

Machine Learning



Autoria do Desafio Profissional: Lucas dos Santos Araújo Claudino

Leitora Crítica: Nathália dos Santos Silva Nolepa

► Proposta de Resolução

A resolução da situação proposta pode ser feita de várias maneiras, pois não existe somente um algoritmo a ser aplicado na filtragem *antispam*. Sendo assim, esta resolução apresentará alguns algoritmos. O importante é se ater às informações que serão fornecidas, como, por exemplo, funcionamento básico, necessidade de treinamento, acurácia etc. A resposta fornecida pelo aluno deverá também conter essas informações mínimas.


1. Descrever como funciona o aprendizado de máquina aplicado ao problema de filtragem *antispam*.

Para lidar com o problema de *spams* e *phishing*, os provedores de e-mail precisam utilizar técnicas de *machine learning* para identificar mensagens indesejáveis por meio da análise de milhares de mensagens similares. Os provedores que desejam uma boa proteção não podem ficar presos a uma simples observação dos e-mails usando regras preexistentes. Eles precisam gerar novas regras a partir do que o algoritmo aprende em seu processo de filtragem.

Uma aplicação de aprendizado de máquina que faz a filtragem de *spam* necessita, basicamente, seguir os seguintes passos:

- I. Preparar os dados de texto.
 - II. Criar um dicionário de palavras.
 - III. Processo de extração de aspectos/características.
 - IV. Treinar o classificador.
2. Propor ao menos duas técnicas de *machine learning* capazes de fazer essa classificação de e-mail.

Existem diversas categorias de técnicas que são aplicadas a esta tarefa. A seguir, é possível verificar as principais técnicas e alguns de seus algoritmos mais conhecidos:

- 
- Filtragem baseada em conteúdo: normalmente utilizada para criar regras de filtragem automática e classificar utilizando técnicas de AM do tipo: classificação Bayesiana (Naive Bayes), máquinas de vetor de suporte, K vizinhos próximos, redes neurais.
 - Filtragem heurística: um grande exemplo deste tipo de filtragem é o filtro *antispam* SpamAssassin.
 - Filtragem baseada em semelhança prévia: esta técnica busca filtrar os e-mails organizando-os em suas classes semelhantes a partir dos grupos obtidos com o treinamento. Geralmente é utilizado o algoritmo KNN para esta filtragem.
 - Filtragem baseada em lista: os tipos de filtros mais utilizados neste grupo são os filtros de *blacklist*, *whitelist* ou *graylist*.
3. Explicar o funcionamento dessas técnicas propostas e por que elas podem ajudar na classificação.
- **Filtragem heurística:** o filtro heurístico SpamAssassin é um sistema de inteligência artificial que consegue analisar toda mensagem que tenta entrar no servidor e, utilizando regras previamente determinadas, classifica o e-mail como *spam* ou não. Esta técnica trabalha com uma pontuação (variando de 0 a 10), que é dada a cada mensagem analisada. Se a pontuação dada pelo algoritmo não atingir a pontuação mínima pedida pelo usuário, o e-mail será considerado *spam*. Uma característica importante deste filtro é que as suas regras não são atualizadas automaticamente. Se o usuário desejar, ele precisa atualizar regularmente o SpamAssassin para que ele possa refazer o treinamento e então criar regras de filtragem mais modernas.
 - **Filtragem baseada em semelhança prévia:** esta metodologia de AM é baseada em semelhança, ou instância, para classificar novas mensagens de acordo com o nível de semelhança que elas possuem com aquelas mensagens já cadastradas em seu banco de dados. Os atributos do e-mail são utilizados para criar vetores multidimensionais, que são usados para plotar novas instâncias como pontos, e então relacionar aos vetores multidimensionais de treinamento. As novas instâncias são alocadas à classe mais semelhante, geralmente utilizando o algoritmo de K vizinhos próximos.

- **Filtragem baseada em lista:** os métodos de filtragem baseados em lista, basicamente, permitem ou não que mensagens de determinados remetentes cheguem ao seu destino. A *blacklist* é uma lista de IPs considerados maliciosos, e as mensagens provenientes de qualquer um desses endereços será bloqueada.
4. Explicar se essas técnicas precisarão de uma etapa de treinamento e o que é necessário fazer para manter o algoritmo sempre atualizado para classificar os novos tipos de *spam*.

Todas as técnicas citadas aqui precisam de uma etapa de treinamento. A **filtragem heurística**, mais especificamente o algoritmo SpamAssassin, utiliza o período de treinamento para criar as regras a serem usadas na pontuação. A **filtragem baseada em semelhança** precisa do treinamento para criar os vetores multidimensionais e então deixar o espaço preparado para receber novos pontos e poder utilizar o kNN para a classificação. A **filtragem baseada em lista** precisa do treinamento para atualizar a lista com os endereços/usuários/IPs permitidos ou não.



Bons estudos!