

# Schema Matching using Machine Learning

Ankita Mehta

Shruti Jadon

Tanvi Sahay

**Abstract**—In this project, we deal with the problem of matching schema of different tables across databases with each other with the help of certain machine learning algorithms. This is done in order to recognize which attributes contain the same or similar values and might map to each other in case the two databases are to be used in unison. We also tackle the challenge of a single attribute mapping to multiple attributes along with the case of basic one-to-one mapping. For this report, one to one mapping using Kohonen Self-Organizing Maps and Multilayer Perceptrons has been explained and experiments carried out have been presented.

## I. INTRODUCTION

Schema matching is one of the key stepping stones for performing data integration and automatic this task has been a topic of research for several years. In simple terms, schema matching can be explained as follows: Given two databases  $X(x_1, x_2, x_3)$  and  $Y(y_1, y_2, y_3)$  with  $x_n$  and  $y_n$  representing their attributes respectively, we match a schema attribute to another either if it is semantically similar or if it represents the same data. Consider Figure 1 which contains two schemas: *Student* and *Grad-Student*.

## II. RELATED WORK

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

## III. FEATURE EXTRACTION

Shruti gave it already, I think.

## IV. ONE TO ONE SCHEMA MATCHING

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut

metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

### A. Kohonen Self Organizing Maps

Theory of SOM and how it was implemented. Image of the network might be good.

### B. Multilayer Perceptron

This section only if we decide to do this finally.

## V. EXPERIMENTATION

### A. Dataset

### B. Data Labelling

We might need to hand label the attributes as a base line to check the accuracy of our system.

### C. Results

Clustering accuracy table, scatter plots etc.

## VI. LIMITATIONS

No many to one mapping and anything else you can think of.

## VII. FUTURE WORK

## REFERENCES

- [1] H. Kopka and P. W. Daly, *A Guide to L<sup>A</sup>T<sub>E</sub>X*, 3rd ed. Harlow, England: Addison-Wesley, 1999.