IUST Faculty evaluation using Data Envelopment Analysis with Web scraping(Spring 2017)

1

Data envelopment analysis(DEA) is a well established method when it comes to evaluating the efficiency. DEA's advantage comes from its ability to compare each unit with one similar to it and considers the limitations and advantages of each unit. Besides, the model can suggest a way to improve the efficiency to each unit. Surprisingly, DEA is not used much in academia. One reason might be the lack of data to do so. There is no user friendly open source to be used to evaluate people in academia. In order to tackle that problem, we used Scopus API to gather all the data needed and saved them in a SQL data set. Now that we have the data needed, we can use it to evaluate professors. In order to have a fair comparison between professors with different area of study, we fuzzified the data. Then, FDEA is used to evaluate professors.

1. Steps

1.1. Scraping professors name from the website

Although we can hand craft the professors' name to the program, it would be very time consuming and we may miss some names. So we scraped http://ie.iust.ac.ir/page/5518/Faculty-membersuniveristy's website to get the name of the professors.

1.2. Getting the Scopus ID of professors and basic information

Now that we have the name of professors we can query from Scopus API to get the scopus ID, H-index, number of citations of each of the professors. Now that we have the Scopus ID, we can query for their paper and more detailed information.

1.3. Mining the papers information and making a SQL dataset

By using Scopus ID we can have access to ID and information of all the papers of all the professors. Due to rapid advances in science, we just use recent 3 years papers. After gathering information for all these papers, we structurize them and save them in an SQLLite dataset to be easier to use for future uses.

1.4. FDEA

Now that we have the structured dataset, we can query to get papers in each area and fuzzify them into high-cited, medium-cited and low-cited. It's important to note that the range of high medium and low would vary for each area of study. Then the data is given to the DEA model to evaluate the professors and each department in industrial engineering.