en code machine compréhensible à l'ordinateur.

Ces machines sont entièrement prévisibles ; si un problème survient, il est lié à un défaut logiciel ou matériel.
Les réseaux neuronaux et les ordinateurs algorithmiques conventionnels ne se font pas concurrence, mais se complètent. Certaines tâches, telles que les opérations arithmétiques, sont plus adaptées à une approche algorithmique, alors que d'autres nécessitent des réseaux neuronaux. Davantage de tâches encore réclament des systèmes adoptant une combinaison des deux approches (en principe, un ordinateur conventionnel réalise la supervision du réseau neuronal), pour un maximum d'efficacité.



On dit qu'un réseau neuronal apprend hors ligne si la phase d'apprentissage et la phase de fonctionnement sont distinctes, et qu'il apprend en ligne s'il apprend et fonctionne en même temps. Le plus souvent, l'apprentissage supervisé s'effectue hors ligne, tandis que l'apprentissage sans supervision s'effectue en ligne.

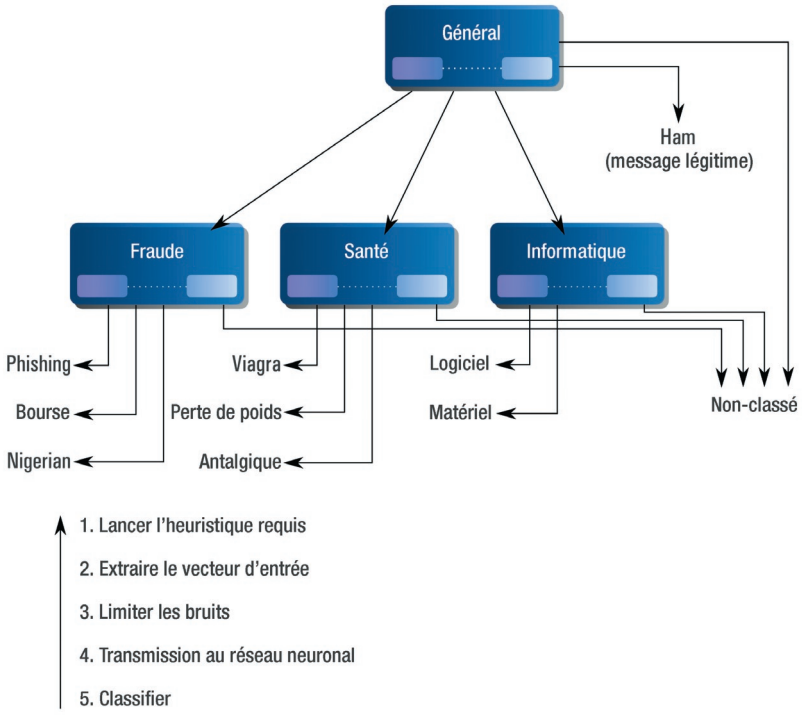
Nouvelle structure articulée sur des réseaux neuronaux

Notre idée était de créer un processus automatisé qui collecterait le corpus de spam et de ham (messages légitimes) sur une certaine période de temps, étudierait ses caractéristiques et apprendrait sans aucune intervention humaine. Plus ce processus est réalisé rapidement, plus la réaction est rapide.

Cependant, les réseaux neuronaux sont confrontés à certains problèmes. En cas de quantités élevées de données, les résultats tendent à chuter. Les réseaux neuronaux à action directe («feed forward») tendent à oublier certaines des informations apprises au début du processus, ou les données de sortie deviennent quelque peu chaotiques. A partir de ce constat, et du fait que le spam peut être séparé en plusieurs catégories distinctes, nous avons développé une arborescence de réseaux neuronaux permettant de classer de grandes quantités de données plus rapidement et sans nuire à la précision de la détection. Chacun des réseaux neuronaux inhérents à cette hiérarchie agit sur un type de spam différent, si bien que l'entrée et la sortie sont suffisamment limitées pour ne pas entraîner de confusion au sein du réseau et pour maintenir les performances à leur niveau le plus élevé.

Ainsi, nous avons créé une sous-catégorie appelée fraude, composée des messages visant à amener par la ruse l'utilisateur à envoyer de l'argent ou à révéler des informations sur sa carte bancaire.

En outre, nous avons identifié un sous-type de fraude, appelé «phishing», et ses variantes ingénieuses : il peut s'agir d'un spammeur tentant d'extorquer des informations sur la carte bancaire d'un utilisateur et se faisant passer pour un riche Nigérian qui cherche à sortir de l'argent du pays et a besoin que vous l'y aidiez en lui envoyant de l'argent ; du coup de la loterie, dans lequel le spammeur essaie d'obtenir des informations personnelles de votre part en vous annonçant que vous gagné plusieurs millions de dollars à la loterie ; ou encore de l'arnaque à la bourse, dans laquelle vous recevez des conseils sur la façon d'acheter des actions, etc. Naturellement, ces messages ont tous quelque chose en commun et peuvent constituer une catégorie à part entière.



Chaque sous-type possède ses propres caractéristiques qui le distinguent des autres, ce qui permet d'établir une sous-catégorisation supplémentaire.

Au cours de la formation, si le module de réseau neuronal ne trouve pas de catégorie dans laquelle classer un modèle particulier, il crée une catégorie entièrement nouvelle. Par conséquent, si un éventail trop vaste de spam est injecté dans le réseau neuronal, le nombre de catégories augmente, ce qui risque de ralentir l'analyse. Toutefois, si le réseau neuronal est spécialisé dans un type de spam unique, le rapide dépassement de capacité des catégories de sorties est évité même avec un nombre d'heuristiques accru, et l'analyse est plus précise et élaborée.

Supposons maintenant que la hiérarchie des réseaux neuronaux ait été formée et qu'elle soit prête à être testée. Dans ce cas, lorsqu'un courrier électronique arrive, le système doit fournir une réponse concernant sa nature : légitime, spam (un certain type) ou « ne sais pas » (réponse qui sera également considérée comme légitime pour éviter les faux positifs). Tout d'abord, le système exécute le type d'heuristique général sur le message, pour vérifier de quelle catégorie celui-ci relève. Si aucune catégorie correspondante n'est détectée, le message est considéré comme légitime ; autrement, il est transmis au réseau neuronal suivant qui traite ce type et l'algorithme se répète. A ce stade, des informations sont extraites, en quantité suffisante pour l'entrée dans le réseau neuronal en question. Si le niveau suivant atteint par le message correspond à un autre réseau neuronal, les informations sont transmises et l'algorithme se répète. Si le message ne peut pas être classé, il compte comme un message légitime. Cependant, si le niveau suivant est une catégorie finale (une feuille, dans notre arbre), ce message a été classé et le processus prend fin.

En conséquence, le processus fonctionne à partir d'une extraction sélective des informations, ce qui accélère l'analyse. De surcroît, l'approche fondée sur des réseaux neuronaux est plus élaborée et potentiellement beaucoup plus précise et fiable dans la réalisation de cette tâche.

Efficacité

Le taux de détection s'améliore de manière régulière avec l'ajout de nouvelles entrées, et peut facilement approcher (voire atteindre) 100 %. De même, le nombre d'heuristiques que l'on peut ajouter est infini, le tout sans devoir se préoccuper des performances en termes de délais d'analyse. En matière de détection, l'essentiel ne concerne pas le nombre d'heuristiques, mais les modèles qui sont découverts pendant la phase de formation. La présence d'un certain mot clé dans un courrier électronique ne signifie pas avec certitude qu'il s'agit de spam, mais ne signifie pas non plus qu'il ne s'agit pas de spam. Un modèle est une liste complète d'éléments clés présents dans le corps du courrier électronique, et le processus d'analyse peut être effectué même si un seul mot peut être considéré comme semblable à du spam. Si, pendant la formation, un courrier électronique de ce type pénètre dans le réseau neuronal, le processus d'analyse l'identifiera correctement.

Nos expériences montrent que l'approche fondée sur les réseaux neuronaux est plus élaborée, plus mathématique et potentiellement beaucoup plus précise et fiable dans la réalisation de cette tâche. A l'aide de ce seul filtre (BitDefender NeuNet - technologie faisant l'objet d'un brevet en instance), sur un ensemble de plus de deux millions de courriers électroniques (dont 80 % n'ont été utilisé que dans la formation et 20 % dans les tests), nous avons atteint une détection de 100 % sur le corpus de formation et de 97,56 % sur le corpus de test, et le système a fonctionné beaucoup plus rapidement qu'un filtre heuristique.

En conclusion, nous considérons que ce filtre franchit un nouveau palier dans la lutte contre le spam menée à l'aide de réseaux neuronaux.

A propos de BitDefender®

Les technologies antivirus BitDefender protègent aujourd'hui plusieurs dizaines de millions d'utilisateurs dans plus de 180 pays, directement ou via leur intégration dans des applications. Les moteurs de détection antivirus BitDefender sont certifiés par les organismes indépendants ICSA Labs, Checkmark, AV-Comparatives et Virus Bulletin. BitDefender a par ailleurs reçu le prix de l'innovation technologique décerné par la Commission Européenne (www.ist-prize.org). Les solutions Linux sont certifiées «RedHat Ready» et Mandriva. Les solutions BitDefender sont rééditées en exclusivité par Editions Profil sur les marchés francophones.

A propos d'Editions Profil

Editions Profil, société indépendante créée en 1989, développe, édite et diffuse des logiciels sur différents secteurs d'activités, professionnel et grand public. L'éditeur a constitué un large catalogue de solutions dans de nombreux domaines, par exemple sur les segments de la bureautique et de la productivité. Editions Profil s'est plus particulièrement spécialisée ces dernières années dans l'édition et la distribution d'outils de sécurité informatique et la protection des données en général. Editions Profil édite notamment les solutions de sécurité BitDefender.

Contacts

Pays : **France, Belgique, Luxembourg, Suisse, Pays-Bas, Afrique du Nord**
Société : **Editions Profil**
Adresse : 49, rue de la Vanne
92120 Montrouge
Tél. : (+33) 1 47 35 72 73
Fax : (+33) 1 47 35 07 09
Email : *bitdefender@editions-profil.eu*

ÉDITIONS PROFIL
www.editions-profil.eu

NEUNET

bitdefender