Prof Halim Sayoud / USTHB University
http://sayoud.net

This research work is presented at the
IVAPP conference,
Berlin 11-14 March 2015

A Visual Analytics based Investigation on the Authorship of the Holy Quran

http://sayoud.net
Halim Sayoud
USTHB University, Algiers.
halim.sayoud@uni.de halim@sayoud.net

Keywords: Visual analytics, Authorship attribution, Pattern recognition, Natural language processing, Religious books.

Abstract: In this paper, we present a visual analytics based investigation for the task of authorship attribution of the holy Quran with regards to the Hadith Author (*the Prophet*). This can be seen as an authorship discrimination task between the two religious books: Quran vs Hadith. The first book represents the Divine book written by Allah (*God*) as claimed by the Prophet Muhammad, whereas the second one represents a collection of certified Prophet's statements.
Two visual analytics clustering methods are employed, namely: a Hierarchical Clustering and Fuzzy C-mean Clustering. On the other hand, seven types of NLP features are combined and normalized by PCA reduction before the classification process. The visual analytics results have revealed interesting results in 2D and 3D disposition. In summary, they show two main clusters in both experiments: Quran cluster and Hadith cluster; and the disposition of the resulting clusters corresponds to a clear authorship distinction between the two religious books.

About my Research Interest
« Computational Linguistics »

Automatic Classification of Text Documents

- **Analysis of some historical documents (600 A.D.)**

- **Authorship verification of antique books:**
  - The 10 Arabic Travelers;
  - Holy Quran (Islamic Book of God);
  - Statements of the prophet Mohamed;
  - And others…

Problem about the authorship of the holy Quran:

Was the Quran only an invention of the Prophet?

or only transmitted to him by God?

No scientific answer for 14 centuries *(in my knowledge)*

?

2

Does an Author have a sort of Fingerprint?

Not exactly!

But there exist typical stylistic features specific to every author

Recent scientific researches have shown that these stylistic features can distinguish between 2 authors

3

About the Quran Authorship

- In the Islamic religion, the Prophet says that the Quran was sent down to him by God.
  He is not the author but only the narrator…
- But some doubts claim that it could be invented (*written*) by the Prophet…Examples:
  - أم يقولون تقوله بل لا يؤمنون
  - أم يقولون افتراه قل فأتوا بسورة مثله وادعوا من استطعتم من دون الله إن كنتم صادقين
  - Reference: Is the Holy Quran Muhammad's invention ?
  - By DR :Abd Al-Raheem Al-Shreef

4

Classification Problem → Author Discrimination

<< Do the 2 Books have the same Author? >>

Author1(*Quran*) =/≠ Author2(*Hadith*)

?

An interesting way to solve this enigma is:

Authorship Verification or Stylometry

5

DEFINITION / RECALL

Author 1 Mohammad → Hadith

Author 2 Allah *(God)* → Quran

- 11 different segments
- 14 different segments

H1, H2, … H11

Q1, Q2, …Q14

Classification

6

Next Page

Old copy of the Holy Quran

Old copy of the Hadith

7

Some difficulties in the Arabic text

- 1. Tachkil (basic voyels): We have removed them…
- 2. Some prepositions/ articles are welded (connected without space between the 2 words)
- 3. Most of existing text-mining softwares are made for the latine alphabet and are not suitable with Arabic letters
- 4. Text direction: from right to left: → Not a great difficulty.

8

Features

In this investigation, a mixture of different features is proposed:

-Author Related Pronouns (ARP),
-Father Based Surname (FBS),
-Discriminative Words (DisW),
-COST value,
-Word Length Frequency (WLF),
-Coordination Conjunction (CC)
-and Starting Coordination conjunction (SCC).

All those features are original and some of them are used for the first time in stylometry.

9

CLUSTERING

In pattern recognition, cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (*ie. cluster*) are more similar to each other than to those in other groups (Wi2, 2014) (Norusis, 2008).

10

VISUAL ANALYTICS

Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces (Thomas 2005). It is often used in cluster analysis to make the analyst's judgment easier to develop and more objective.

11

DATASET DESCRIPTION

The two books have been segmented into 25 several text segments (14 for the Quran and 11 for the Hadith). Do, there are 2sets of texts:
{Q1, Q2, …Q14} for the Quran and **{H1, H2, …H11} for the Hadith**.

12

Hierarchical clustering

The hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters (Greenacre, 2014). In our case, we used the Agglomerative clustering with a Manhattan distance measure: $d(X,Y) = \sum_i |x_i - y_i|$

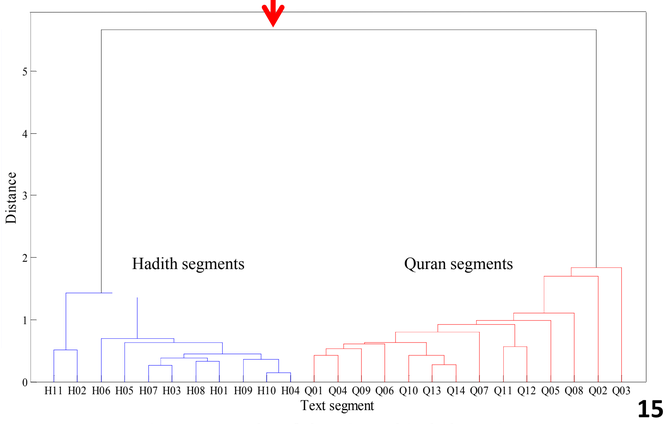
13

Fuzzy C-Means clustering

Fuzzy clustering is a cluster analysis in which the allocation of data points to clusters is not hard but "fuzzy "(Suganya, 2012). So, every point has a degree of belonging to clusters, rather than belonging completely to just one cluster. That is, any point x has a set of coefficients giving the degree of being in the kth cluster $w_k(x)$. With fuzzy c-means, the centroid of a cluster is the mean of all points, weighted by their degree of belonging to the cluster: $C_k = \frac{\sum_x w_k(x)^m x}{\sum_x w_k(x)^m}$

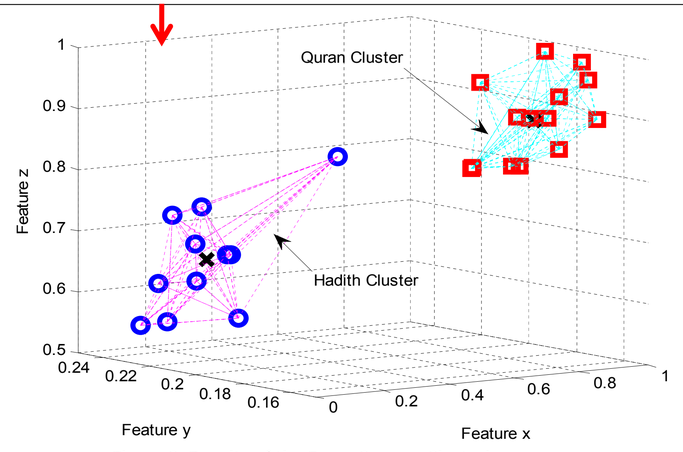
14

Automatic classification by the Hierarchical Clustering



15

Figure 1: Results of the Hierarchical Clustering.



Automatic classification by the Fuzzy C-Mean Clustering

16

Figure 2: Results of the Fuzzy C-mean Clustering.

Conclusion/Discussion

Observation: -We notice that there are 2 distinct clusters: Quran cluster (in red) and Hadith cluster (in blue);
-Furthermore, all the Quran texts are grouped together in one cluster and all the Hadith texts are grouped together in another distinct cluster.

Consequently: This implies that **the two books** (Quran and Hadith) are **written by 2 different authors** or at least with 2 different styles.

17