Big Data - Exercise 1

Svein Jakob Høie
Jonas Brunvoll Larsson

a)

"Describe the three V's of the data sources used in the IT service requests management system at Florida State University."

The three V´s consisting of volume, velocity and variety, are the general way of defining big data. Big data is used to describe the exponential growth and availability of data - including both structured, semi-structured and unstructured data. Information Technology Services (ITS) is the central IT Organization for Florida State University (FSU). Including Florida State University, Information Technology Services provides IT support and IT services to students, staff and different institutions from 16 different colleges. Let's talk about how the ITS handles big data.

Firstly let us discuss how the ITS handles the velocity of data. ITS has one system who captures and manages all the different IT service requests from the 16 different colleges. Since the system was first used until this article was written five years later, more than 100 000 service requests have been logged in the system. The article does not say anything about how ITS handled all the IT requests before 2011, but since the system was still used in 2016 we assume that the system ensures better operational efficiency than the older solution.

Secondly let us discuss how the ITS handles the variety of data. The ITS offers more than 100 different IT services, including email, desktop support, file storage and system support just for mention a few. In other words the system manages to collect data from many different sources. The data itself collected from these sources vary from structured (example software licensing), to unstructured data (example email). The ability to analyze unstructured data is important because it results in more accurate analysis and therefore hopefully better decision making.

Lastly let us discuss volume. When discussing volume we discuss the share amount of data ITS system receives and manages over a period of time. As a tool to better understand the volume of the tasks the ITS IT support groups solve, ITS prepears a weekly report. The weekly report gives a good overview of the major incoming request and how they were solved / should be solved.

b)

"Can you relate these to the cases presented in the paper and give one example for each of the three benefits?"

i) FSU uses text mining to gain insight in the what and why of service requests, in order to group them into correct topics.

ii) The insight gained from text mining the service requests is used to identify the most prevalent topics and their evolution over time. By knowing the most prevalent topics, the right amount of resources and time can be invested in order to attempt to decrease the number of service requests and fix the problems that would cause the large number of requests.

iii) Iventx decided to improve their IT service management to the point that their text analytics now not only understand the users issue, but also solves the issue directly. This method naturally does not work 100% of the time, but currently solves the service requests 60% of the time.

c)

"Discuss whether each of the following activities is a database query or a data mining/machine learning task."

- Dividing all the service tickets into two groups: open and closed requests.
    - This activity could be done with a simple database query for a field set to either true or false.

- Computing the total number of requests issued in the past week.
    - This activity could be solved with a database query to sum the number of requests issued in the last 7 days.

- Group more than 100,000 service tickets into topics (printer, email forwarding, etc…)
    - This task seems more like a text mining task, where we would have a program check the contents of the service requests and from the words and sentences in

that request, decide which topic is most likely to be the most descriptive.

- A ‘recommender’ system that suggests the most appropriate expert for solving a problem and the possible solutions based on how requests were handled in the past.
    - A recommender system like this is probably a task for a machine learning program. It would have to be a system that learns which expert handles which problem best, and would need a lot of data from previous service requests in order to be an optimal system.