Web Information Retrieval

Assignment 1

Team Name : Gamma

Members

Name	Matriculation Number
Mohammad Nizam Uddin	**216101140**
Md. Shohel Ahamad	**216203438**
MD Jakaria Nawaz	**216203442**
Shreya Chatterjee	**216100848**

1 Machine Learning and its relation to Information Retrieval (14 Points)

Answer:

1.1 - Computer program to *learn* in the context of Machine Learning.

Computer programs are group of instructions that a computer performs accordingly. Over the past three decades, research has been done on machine learning. Computer programs can adapt, to learn from data. They can have a huge practical impact in the field of machine learning if they start to improve their ability to learn.

A machine can't learn something by himself unless someone tells it what they want it to learn. In summary, we can say that predicting about data based on the statistical model of existing data is how a computer program learns. And over time the prediction of a computer program gets more precise as it gets to evaluate thousands or even millions of times on its data set.

1.2 - Difference between *supervised* and *unsupervised* learning.

Supervised learning:

In machine learning, supervised learning is the task of learning from labeled training data and apply that knowledge in new test data to produce the output. Labeled training data gets analyzed by supervised learning algorithms and thus it generates function, which can be used to identify test data accordingly. Classification is required for the learning algorithm to generalize from the training data in a reasonable way. *(Dataconomy. [ONLINE])*

Unsupervised learning:

Unsupervised learning doesn't have neither trained labeled data nor any response variable. It models the data in a structure to learn more about the data. Unsupervised learning doesn't have any correct answer as there is no train data. Algorithms are on their own to find and determine a structure in the data. Clustering analysis is the most common unsupervised learning method to find hidden patterns or groups in data. *(Dataconomy. [ONLINE])*

1.3 - Difference between *flat* and *hierarchical* clustering.

Clustering is the process of identifying objects of same characteristics and combine them based on their similarities. Documents from different cluster should not be similar in any way.
(*Big Data Made Simple - One source. Many perspectives.. 2017.[ONLINE]*)

Flat clustering

A flat set of clusters that are not structured to relate to each other are created by flat clustering. The process starts with a random separation of documents into groups. It refines iteratively. K-means is the main algorithm of flat clustering. Flat clustering is simple and efficient, but it returns unstructured clusters. It has a number of drawbacks. *(Flat clustering. [ONLINE])*

Hierarchical clustering

Hierarchical clustering creates a hierarchically structured set of clusters that are more informative than the unstructured set of clusters. It solves the problem of labeling clusters. It doesn't require any prespecified information regarding number of clusters. Deterministic hierarchical clustering are widely used in IR. But it is not as efficient as flat clustering. *(Hierarchical clustering. [ONLINE])*

1.4 - Machine learning techniques in Information Retrieval.

In information retrieval, the main goal is to substance information resources from the web, relevant to an information need. When someone is searching for anything on the web it can show related and convenient search results for that user by information retrieval. Either way, the end goal is to get out the relevant resources, For an example, while someone is searching for a shoe online it can bring several style and brand's shoes, however, by information retrieval it can show only products that person can may based on that person's previous search history.

References & Bibliography:

- Bhatia, S.J., Deogun, J.S. and Raghavan, V.V. (1995) “Conceptual query formulation and retrieval.” Journal of Intelligent Information Systems 5(3), pp. 183–209
- Big Data Made Simple - One source. Many perspectives.. 2017. What is Clustering in Data Mining?. [ONLINE] Available at: http://bigdata-madesimple.com/what-is-clustering-in-data-mining/. [Accessed 06 May 2017].
- Dataconomy. 2017. What's The Difference Between Supervised and Unsupervised Learning? - Dataconomy. [ONLINE] Available at: http://dataconomy.com/2015/01/whats-the-difference-between-supervised-and-unsupervised-learning/. [Accessed 06 May 2017].
- Flat clustering. 2017. Flat clustering. [ONLINE] Available at: https://nlp.stanford.edu/IR-book/html/htmledition/flat-clustering-1.html. [Accessed 06 May 2017].
- Hierarchical clustering. 2017. Hierarchical clustering. [ONLINE] Available at: https://nlp.stanford.edu/IR-book/html/htmledition/hierarchical-clustering-1.html. [Accessed 06 May 2017].
- Krulwich, B. (1995a) “Learning document category descriptions through the extraction of semantically significant phrases.” Proceedings of the IJCAI Workshop on Data Engineering for Inductive Learning, (Montreal, Canada).
- Machine Learning Mastery. 2017. Supervised and Unsupervised Machine Learning Algorithms - Machine Learning Mastery. [ONLINE] Available at: http://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/. [Accessed 06 May 2017].
- Martin, J.D. (1995) “Clustering full text documents.” Proceedings of the IJCAI Workshop on Data Engineering for Inductive Learning at IJCAI-95, (Montreal, Canada).
- Petr Sojka, Hinrich Schütze et al.. 2017. Introduction to Information Retrieval. [ONLINE] Available at: https://www.fi.muni.cz/~sojka/PV211/p16flat.pdf. [Accessed 7 May 2017].