View Reviews

Paper ID
495

Paper Title
Using Metric Space Indexing for Complete and Efficient Record Linkage

REVIEWER #1

REVIEW QUESTIONS

1. Please rate your overall rating for this paper.
Weak Reject

2. Please list three, or more, strong aspects of this paper. Key aspects that you might want to consider are in terms research quality and contribution, originality, technical novelty as well as novelty in the application and/or problems being solved.
1. The paper is easy to read.
2. Contributions are clearly claimed.
3. Experiments are conducted to evaluate the performance of the proposed method.

3. Please list three, or more, weak aspects of this paper.
1. The diagrams in Fig. 2 are hardly readable/distinguishable if printed in B/W.
2. References are not recent. The LSH method chosen for comparison with the proposed method was published in 1997 [2].
3. The proposed M-tree method does not really outperforms LSH in both the run time and accuracy measures.

4. Please enter your detailed comments here. Please be critical, polite as well as constructive.
Table 2, on the Cora dataset, shows that M-tree has a bit lower F-measures than LSH-5-2 and LSH-10-2 (0.68 vs. 0.69) and requires more comparison.

Tables 3 and 4, on the other two datasets, show that M-tree gives higher F-measures than LSH but also requires more comparisons.

REVIEWER #2

REVIEW QUESTIONS

1. Please rate your overall rating for this paper.
Strong Accept

2. Please list three, or more, strong aspects of this paper. Key aspects that you might want to consider are in terms research quality and contribution, originality, technical novelty as well as novelty in the application and/or problems being solved.
The authors propose using metric space indexing techniques for record linkage. In this paper, one of the metric space indexing structure, M-tree, is used for investigation. The proposed approach is compared against three indexing methods including brute force, traditional blocking and LSH. Experimental results shows that metric space indexing techniques achieve high quality and efficient record linkage without the need of complex parameter tuning.

The paper is well written and easy to follow. The novel application of MSI to record linkage has the advantages of simple parameter tuning. The experiment demonstrated the efficacy of MSI for record linkage.

3. Please list three, or more, weak aspects of this paper.
I feel this paper should be accepted as it is

4. Please enter your detailed comments here. Please be critical, polite as well as constructive.
I feel this paper should be accepted as it is

REVIEWER #3

REVIEW QUESTIONS

1. Please rate your overall rating for this paper.
Accept

2. Please list three, or more, strong aspects of this paper. Key aspects that you might want to consider are in terms research quality and contribution, originality, technical novelty as well as novelty in the application and/or problems being solved.
1. The paper is well written.
2. The experimental setup is well done.

3. Please list three, or more, weak aspects of this paper.
1. Some design choices (e.g., similarity measure) could be discussed and motivated more.

4. Please enter your detailed comments here. Please be critical, polite as well as constructive.
The paper proposes a new indexing technique for record linkage. The advantage of the new proposed technique is that it has less parameters and can perform record linkage in a single step, although this is true only for a threshold-based record linkage system. The idea is interesting and can lead to some advances in record linkage, but there are still some questions that the authors should address. How does the similarity measure chosen affects the overall performance? Why did you use edit distance as the similarity measure? Does it make sense to sum the edit distances over all the attributes? what about other weighting schemes? Although selecting parameters is generally hard, LSH seems to give good results with less comparisons and even the authors state that LSH was not particularly sensitive to one of the parameters. It makes one wonder how difficult it is to set those parameters for LSH.