**Intelligent E-mail Prediction System**



**Synopsis Report**

Submitted by:

Ankit Gupta

(2K10/IT/016)

Chirag Gupta

(2K10/IT/023)

Sanchit Gupta

(2K10/IT/056)

Vaseem Ahmed Khan

(2K10/IT/068)

**Mentor**:-

**Prof. N S Raghava**

Information Technology

Delhi Technological University

**ABSTRACT**

Nowadays, email has become one of the most critical personal and business applications and email users would experience serious consequences if email messages could not be available or experience high volume of messages which lead to congestions, overloads and limited storage space coupled with unstructured messages in mail boxes. A few years ago, the means of communication are via letters by post, telegraph, fax, couriers to mention a few but now the focus has changed to a faster means of obtaining quick responses and faster ways of communication-emails.

We propose a new framework to help organise and prioritize email better; An intelligent email prediction system (IEPS). The goal is to organise emails better in mail boxes, prioritise emails based on the focus of the email content. The intelligent email prediction system helps to improve email users’ performances, saves time, very effective and efficient tool and is cost effective for businesses and for personal use. The system is evaluated against a corpus of human-judged predictions, reaching satisfactory level of performance.

**INTRODUCTION**

The aim of IEPS is to help email users save lots of time while checking, reading and searching for email messages reduce email overloads and congestions which are caused as a results of high volume of email messages in mail box, organise mail boxes better and all these makes life easier and improves user’s performances and productivity.

One of the lines of work developed within IEPS is the use of machine learning techniques for information management namely, text classification of email messages and determination of messages that needs reply and why. IEPS is an automated system that learns to determine whether email messages received in a mail box needs a reply or no action is to be taken.

Whittaker and Sidner analyzed the use of email to perform task management, personal archiving, and asynchronous communication and referred to the three as “email overload”. They concluded: (1) Users perform a large variety of work-related tasks with email. (2) As a result, users are overwhelmed with the amount of information in their mailbox. IEPS will enable email users to both manage their email inboxes and at the same time manage their time more efficiently.

**FEATURES EXTRACTED**

A. **Sender’s email address**: If an email message is from certain people: CEO, Manager, Head of department, debt collector, hospital etc then assume it may require a reply and assign a score of 1 to this on the scoring board provide with the predictor system

B. **Subject of each email messages**: If the subject of the email messages are similar or almost related to the phrases stored in the database of words and phrases on our system- please reply soon, let me hear from you, Is there any news today?, Are we on the same project or not. A score of 3 is allocated to the scoring board.

C. **Cc/Bcc Filed**: If a mail is Cc or Bcc to others, such a mail may require attention because mail copied to others may be as a result of a group project or task that an individual or groups need to be aware of or act upon as soon as possible. Such a mail may require reply. A score of 2 is allocated to the scoring board.

D. **Email content**: If a mail contain words and phrases such as: interrogative words- could, when, where, how? Rarely used phrases: meeting at noon, is it alright, because of the yearly budget, based on what we saw, etc. such mail may denote a request and may denote a reply. A score of 3 is allocated to this area because content of email messages focus on what the email is all about and that is why human participant spent more time in analysing this area of research.

E. **Previous conversation**: An existing email conversation within the same subject indicated that majority of such messages require attention. A score of 1 is assigned to the scoring board making a total of 10 points.



**CLASSIFIED OUTPUT**

**Need reply**: Email messages that are categorised to need a reply indicated that such a mail passed the threshold set by human analyser. The threshold value assigned for email messages that require a reply is 7 and any messages that score 7 or above out of total score of 10 will be assigned the tag “need reply -1”.

**Do not need reply**: Messages that scoreless that the threshold value 7 will be assigned a tag “Do not need reply- 0”.

**Others**: Messages that could not belong to either of the categories above will be categorised here. These email messages that are in this categories are: email messages from friends that does not require any urgency, auto respond reply messages, advertisement email messages, junk emails, email messages with email address: nonreply@myname.com, informative emails etc.

**REFERENCES**

Intelligent Email Prediction System (IEPS) by Taiwo Ayodele, Shikun Zhou, Rinat Khusainov, *Department of Electronics and Computer Engineering*, *University of Portsmouth, United Kingdom*

Mark, D., John, B., Fernando, P. Reply Expectation, Prediction for Email Management In the 2nd Conference on Email and Anti-Spam (CEAS 2005). 2005. Stanford University, California, USA.