



# Towards Email Importance based on User Profile and Message

Fatemeh Razzaghi and Ali A. Ghorbani

## Abstract

Towards Email Importance based on User Profile and Message is a model for mining emails and identifying their importance. The underlying theory is that there exists an intuitive link between the relevance of information (email content), the role of the sender, the circumstances under which the information is received and the semantic features of the email. In this system we are combining the user profile in an email social network, including role, interests and habits, and responsibility, with email features, including content relevant keywords, message length, and other chosen features, using a common relevance weight. Emails classified using Fuzzy algorithm.

## Introduction

An email network is essentially a directed or undirected graph with nodes and links representing users and their messages respectively. Email can be costly and extremely time consuming, especially for those who track multiple email accounts and have to deal with a full Inbox every day. The negative effects of email overload are obvious: loss of time, loss of productivity, and unanswered email or delays in email replies.

Email in its current situation:

- Great tool for effective collaboration
- Go across time and distance
- Email overload
- Prioritizing messages
- Unanswered emails

By combining a set of selected features from email header and content with some features extracted from email social network analysis, we are developing an effective email management tool that would help us to overcome the above challenges and assist email clients in detecting important emails.

## Techniques

- Information Retrieval
- Information Filtering
- Text classification
- Unstructured Text Mining
- User Profiling
- Social Network Analysis
- Email Retrieval

## Dataset

The only publicly available dataset which has been used in email research is the Enron corpus. We used this data set and there are 255636 email and 87474 users that only 103 of their organizational roles are available.

## Analyzing User Profile

We are consider Enron email dataset as a graph that depicts the relationships between their employees based on who emailed whom and how often. A primary use of graph theory in social network analysis is to identify "important" actors. Centrality and prestige concepts seek to quantify graph theoretic ideas about an individual actors prominence within a network by summarizing structural relations among the nodes. One way to start is to first seek out nodes that either send or receive messages from numerous employees, thus possibly representing either hubs or authorities in the network. Three different measures are used to insert new weights improving the identification of the importance concept.

- Degree centrality
- Closeness
- Between ness

$$C_D(i) = \frac{d(i)}{n-1}$$

$$C_C(i) = \frac{n-1}{\sum_{j=1}^n d(i,j)}$$

$$C_B(i) = \sum_{j < k} \frac{P_{jk}(i)}{P_{jk}}$$

## User profile Folder

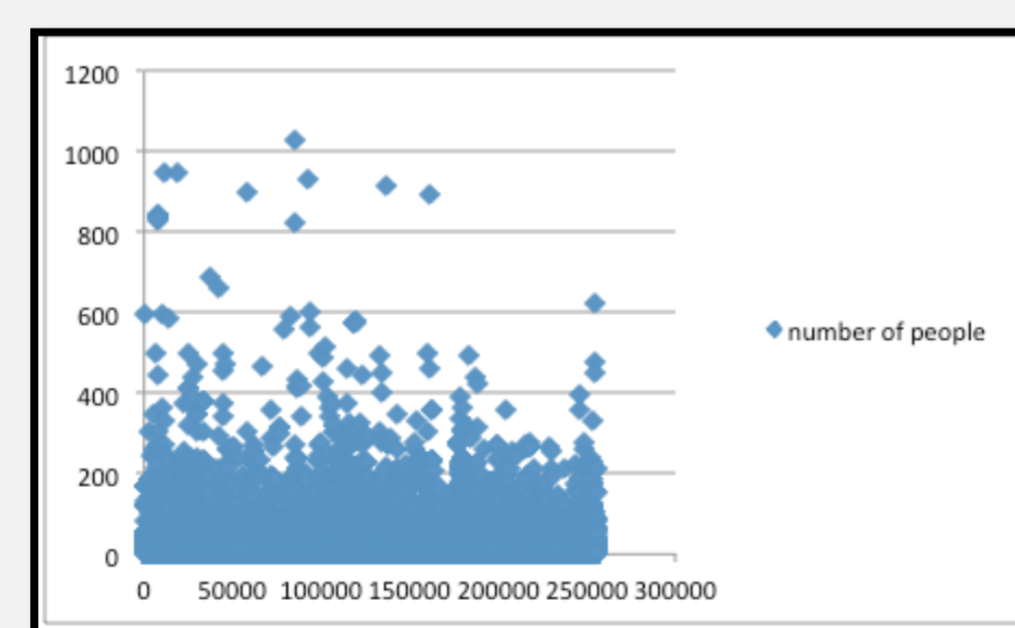
The user profile folder consist of two different types of features:

- Interest lists
- User roles

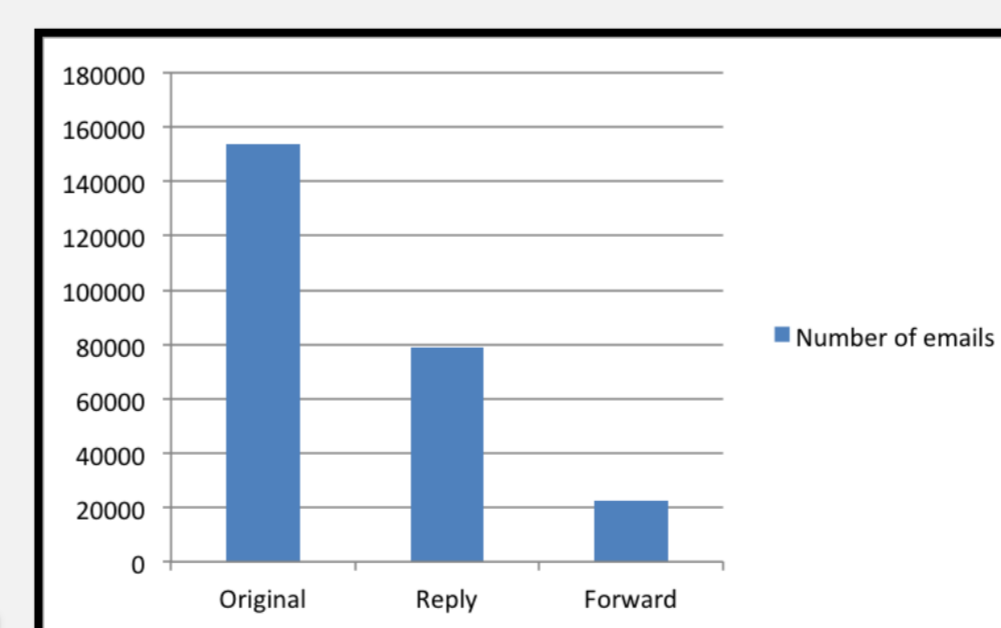
## Analyzing Email Interactions

We have also done analysis to select the most informative interaction based features considering each email independently .

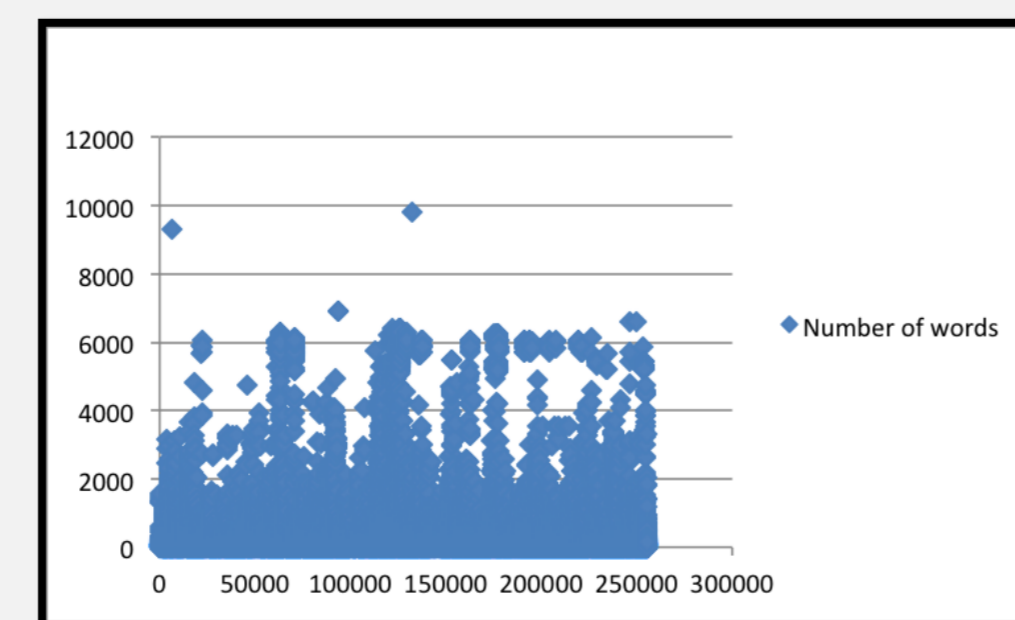
People Count



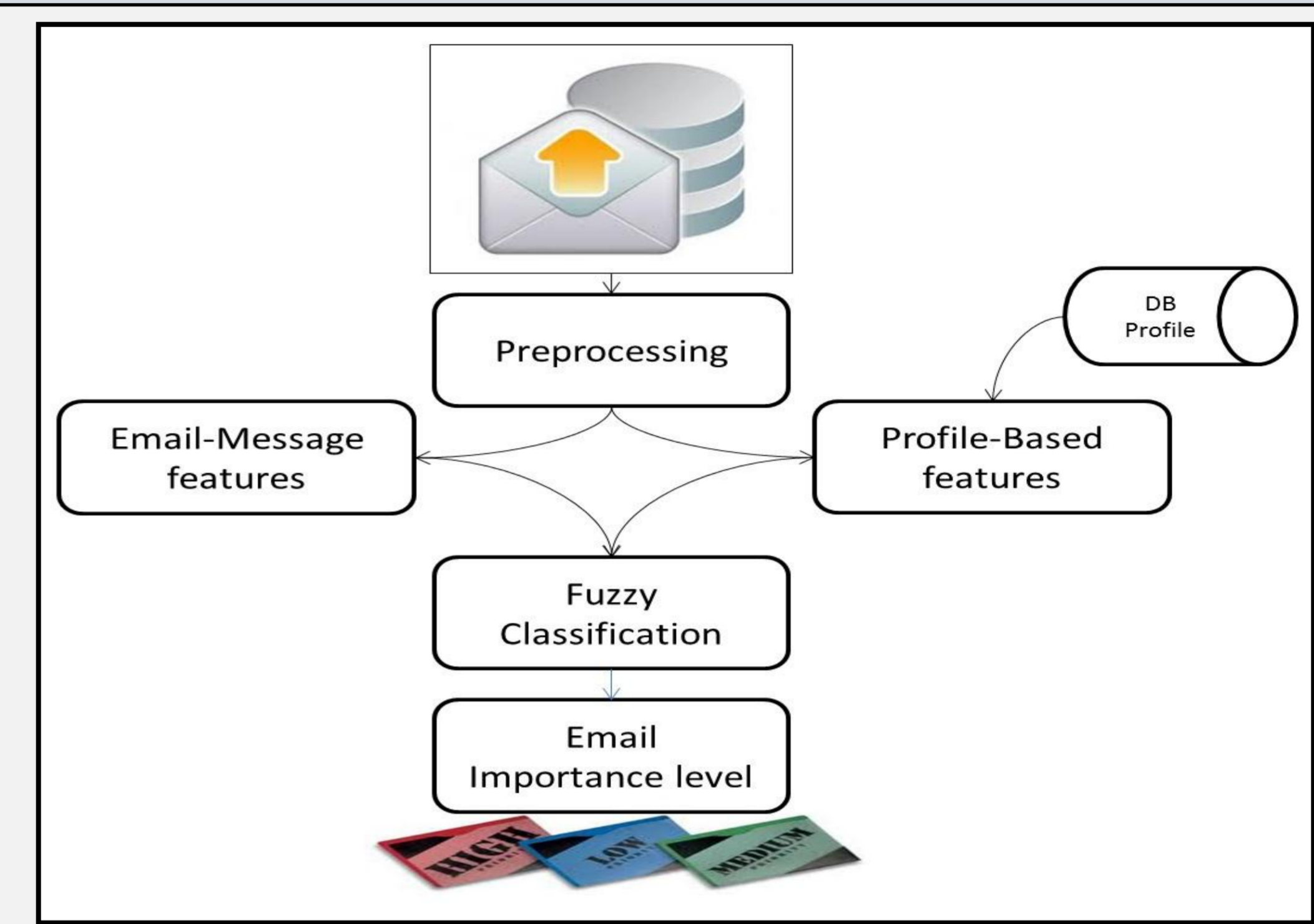
Answer Type



Message Length

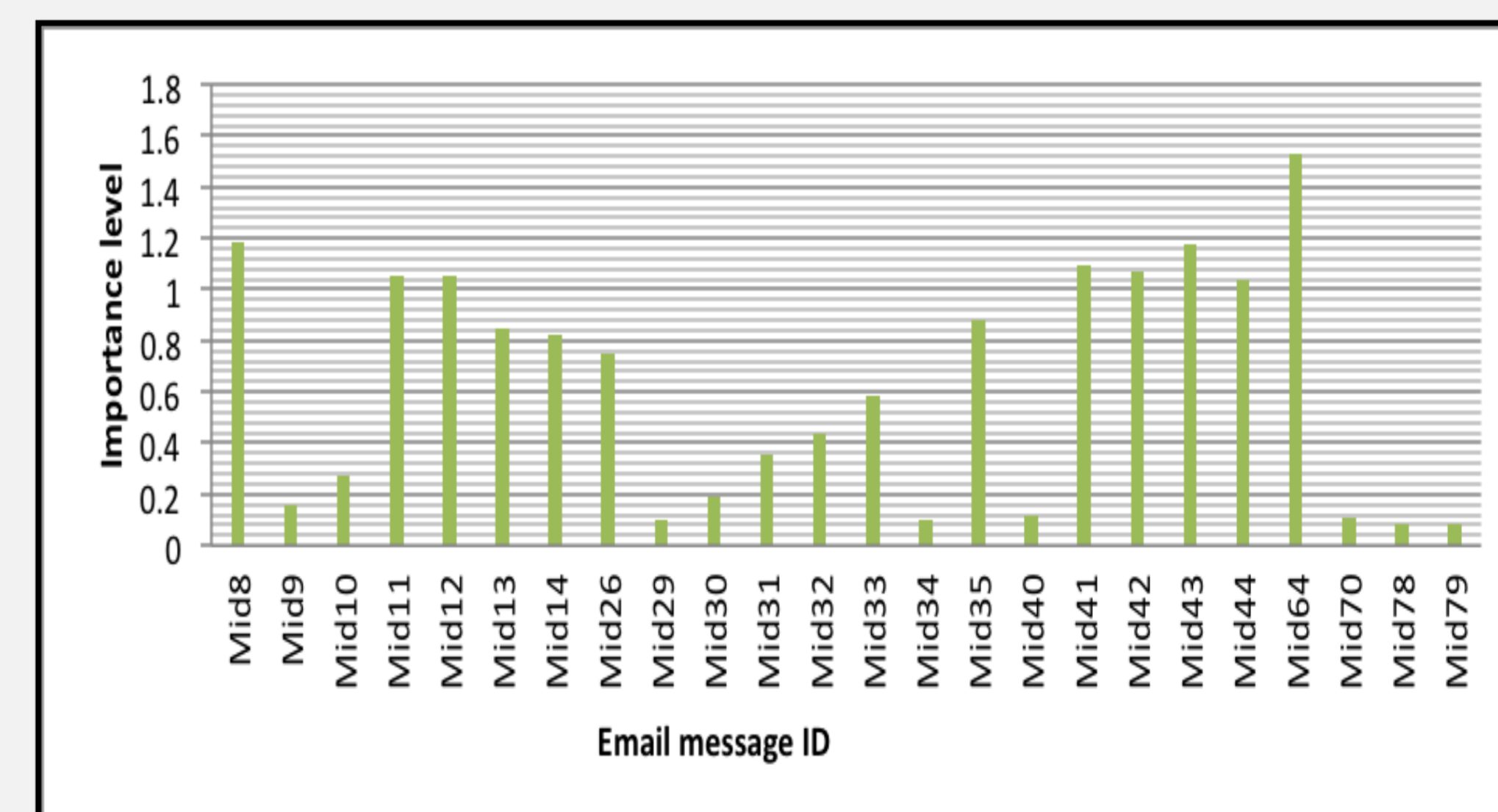


## Design



## Experiments and Results

We classify rank emails based on their importance using normalized weighted feature vectors. We are using the Fuzzy theory that presents the sets whose elements have degrees of membership. For each email the vector of features, based on the information contained in email and user profiles, computed. The assessment was done on 3-point scale levels 0, 1, and 2 (and more) corresponding to Low, Medium, and High level of importance respectively. Therefor social network measures along with some Interaction-based features can make a contribution to classifying emails.



Up to our knowledge, there is no Email Retrieval Importance Ranking system available to compare with its performance. For the future scope of this study we want to apply and improve clustering performance. Also we plan to investigate the ability of a system that can learn a significant portion of the users information preferences such as contexts, interests, disinterests, and contact trust preferences over time.