سمینار

تمرین چهارم

مینا فریدی
۸۱۰۱۰۰۴۳۰

سوال اول

بخش‌های چکیده مقاله‌هایی که دراختیار ما قرار گرفته را بر اساس رنگ‌های زیر مشخص می‌کنیم:

Background
Perpose
Methodology
Results
Conclusion

مقاله اول

Consider the following problem: We have k players each receiving a stream of items, and communicating with a central coordinator. Let the multiset of items received by player i up until time t be Ai(t). The coordinator's task is to monitor a given function f computed over the union of the inputs ∪i Ai(t), continuously at all times t. The goal is to minimize the number of bits communicated between the players and the coordinator. Of interest is the approximate version where the coordinator outputs 1 if f ≥ τ and 0 if f≤ (1−ϵ)τ. This defines the (k,f,τ,ϵ) distributed functional monitoring problem. Functional monitoring problems are fundamental in distributed systems, in particular sensor networks, where we must minimize communication; they also connect to the well-studied streaming model and communication complexity. Yet few formal bounds are known for functional monitoring.

We give upper and lower bounds for the (k,f,τ,ϵ) problem for some of the basic f's. In particular, we study the frequency moments Fp for p=0,1,2. For F0 and F1, we obtain monitoring algorithms with cost almost the same as algorithms that compute the function for a single instance of time. However, for F2 the monitoring problem seems to be much harder than computing the function for a single time instance. We give a carefully constructed multiround algorithm that uses “sketch summaries” at multiple levels of details and solves the (k,F2,τ,ϵ) problem with communication Õ(k2/ϵ + k3/2/ϵ3). Our algorithmic techniques are likely to be useful for other functional monitoring problems as well.

بخش چکیده در این مقاله همه قسمت‌ها را دارا است.

مقاله دوم

The problem of building an `0-sampler is to sample nearuniformly from the support set of a dynamic multiset. This problem has a variety of applications within data analysis, computational geometry and graph algorithms. In this paper, we abstract a set of steps for building an `0-sampler, based on sampling, recovery and selection. We analyze the implementation of an `0-sampler within this framework, and show how prior constructions of `0-samplers can all be expressed in terms of these steps. Our experimental contribution is to provide a first detailed study of the accuracy and computational cost of `0-samplers.

تنها بخشی که این چکیده کوتاه ندارد جمع‌بندی است.

مقاله سوم

We introduce a new sublinear space data structure—the count-min sketch—for summarizing data streams. Our sketch allows fundamental queries in data stream summarization such as point, range, and inner product queries to be approximately answered very quickly; in addition, it can be applied to solve several important problems in data streams such as finding quantiles, frequent items, etc. The time and space bounds we show for using the CM sketch to solve these problems significantly improve those previously known—typically from 1/ε2 to 1/ε in factor.

بخش‌های Background و Conclusion در این مقاله نیست.

مقاله چهارم

DBSCAN is a popular method for clustering multi-dimensional objects. Just as notable as the method's vast success is the research community's quest for its efficient computation. The original KDD'96 paper claimed an algorithm with O(n log n) running time, where n is the number of objects. Unfortunately, this is a mis-claim; and that algorithm actually requires O(n2) time. There has been a

fix in 2D space, where a genuine O(n log n)-time algorithm has been found. Looking for a fix for dimensionality d ≥ 3 is currently an important open problem.

In this paper, we prove that for d ≥ 3, the DBSCAN problem requires Ω(n4/3) time to solve, unless very significant breakthroughs---ones widely believed to be impossible---could be made in theoretical computer science. This (i) explains why the community's search for fixing the aforementioned mis-claim has been futile for d ≥ 3, and (ii) indicates (sadly) that all DBSCAN algorithms must be intolerably slow even on moderately large n in practice. Surprisingly, we show that the running time can be dramatically brought down to O(n) in expectation regardless of the dimensionality d, as soon as slight inaccuracy in the clustering results is permitted. We formalize our findings into the new notion of ρ-approximate DBSCAN, which we believe should replace DBSCAN on big data due to the latter's computational intractability.

این چکیده هم قسمت بک‌گراند را که اجباری نیست ندارد.

مقاله پنجم

In the context of the sliding-window set membership problem, and caching policies that require knowledge of item recency, we formalize the problem of Recency on a stream. Informally, the query asks, “when was the last time I saw item x?” Existing structures, such as hash tables, can support a recency query by augmenting item occurrences with timestamps. To support recency queries on a window of W items, this might require Θ(W log W) bits. We propose a succinct data structure for Recency. By combining sliding-window dictionaries in a hierarchical structure, and careful design of the underlying hash tables, we achieve a data structure that returns a 1 + ε approximation to the recency of every item in O(log(εW)) time, in only (1 + o(1))(1 + ε)(B + W log(ε −1)) bits. Here, B is the information-theoretic lower bound on the number of bits for a set of size W, in a universe of cardinality N.

این مقاله قسمت جمع بندی را در بخش چکیده خود ندارد.

مقاله ششم

۶-۱

In this paper, we study the (approximate) Range Thresholding (RT) problem over streams. Each stream element is a d-dimensional point and with a positive integer weight. An RT query q specifies a d-dimensional axis-parallel rectangular range $R(q)$ and a positive integer threshold $\tau(q)$. Once the query q is registered in the system, define $s(q)$ as the total weight of the elements that satisfy: (i) they arrive after q's registration, and (ii) they fall in the range $R(q)$. The task of the system is to capture the *first moment* when $s(q) \geq \tau(q)$. In addition, it admits a more general approximate version: given a real number $0 < \varepsilon < 1$, the task is to capture an arbitrary moment during the period between the first moment when $s(q) \geq (1 - \varepsilon) \cdot \tau(q)$ and the first moment when $s(q) \geq \tau(q)$. The challenge is to support a large number of RT queries simultaneously while achieving *sub-quadratic* overall running time.

We propose a new algorithm called *FastRTS*, which can reduce the exponent in the poly-logarithmic factor of the state-of-the-art *QGT* algorithm from $d + 1$ to d, yet slightly increasing the log term itself. A crucial technique to make this happen is our *bucketing technique*, which eliminates the logarithmic factor caused by the use of heaps in *QGT* algorithm. Moreover, we propose two extremely effective optimization techniques which significantly improve the performance of *FastRTS* by orders of magnitude in terms of both running time and space consumption. Experimental results show that *FastRTS* outperforms the competitors by up to *three* orders of magnitude in both running time and peak memory usage.

این مقاله بخش جمع‌بندی را ندارد. چون فایل مقاله از طریق لینکی که در صورت سوال تعبیه شده بود قابل دسترس نبود برای من این مقاله را با ادیتور علامت زدم که زیاد زیبا نیست و از این بابت پوزش.

سوال دوم

عنوان مقاله از قسمت‌های تاثیر گذاری اولیه مقاله است و به همین دلیل از حساسیت بالایی برخوردار است.

ویژگی‌های یک عنوان خوب برای مقاله عبارت اند از:

- دقت محدوده پوشش دهنده
- اندازه مناسب (نه کوتاه و نه بلند)
- هدایت خواننده به سمت مطلب
- عدم استفاده از قیود کیفی بلکه توصیفی
- عدم استفاده از لغات بسیار تخصصی
- مختصر و فشرده
- همخوانی عنوان و چکیده با محل چاپ مقاله

اجزای مهمی که در عنوان مقاله بر اساس مطالب درس باید وجود داشته باشند موارد زیر است که این موارد را در هر عنوان از مقالات تمرین با رنگ متمایز نشان می‌دهیم و در صورت لزوم بر اساس اصولی که در درس به آنها پرداخته شد عناوین را اصلاح می‌کنیم.

Product
Objective
Method
Context

Architecture Anti-patterns: Automatically Detectable Violations of Design Principles

این مقاله بخش هدف ندارد و نمیتوانیم به ان بیفزاییم.

A Light-Weighted CNN Model for Wafer Structural Defect Detection

در این چکیده کانتکست کار ذکر نشده است. مثلا یک حالت اصلاح شده برای این مقاله میتواند نمونه زیر باشد:

A Light-Weighted CNN Model for Wafer Structural Defect Detection *for Earthquake resistance*

An Artificial-Intelligence-Driven Predictive Model for Surface Defect Detections in Medical MEMS

این مقاله مورد مشکل داری ندارد و بخش هایی که استاد گفتند را دارا است.

Covering orthogonal polygons with sliding k-transmitters

عنوان این مقاله خیلی کوتاه و ناکافی است. برای این که بتوانیم عنوان بهتر پیشنهاد بدهیم باید مقاله را کامل مطالعه کنیم اما اگر بخش هایی که ندارد مثل هدف و همبافت را به آن اضافه کنیم میتوان ان را بهبود داد

Proactive Deployment of Aerial Drones for Coverage over Very Uneven Terrains: A Version of the 3D Art Gallery Problem

این مقاله عنوان بدی ندارد فقط یکم طولانی است. تنها بخشی که در این عنوان وجود ندارد روش است که اگر روش هم ذکر می‌شد عنوان طولانی‌تر شده که مطلوب نیست.

Type-based analysis of logarithmic amortised complexity

با وجود نداشتن بخش هدف این عنوان در عین کوتاه بودن به نظر خوب است

Bounded Expectations: Resource Analysis for Probabilistic Programs

عنوان این مقاله هم مناسب است و ایراد خاصی درش نمی‌بینم. تنها بخش متد را ندارد که لزومی بر ذکرش در این مقاله به نظر نمی‌رسد که حس شود.

A Simple and Scalable Static Analysis for Bound Analysis and Amortized Complexity Analysis

این مقاله هم با وجود نداشتن بخش روش در عنوان خود قابل قبول است.

Complexity and Resource Bound Analysis of Imperative Programs Using Difference Constraints

عنوان مقاله مناسب و کامل است.

Beyond Symbolic Heaps: Deciding Separation Logic with Inductive Definitions

عدم وجود بخش کانتکست باعث بد بودن عنوان این مقاله نیست.

ارزیابی میزان تطابق راهکارهاي مدل سازي موضوعی بر پایگاه هاي داده تحت وب گراف محور متن کوتاه پویا

این مقاله بخش متد را ندارد.

تشخیص متون توهین آمیز در موتورهاي جستجو با استفاده از یادگیري ماشین

عنوان کوتاه و کاملی است که بخش محصول را ندارد.

یک رهیافت خودتطبیق پذیر مدیریت منابع در محیط هاي رایانش ابري

توضیحات در مورد رهیافت کم و ناقص است. بجز این مورد مشکل دیگری ندارد.

بهبود عملکرد سیستم هاي توصیه گر مبتنی بر تکنولوژي بلاك چین

یک عنوان ناقص که متد را نگفته و بجز این هم خیلی خیلی کوچک است که از روی ان اطلاعات کمی نصیب مخاطب می‌شود.

بهبود سیستم هاي توصیه گر مبتنی بر برچسب با استفاده از آنتولوژي

عنوان مختصر و کاملی است که همه قسمت های عنوان را پوشش داده.

بررسی چالش هاي استقرار سرويس داده کاوي در حوزه سيستم هاي اطلاعات سلامت

متد انجام کار در این هوان ذکر نشده، مثلا میتوانست بگوید:

بررسی موردی چالش هاي استقرار سرويس داده کاوي در حوزه سيستم هاي اطلاعات سلامت

ارائه يک روش فرا ابتکاري براي حل مشکل اقلام دنباله طولانی در سيستم هاي توصيه گر

عنوان این مقاله خوب است.

ارائه يک مدل ترکيبی داده کاوي جهت بررسی شکست و يا موفقيت استارتاپ هاي ايرانی با انتخاب ويژگی و طبقه بندي

بخش عنوان این مقاله همه قسمت ها را دارد.

ارزيابی ميزان تطابق راهکارهاي مدل سازي موضوعی بر پايگاه هاي داده تحت وب گراف محور متن کوتاه پويا

بخش روش در این عنوان وجود ندارد.

يک روش مبتنی بر انتخاب ويژگی با الگوريتم بهينه سازي پروانه به منظور پيش بينی زمان اجراي Job هاي مبتنی بر نگاشت-کاهش

بکار بردن لغت انگلیسی در عنوان فارسی خوب نیست و باید معادل فارسی برای این کلمه بکار میرفت.