Existing Cate:

-Content Based Filtering Technique: Content based filtering is usually used to create automatic filtering rules and to classify emails using machine learning approaches, such as Naïve Bayesian classification, Support Vector Machine, K Nearest Neighbor, Neural Networks. This method normally analyses words, the occurrence, and distributions of words and phrases in the content of emails and used then use generated rules to filter the incoming email spams

-Case Base Spam Filtering Method: Case base or sample base filtering is one of the popular spam filtering methods. Firstly, all emails both non-spam and spam emails are extracted from each user's email using collection model. Subsequently, pre-processing steps are carried out to transform the email using client interface, feature extraction, and selection, grouping of email data, and evaluating the process. The data is then classified into two vector sets. Lastly, the machine learning algorithm is used to train datasets and test them to decide whether the incoming mails are spam or non-spam

-Heuristic or Rule Based Spam Filtering Technique: This approach uses already created rules or heuristics to assess a huge number of patterns which are usually regular expressions against a chosen message. Several similar patterns increase the score of a message. In contrast, it deducts from the score if any of the patterns did not correspond. Any message's score that surpasses a specific threshold is filtered as spam; else it is counted as valid. While some ranking rules do not change over time, other rules require constant updating to be able to cope effectively with the menace of spammers who continuously introduce new spam messages that can easily escape without been noticed from email filters. A good example of a rule based spam filter is SpamAssassin

-Previous Likeness Based Spam Filtering Technique: This approach uses memory-based, or instance-based, machine learning methods to classify incoming emails based to their resemblance to stored examples (e.g. training emails). The attributes of the email are used to create a multi-dimensional space vector, which is used to plot new instances as points. The new instances are afterward allocated to the most popular class of its K-closest training instances. This approach uses the k-nearest neighbor (kNN) for filtering spam emails.

-Adaptive Spam Filtering Technique: The method detects and filters spam by grouping them into different classes. It divides an email corpus into various groups, each group has an emblematic text. A comparison is made between each incoming email and each group, and a percentage of similarity is produced to decide the probable group the email belongs to.

SVM vs Log.Reg

Boundary vs Margin

(…)

Problem

* Image embedded
* Lack of security threat fliter(high usage of url in both spam and ham)
* Unable to handle drift problem
* Lack of meaningful development against ongoing development of spammers
* Lack of filters that have the capacity to dynamically update the feature space. Unable to incrementally add or delete features without re-creating the model totally to keep abreast of current trends in email spam filtering.
* While the amount of spammer has reduced by years but it still persist thus provide more samples, dimension,… thus make Deep Learning getting better as time go on compare to usual ML(Dimension, sample bound)

State of the art:

Perceptron Neural Networks (MLPNNs) and Radial Base Function Neural Networks (RBFNN)

* SVM