Use data mining methods in quality measurement in the education systems

Sándor Balázs Domonkos, Németh Tamás

Abstract: Our basic problem is rooted in the education systems, where they want to measure and evaluate the pedagogical work from year to year and watch for pedagogical developments. These measures can be used for the individuals to get informations on which field they need to improve and can be used for rewarding systems. These measurements can be achieved by different methods, in this case our method is the surveying in the classes,direct approach person to person surveys and demand and satisfaction measurements for every person. These surveys are precomposed with discussions about the needed attributes. We get real datas from 58 different schools from 2007-2016 , nearly 8000 educators. All these surveys-ratings and other datas get collected and processed to get in a usable form. The schools make a statistics with the collected datas every year, but the statistics has really person dependents and because of that these has a lots of distorsions. For example a human focused teacher can give bad points for a informatics teacher class, because they dont share a territorial interest.To get a workaround for these statistics "personal" dependencies, we use the pagerank algorithm with the Comparability graph[1]. With the comparability graph we could compare two attributes with each other not heavily depended on the people who fill the survey and we could make a new graph for every attributes. After that on those graphs we can use the pagerank algorithms to get the datas out that we want to examine and get further consequences. For these datas we should tell they are lead to a much more usable development curves about the educators qualities. Important thing is these datas dont suffer the distorsions of the statistic ones has.

Keywords: data mining, pagerank, comparability graph

Introduction

The main problem is raising from the education system qualilty and valuation system about the educators. This system is really important for the schools because only this way they can monitor the educators and other workers qualities and working capabilities. These datas that they collected via surveys are not just used to person evaluation and ratings, but for to get a view about the workload for each individual, so it can help for example to make a better timetable for everyone with this datas too. For bigger view we can get more datas about the leadership and the lecturer and parents connections too and the way they communicate with each other. But this article is focusing on the quality measurement on the lecturers, with our method they get more accurate data. So they can get well earned rewards and get a better feedback on which area they are on a good level and where they need to improve. So we examined the datas and the statistical style old system and we get the conclusion that the statistical based system is can be very noisy and heavily depends on the person to person connections and sometimes the given teacher connection to another subject. For example literature teacher can give only bad points for informatics and math teachers and good points only for literature teachers, because they dont share a territorial interest or they are not in a good connection in the workplace or any other issues they can have. So statistics methods cannot overcome this types of noise and predilection. Because of this the statistics method have a very high dispersion on the results and for year to year compliment the fluctuation is very heavy on this results.

Graph Based Solution

Our solution for this problem is based on graphs. We analize the raw datas try out different methods to break up the datas for different viewpoints and build different types of graphs. After a lots of experience we decide to use Comparability graphs for every attributes. First of all we need to break up the datas for 3 parts: evaluating lecturer, attributes for the evaluated lecturer, evaluated lecturer. So in this way we can construct comparability graphs for every attributes. In this way 2 evalutaion can be compared if they are from the same evaluator lecturer and for the same attributes, in this way the data noise is dampening down,because its compare data from the same evaluator. From these datas we construct

Comparability graphs for every attributes. On these new graphs we can use the PageRank algorithm to get quantified results for every evaluated lecturer for every attributes without the distorsion of the personal conflicts or any other type of distorsions mentioned above. The pagerank algorith give a more data about the evaluating lecutrers too for each attributes and after the pagerank is done these values are just more well adjusted for every attributes for ever evaluated lecturer. For example on a statistics old styled method if someone is get a few bad results from personal conflicts, these bad ratings can really pull down the person numbers , but with our method the comparability graph for attritbutes processed with pagerank methods these bad ratings can ease down the datas, cause the algoritm will ease out if someone give only bad points for an individual and good points for everyone else or give bad points for everyone and this can be takint into account with a less impact, and these are all for the good cases too, so not just only the bad ratings.

Results

The quantified values are less likely to depend on the evaluator person and much more like depends on evaluated person. This method the year to year improvements are take shape in a better way and more readable way. This way we not just only get the results for the evaluated people attributes it can work backway too, we can get datas from the evaluators too, so we can make some conclusion in the lecturers have a personal or work related problems with each other. For results we get a much lower dispersions level and with a lower dispersion and very low fluctuation level we can much more easier fit line to yearly improvement attributes too. The average dispersion level for the results can be 30-50 percent better than the statistics ones.

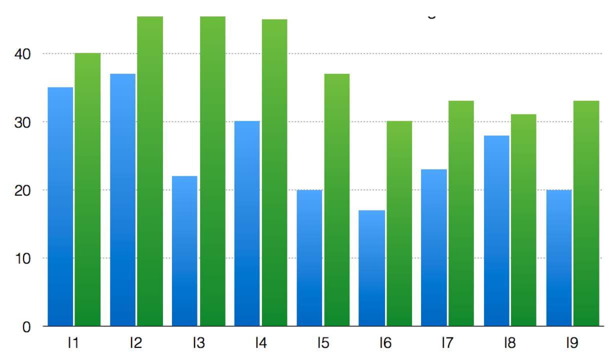

Figure 1: Blue pagerank, green statistical

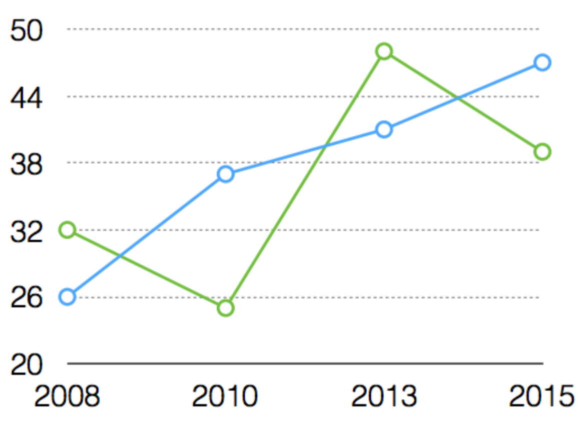

Figure 2: Blue pagerank, green statistical

Acknowledgements

This work was supported by Enaplo and snw systems.

References

[1] A. London. A local PageRank algorithm for evaluating the importance of scientific articles , *Annales Mathematicae et Informaticae* 44:131-140 2015.