

発表概要

射影積算法による
劣線形メンバシップサマリの構築に向けて山本 泰生^{1,a)}

2020年1月15日発表

要素の集合 S に対し、ある要素 q が $q \in S$ であるかどうか答えるデータ構造を S のメンバシップサマリ (*membership summary*, 以降 MS と略す) と呼ぶ。本発表では、判定の誤り発生率を確率的に抑制する枠組みの下で、領域計算量が $O(|S|)$ 未満となる劣線形メンバシップサマリを構築する手法を検討する。偽陽性を許容する確率的メンバシップサマリとして、ブルームフィルタ (*Bloom filter*) がよく知られる。ブロックチェーンの SPV ノード管理など幅広い用途に利用される反面、領域計算量は $O(|n|)$ であり、データ数に対し、メモリ消費が線形的に増加する課題がある。本研究では偽陽性と偽陰性の2つを許容するよう制約を緩和したうえでメンバシップサマリの劣線形化を試みる。本発表では、キーとなる一様独立なハッシュ関数による高次ベクトルへの射影 (Projection) とそれらの積算ベクトルを用いた判定法について述べ、集中不等式を用いて誤り発生率の上界を明らかにする。

Presentation Abstract

Towards a Construction of Sub-linear Membership Summary with
Projection SummationYOSHITAKA YAMAMOTO^{1,a)}

Presented: January 15, 2020

Here, we investigate so-called “membership summary” of a given set of elements S . A membership summary is defined as a memory-efficient data structure which enables to answer the membership query if $q \in S$ for any element q . *Bloom filter* is such a well-known summary with a variety of applications, such as SPV node construction in block-chain. In modern bigdata processing point of view, however, Bloom filter may face a fundamental drawback on the memory consumption, that is $O(n)$ space for the element number n . In this talk, we present a relaxation approach to construct a sublinear membership summary, where both false positives and false negatives can happen. The key idea lies in projection technique with pair-wise uniform and independent hash functions. Each element of S is randomly projected to a high-dimensional vector. We then show the summation of those vectors from S can be used to answer the membership query. Finally, bound of false positive and false negative occurrence rates would be revealed based on concentration inequalities.

This is the abstract of an unrefereed presentation, and it should not preclude subsequent publication.

¹ 静岡大学
Department of Computer Science, Shizuoka University,
Hamamatsu, Shizuoka 432-8011, Japan

a) yyamamoto@inf.shizuoka.ac.jp