# CS848 Paper Submission Site

Go To CS848 Site

Submit or Update A Review For Paper #19

Go To CS848 Submissions

It is currently Thursday 11th of November 2004 05:42:07 PM EST

| Paper # 19 (Download paper of type application/pdf, 1586224 bytes) | |
|---|---|
| Title: | Selectivity Estimation Without the Attribute Value Independence Assumption |
| Abstract: | |

**You have already finalized your review for this paper.**
**You can no longer modify it, but you may view it.**

**If you made a mistake in your review and you want it "unfinalized",**
**you may send mail to the program chair asking them to unfinalize it**

Send yourself this review by email

| Attribute | Value |
|---|---|
| Are you finished with this review? | Finalize, I am done editing |
| Provide a short summary of the paper | THis paper presents two methods of estimating selectivities when queries rely on selections from dependent column data. The authors begin with a quick overview of the motivation for their work, as well as a brief outline of the solutions they will present. Some related work is also presented.<br><br>The authors next present the notation used in their paper, as well as an introduction to the attribute value independence assumption. Definitions and examples are given. The idea of a histogram is introduced and explained in great detail.<br><br>The authors next begin discussing multidimensional histograms (MH). A MH on two attribute columns first partitions the data on one column, generating a one dimensional histogram. Then, the data within each bucket is further partitioned into a second set of buckets, thus extending the histogram into a second dimension. A selectivity estimate can be made more accurate if it is performed using this MH on these predicate columns. MH's can theoretically be extended into an arbitrary number of dimensions. The authors present an example to illustrate the concept further before presenting some different methods of MH construction.<br><br>The authors then move to discuss the method of singular value decomposition SVD to estimate dependent column selectivities. After presenting some mathimatical background, the authors discuss briefly a technique for estimating selectivities from the resultant frequency matrix.<br><br>The authors present a brief performance analysis as well as some concluding remarks. |
| What is the strength of the paper? (1-3 sentences) | This paper details a few methods for calculating selectivities between dependent columns. The authors' writing style is, for the most part, quite clear (see weaknesses). The performance evaluation presented is fairly comprehensive. |
| | Missing are methods of developing the SVD matrix. The reader is refered to another |

| | |
|---|---|
| What is the weakness of the paper? (1-3 sentences) | paper, but the methods should have been presented in this paper for completeness. Also, the mathematical and notational models are confusing and not intuitive. For example, k is meant to be less than j in one formula, but j comes before k in the alphabet. |
| Your qualifications to review this paper | I know the material, but am not an expert |
| Writing Quality | Good |
| Relevance to query processing? | The paper is relevant to query processing |
| **Experimental Methodology** | **Good** |
| Novelty of paper | This is a new contribution to an established area |
| **Overall paper merit** | **A novel or new contribution to this area with good methodology, or an incremental contribution paper that has excellent methodology. A must read for anyone in the area.** |
| In your opinion, will this paper be important over time? | Good |
| Provide additional detailed comments to the author | You have presented two very good methods of calculating selectivities for dependent columns. Your writing style is generally quite good. However, some discussions of the mathematical concepts and foundations was confusing and unintuitive.For example, k is meant to be less than j in one formula, but j comes before k in the alphabet. More care should be taken in future writing to be clearer and present your material at a technical level suitable for all readers.<br><br>I thought your performance anaylsis was quite comprehensive. However, it would have been useful to see the overhead and accuracy of your methods when the relation size is scaled up (say to one million tuples, etc). As well, how does your estimate perform when there is no dependency between columns?<br><br>Missing from your paper are methods of developing the SVD matrix. The reader is refered to another paper, but the methods should have been presented in this paper for completeness.<br><br>Overall, your work is quite good. However, in future, please fix the above deficiencies. |
| **Additional comments to PC (not seen by author)** | **This paper presents novel work and should be published. I do not belive the deficiencies I outline to the authors warrant rejection from publication.** |

Goto Main Index  Close Window