Result explaination:



1. Original type of data

From the paper Deep Learning for Entity Matching: A Design Space Exploration, we can find the source of the initial data, and the original types of data are structured data and textual data. The abt\_buy is textual data, the amazon\_google is structured data, and the dirty\_amazon\_itunes, dirty\_dblp\_acm, dirty\_dblp\_scholar, dirty\_walmart\_amazon are structured data, which has been specially processed to become dirty data.

From the results show in the table, we can find that all the structured data get higher f1-score than the textual data, the reasons might be as follows:

1. In the textual data, people might use some synonyms to replace the words and have some more informations, but for the structured data, people only need to give the information that the attributes need.
2. Analysis of predicted results

Precision = True Positive / (True Positive + False Positive)

Recall = True Positive / (True Positive + False Negative)

F1 Score = 2 \* (Precision \* Recall) / (Precision + Recall)

For all the datasets that we run, there are two main reasons that caused the low F1-score:

(1). Precision and recall are both quite low in one dataset.

(2). The Flase Positive is very high which causes precision is much smaller than recall, i.e. plenty of the prediction is the same but the truth is different.

(3) The possible reasons for (2)’s situation might be both sentences use many of the same root words or the same substrings, but the use of negation is not the same in both sentences.

(4)?When the threshold is smaller, the f1-score is higher.