




CoSi: Visual Comparison of Similarities in High-Dimensional Data Ensembles

Supplemental Material

A. Heim^{1,2} , E. Gröller^{2,3}  and C. Heinzl¹ 

¹ University of Applied Sciences Upper Austria, Austria

²TU Wien, Austria

³VRVis, Austria

1. Introduction

In the following paragraphs we present additional information that accompanies our paper which was left out due to space constraints. We provide a more detailed explanation about our definitions of terms and describe an additional scenario presenting the generalizability of our framework.

2. Definition of Material Domain Terminology

In order to provide a better understanding of the materials science domain, the data as well as our design choices, we illustrate the terms used in this paper in Figure 1.

3. Generalizability

To demonstrate the generalizability of our framework, we describe a scenario in which we compare a generic data ensemble that does not come from the domain of materials science. The data was published by ILOSTAT [Org20] on the Gapminder platform [Gap20]. In this scenario we visualize the unemployment of men and women of different age groups. The ensemble consists of six different datasets. Three datasets describe the unemployment of women in the age groups between 15-24, 25-54 and 55-64 years of age. The remaining three groups describe the unemployment of men in the same age groups. The datasets are organized as follows: For about 165 countries of the world, the unemployment rate, expressed as a percentage, is recorded for 29 years. Our tool enables different analysis goals depending on the structure of the data. If we were to define the years as features and the countries as characteristics, we could determine in which countries unemployment has developed similarly over the years. In this scenario, we specify that the different countries represent the features, while the years describe the characteristics. With CoSi we can identify if the unemployment rate of men and women behaves similarly. Furthermore we can assess how the unemployment developed in the various countries through the years.

Scenario 3 - Compare Unemployment of Men and Women

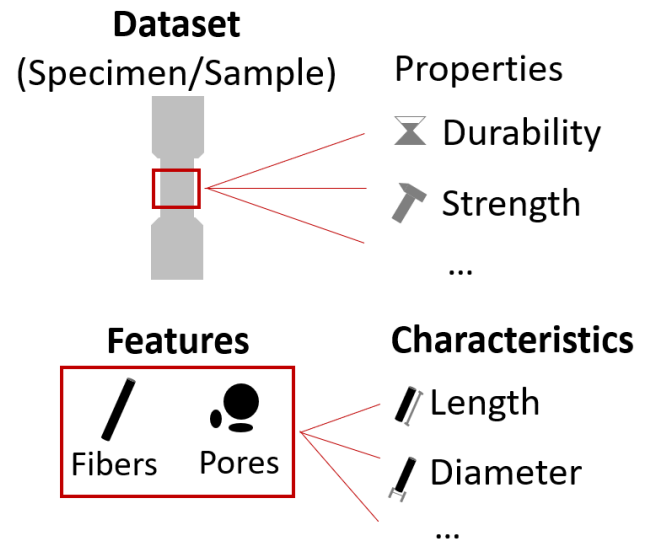


Figure 1: A specimen, also referred to as sample, is a material system that possesses certain properties such as a particular durability or strength. Specimens contain features such as fibers and pores. The number of features and their characteristics, like their length, diameter, etc., affect the specimen's properties.

in different Age Groups. When comparing the datasets in the ensemble similarity explorer, we can see that the unemployment rates in the different countries behave similarly within a gender group, while a clear difference between the sexes can be observed (see Figure 2). In many countries, the unemployment rates for men and women are nearly equal for each age group, as indicated by the darkest red bins in the center of the visualization across all datasets. If we look at the excerpt of the bar chart, showing the similarity of the unemployment rate in the various years, we can infer, that from the 1990s to the 2010s the unemployment rate became increasingly different, indicated by the decreasing size of the bars (see Figure

3 A). Observing the box plot next, we perceive that the unemployment rate increases in various countries (see Figure 3). This finding is illustrated by the growing boxes, which become bigger from the 1990s to the 2010s.

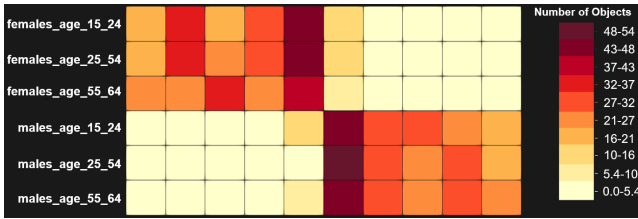


Figure 2: Scenario 3 - Comparison of the unemployment rate between men and women of three different age groups. The ensemble similarity explorer shows a clear difference between the unemployment of men and women in the different countries.

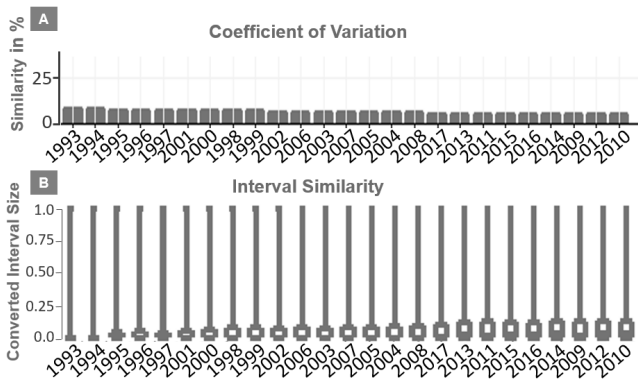


Figure 3: An excerpt of the similarity widget describing how the unemployment rate develops in the various countries. (A) In the bar chart, the more the bars decrease in size, the more different the unemployment rate develops in the various countries from the 1990s to the 2010s. (B) The boxes that increase in size in the box plot, indicate the rise in unemployment rates from the 1990s to the 2010s in the various countries.

References

- [Gap20] Unemployment data. <https://www.gapminder.org/data/>, 08 2020. 1
- [Org20] ORGANIZATION I. L.: Employment by sex and age — ilo modelled estimates. <https://ilostat.ilo.org/data>, 08 2020. 1