Abstract

Bloom filter is a probabilistic and space efficient data structure designed to check the membership of an element in a set. The trade-off to use Bloom filter may have configurable risk of false positives. The percentages of a false positive can be made low if the hash bit map is sufficiently massive. Spam is an unsolicited or irrelevant message sent on the internet to an outsized range of users or newsgroup. A spam word may be a list of well-known words that usually appear in spam mails. In the proposed system, Bin Bloom Filter (BBF) groups the words into number of bloom filters that have different false positive rates primarily based on the weights of the spam words. Clonal Selection Algorithm is one of the methods in Artificial Immune System (AIS) involved with computational methods inspired by the process of the biological immune system. This paper demonstrates the CSA algorithm for minimizing the total membership invalidation cost of the BBF which finds the optimal false positive rates and number of elements to be stored in bloom filters of Bin. The experimental results demonstrate the application of CSA in BBF and compare the results with Genetic Algorithm (GA).

Index Terms

Computer Science  
Artificial Intelligence
Keywords
Clonal Selection Method  Bloom Filter  Spam Filter  False Positive Rate  Hash Function