Applying Machine Learning Techniques for Email Reply Prediction

Taiwo Ayodele , Shikun Zhou, Rinat Khusainov

Abstract—For several years now, email has grown rapidly as the most-used communications tool on the internet. One advantage of the Internet is the ease with which people can communicate online. The popularity of online communication has created an explosion of users who regularly access the internet to connect with others. Many people use email to stay in touch with relatives and friends who live far away geographically. We propose a new framework to help prioritised email better using machine learning techniques; an intelligent email reply prediction system. Our goal is to provide concise, highly structured and prioritised emails, thus saving the user from browsing through thousands of emails and help to reduce time spent on checking and reading email messages.

Index Terms—**Email reply prediction, email messages, machine learning, and email features**

I. INTRODUCTION

Email is the most common method of communicating online and it is also a method of creating, transmitting, or storing primarily text-based human communications with digital communications systems. Every day, Internet users send each other billions of email messages. If you are online a lot, you yourself may send a dozen or more emails each day without even thinking about it. Obviously, email has become an extremely popular communication tool. Have you ever wondered how hundreds of emails received per day could be well organised, highly prioritized and be predicted if they require a reply.

Email prediction is a method of anticipating if email messages received require a reply or did not require any urgent attention. Our email prediction system will enables email users to both manage their email inboxes and at the same time manage their time more efficiently. Bradley et al [2] implemented remembrance agent to analyse documents and predict useful information from documents that users frequently use and explicated that Remembrance Agent (RA) is a program which augments human memory by displaying a list of documents which might be relevant to the user's current context. Unlike most information retrieval systems, the RA runs continuously without user intervention. Its unobtrusive interface allows a user to pursue or ignore the RA's suggestions as desired. This idea was implemented in information retrieval and his approach relies on continuous searches for information that might be of use in its user's current situation. For example, while an engineer reads email about a project the remembrance agent reminds her of project schedules, status reports, and other resources related to the project in question. When she stops reading email and starts editing a file, the RA automatically changes it recommendations accordingly.

The existing solutions by Joshua et al [7] explained that "A study of email responsiveness was conducted to understand how the timing of email responses conveys important information. Interviews and observations explored users' perceptions of how they responded to email and formed expectations of others' responses to them. We identified ways in which users maintain and cultivate a responsiveness image for projecting expectations about their email response" This work grew from the belief that an interesting, relatively unexplored aspect of email usage is its implicit timing information". Also Dredze et al [3] provided solutions to email reply prediction by assessing date and time in email messages as email containing date and time are time sensitive and may require a reply, and finally used logistic regression with other feature like questions in email message and many more to provide solutions to email reply predictions. Other studies have focused on how people save their email, what purposes it serves for them, and its importance as a tool for coordination in everyday life [5, 6, 8, 9, 10, 11].

This paper proposes to solve the problem of un-structured email messages and overload by determining if email received needs reply. Our intelligent prediction system provides a better and efficient way of prioritizing email messages and provides a new solution with new approaches to email reply prediction.

ISBN: 978-988-17012-5-1

II. PREVIOUS WORK

Because email is one of the most used communication tools in the world, Sproull and Kiesler [9] provide a summary of much of the early work on the social and organizational aspects of email. Here we will focus on work about email reply prediction strategies, as well as research dedicated to alleviating the problem of "email overload and prioritization." Mackay [8] observed that people used email in highly diverse ways, and Whittaker and Sidner [1] extended this work. They found that in addition to basic communication, email was "overloaded" in the sense of being used for a wide variety of tasks-communication, reminders, contact management, task management, and information storage.

Mackay [8] also noted that people fell into one of two categories in handling their email: *prioritizers* or *achievers*. Prioritizers managed messages as they came in, keeping tight control of their inbox, whereas achievers archived information for later use, making sure they did not miss important messages.
Tyler and Tang [7] in a recent interview study identified several factors that may influence likelihood of response. These empirical studies were qualitative, generally based on 10 to 50 interviews.

III. METHODOLOGY

We used machine learning techniques to learn and extract email features: *subject field, senders' domain address, CC/BCC field, email content* and be self-improved for increase in the efficiency and effectiveness of email message classification. Our machine learning evolves round determination of which email require a *reply* and those that does not require a *reply*.

We used interviews and qualitative observations to study features of email massages. This study was conducted in two phases using ***survey system***, ***observation-based interviews*** as a way to elicit users' perceptions and attitudes about email usage. The two stages are:

- **Survey System:** We conducted several feed back on what email users want their email client to do- *organised email messages better, prioritized their email messages, check the header fields, check the attachment fields* etc. Our web based survey system was filled by over 8000 email users from across the world. The survey results represent users from various professional fields- IT Professionals, Engineering, academic (Art, Science majors, Business, Leisure and tourism, Banking and Finance etc), Business owners, heath care and many more which represents the idea of most email users as explained below:
  - 98% want their email client to tell them which email requires reply and which one does not require a reply.
  - 90% want an email intelligent system to access their email header fields- *subjects fields*, *CC/BCC field*, *Attachment field*, *email content* which will allow decision making on Email Reply Prediction System.
  - 89% say they want email client that could classify emails with a word with real meaning (Critical, Urgent, very important, Important)
- **Observation based interviews:** These interviews and observations were designed to broadly explore the concept of email features extractions and how users convey information through the subject field, vocabularies and phrases used in their email, and what they learn from other selected features (senders' email address, CC/BCC, Attachment if there is any) extractions of emails. We asked about:
  - The usage of their emails
  - How import is their emails to their life, business and leisure
  - Have they suffered any loss from not replying or responding to some mails
  - What do they use email for- *task management tool, archiving tools* etc.
  - How often do they reply their emails
  - How many emails do they receive in average a day
  - Will they prefer an email system that can predict those emails that require responses
  - If they have an email client that can rank emails, will they prefer numeric 1,2,3) ranking, alphabetical (A, B, C) or word with meaning (Urgent, Important)
  - When and how often do they reply emails etc.

We implemented a machine learning approach to solve the problem of email reply predictions. Machine learning is learning the theory automatically from the data, model fitting, or learning from examples. It is also an automated extraction of useful information from a body of data by building a good probabilistic model.

A. Importance of Machine Learning

Our work involves machine learning because it is the underlying method that enables us to generate high statistical output. These are the importance of machine learning as applied in our work:

- New knowledge about tasks is constantly being discovered by humans. Like vocabulary changes, and there is constant stream of new events in the world. Continuing redesign of a system to conform to new knowledge is impractical, but machine learning methods might be able to tract much of it.

- Environments change over time, and new knowledge is constantly being discovered. A continuous redesign of the systems "by hand" may be difficult. So, machine that can adapt to changing environment would reduce the need for constant redesign.

- Some tasks cannot be defined well, except by examples and large amounts of data may have hidden relationships and correlations. Only automated approaches may be able to detect these. Figure 1 show email feature extraction approach. A schematic diagram of the architecture for email words extraction from incoming email messages for efficient reply prediction proposes in this work.

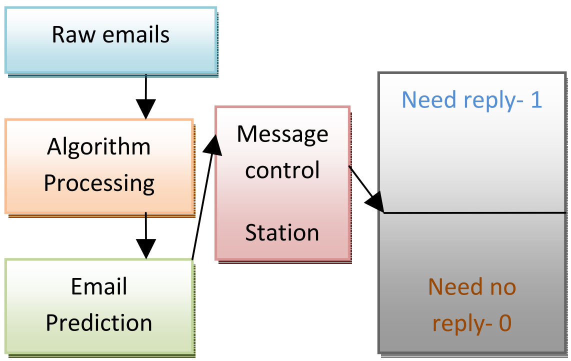

Figure 1: Architecture for words extraction from incoming email

Our proposed prediction system accept email messages as input data and emails are passed unto our machine learning prediction algorithm system, email header features are obtained from each emails and the predictor determines in numeric values the mails that require replies and the emails that does not require replies as shown above.

B. *Email Reply Prediction System (ERPS)*

This is a decision making system that could determine if emails received require a reply. For any given email datasets, there are multiple email conversations and to capture these different conversations, we assume that if one email was a reply to the sender's original message, then such a mail may require attention as this may have element of *request* and section 3.2.1 explains more. We also developed a scoring method for the extracted email features to determine the accuracy of our prediction system.

C. *Email Header Fields*

One approach we used was a "Subject content walkthrough." Each of the human participants opened their mail boxes with the subject field opened and ask each of them to describe the relationship with the sender of each message. In some cases this walkthrough was done on the email outbox as well as current emails messages.
At first, we conducted the subject content walkthrough on a per-message basis. As we discovered that email subject fields are based much more on relationships than isolated messages, we began asking users to sort the inbox by sender (to see many or all messages from a particular sender). We then focused the interview on the relationships represented by the messages—it became more of a "relationship walkthrough." And this gave us a better idea about the importance of subject fields in relations to how close you are to the sender and helped us to conclude that subject field of an email is important to our algorithm prediction system

Also, another approach we used was "Senders' email address, CC/BCC, Attachments walkthrough". Each participant checked who sent the mail and most of them realised that some emails are from personal friends and some are from work places and other are from online transactions- flight booking confirmations, online orders etc. Only few emails based on their job description or roles are from- *head of departments*, *manager of a company*, *administrators* e.g. admin@edu.ac.uk,hod@uni_port.ac.uk, finance@bank.net etc. Such a mail from head of departments, financial institutions-*credit card company* may require a response. We are aware that sender's email address is not enough to determine if such a mail require a reply and that is why we develop a scoring mechanism to assign values to what field should be valued higher and what should not. Cc/Bcc header field according to our participants show that if there is any email address in this fields, there is high possibility that such a mail require other people's attention and that's why the sender copied those concerned. Some of the participants that have attachments in their emails explained that most of them are work related e.g. PDF files, some are photos, majority are urgent CVs and Letters. This shows that attachments could be *work related walkthrough* or *personal related walkthrough* This makes email attachment fields to be relevant and important to our solution because any attachment found in email messages could be work related as *an human recourses officer, personnel manager, job recruiter agent* or could be personal walkthrough related as someone who uploads family photos or historical photos for people to see and admire.

D. *Email Content*

One of the techniques used here was "phrase selection". The participants opened their mail boxes and check contents of each email messages and select relevant phrase such as: *please reply soon, when should I send my CV, is there any vacancy, when is the meeting, looking forward to your reply,* *confirm your booking, please make a payment, your deadline is, only three days left* and many more. These aforementioned phrases were identified from most email boxes and we realised that email users use some phrases or words to create attention and also most frequent vocabularies are checked and this shows a *vocabulary to phrases walkthrough* in the content of the email messages. From the participants, we conclude that phrases used in email messages are very important and will be useful to determine if a mail needs a reply or if a mail does not need a reply. We further implement a scoring mechanism to handle all the email header fields that we explore.

IV. SCORING METHOD

Our approach analysed the feature of email contents and other header fields namely; *phrases, interrogative words, questions and question mark, attachments, early communications of senders* and many other aforementioned features in section 3.2 above. Our algorithm prediction system (APS) performs unsupervised scoring methods using weighting measures [4]. All new emails have number – score. Then more score then more email need reply. We calculate the weighting scores on the features of the email by implementing a method called "**the inner product**" with its elements. We collect n numbers of emails using this function below:

$$\boxed{S_{q,e} = \sum_{t \in T_{q,e}} \left(w_{q,t} . w_{e,t} \right)}$$

Here, $w_{e,t}$ is the email-term weight while query-term weight is denoted by $w_{q,t}$ and we also denote these various set:

- the set E of emails;
- for each term t, the set Et of emails containing t;
- the set T distinct terms in the database and
- the set T_e of distinct terms in emails e, and similarly T_q for queries and $T_{q,e} = T_q \cap T_e$

The terms are the features extracted to determine the email prediction namely: *phrases, interrogative words, question marks, attachments* etc. When the formula above is applied, the average weighting score is calculated for each email and if it is above the set threshold, then that mail will be categorized as need reply or do not need reply (*need no reply)* as given relevant item is retrievable without retrieving number of irrelevant items.

Our predictor assigns a weight score to any question (s), question mark (s) found in email subject as well as contents of the mail. For example: A question in the subject has a weight score of 3 point of value and a weight score of 2 in the body of the email message. Do note that a question is a sentence that ends with the sign "?" and start with an interrogation pattern like: "where", "when", etc. Also, a score of 1 is assigned to the following sample features: "*if communication with sender was earlier (*"Re:"-letters)", emails from specific domain (.ac.uk, .edu), phrases such as "*please reply soon",* if there is email address in cc or bcc, all these are assigned a score of 1. The prediction analysts concluded that the maximum weight score that could be assigned to every email is 10 and choose 7 as the threshold weighting score that a mail must attain before it could be grouped as "*need reply- 1"* and any email that does not measure up to the threshold will be re-examined and if other factors have been re-assessed and could not meet up with the threshold at the second attempt, then it will be grouped as "*need no reply- 0.*

A. Email Prediction Methods (EPM)

Email space is a function of the manner in which terms and term weights are assigned to the various emails with an optimum email space configuration that provides an effective performance. If nothing is known about the emails under consideration, it suggests that ideal email space is one where emails are jointly relevant to certain user queries and such mails are predicted together ensuring that they will be retrievable jointly in response to the corresponding queries.

Inner product space [4] is a vector space of arbitrary (possibly infinite) dimension with additional structure, which, among other things, enables generalization of concepts from two or three-dimensional Euclidean geometry. The additional structure associate to each pair of vectors in the space is called the inner product (also called a scalar product) of the vectors as shown in the formula below:

a = [a1, a2, , an] and b = [b1, b2, … , bn] is defined as:

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{n} a_i b_i = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$$

For example, the dot product of two three-dimensional vectors

[1, 3, −5] and [4, −2, −1] is

$$\begin{bmatrix} 1 & 3 & -5 \end{bmatrix} \cdot \begin{bmatrix} 4 & -2 & -1 \end{bmatrix} = (1)(4) + (3)(-2) + (-5)(-1) = 3.$$

For two complex vectors the dot product is defined as

$$a \bullet b = \sum a_i \overline{b}_i$$

Where is the complex conjugate of bi. The absolute avoids two weights cancel each other and that enables us to avoid negative weight measures and correct errors in the weighting system

Since our annotated emails from Enron corpus [12] are treated like a bulk of dataset, we used term weighting with unsupervised techniques with our approach of heuristic techniques to provide a well organised and prioritised email prediction system.

B. Algorithm Prediction system (APS)

Algorithm prediction system uses a heuristics-based approach with aforementioned email features that was extracted, with weighting measures. The assumption is that if interrogative words, questions, questions mark(s), phrases such as *do reply, when will you, if time* is found in email messages, such a mail is important and will be assigned some score. Figure 2 shows our algorithm as other intelligent technique is kept as black box behind the system. Figure 2 shows the algorithm below.

Reply Prediction algorithm

1. Define X as the number of matching needed to mark the message needs reply
2. Define Count as the number of matching =0
3. If CC or BCC contains email addresses then
   - **a. Count = Count+1**
4. **create a rule that**
   - a. If the contents contains some of these words
     - i. **Count** = **Count**+1
   - b. must, should, what about, meeting ,priority,
     - i. **Count** = **Count**+1
   - c. Dear, hello, hi
     - i. **Count** = **Count**+1
   - d. Multiple of "?"
     - i. **Count** = **Count**+1
   - e. Dates or months names
     - i. **Count** = **Count**+1
   - f. AM,PM
     - i. **Count** = **Count**+1
5. if(**Count** > **X**)
   - a. then mail need reply
   - b. Else
   - c. mail doesn't need reply

}

Figure 2: Algorithm prediction System

Algorithm prediction system (APS) for email management is a new unsupervised machine learning techniques that is implemented. APS described above uses a precision and recall to evaluate this new technique in comparison with gold standard- *human participants.*

V. EVALUATIONS AND RESULTS

In order to compare different approaches of email reply prediction, a gold standard is needed. In practice, for comparing extractive predictor, we tested our algorithm performance with 7000 annotated emails from 70 human participants to:

- ➢ *Need reply*
- ➢ *Need no reply*

We tested our algorithm with the embedded similarity measure approach on the 7000 email datasets. To measure the quality and goodness of the email prediction, gold standards are used as references. It is noticed that our unsupervised machine learning approach achieved 98% accuracy in comparison to the gold standard. Our sample graphical prediction client output is shown in figure 3.

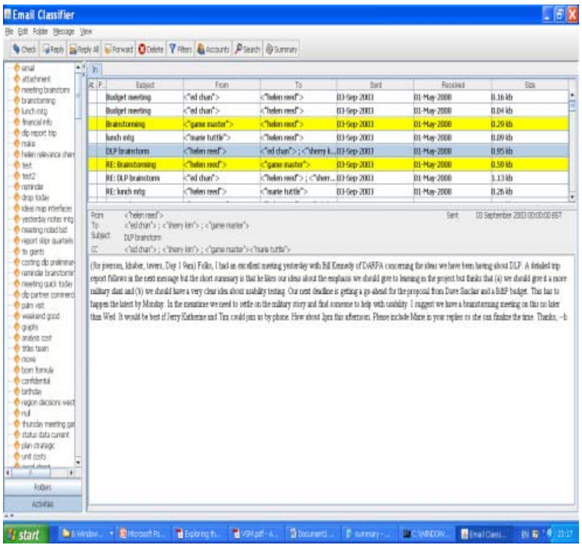

Figure 3: A sample email prediction system

This section describes experiments using APS system to automatically induced email features classifiers, using the features described in Section 4. Like many learning programmes, APS takes emails as input and the classes to be learned, a set of features names and possible values and training data specifying the class and feature values for each training example. In our case, the training examples are the Enron email datasets [12]. APS outputs a classification model for predicting the class (i.e *need reply- 1, need no reply- 0).* We obtained the results presented here using precision and recall. In this paper, we evaluated APS system based on weighting measures, and human judgments. We show results of 7000 annotated emails and different feature set in figure 4.

We evaluated our email reply prediction system on over 7000 email Enron datasets from over 120 email boxes owned by 200 people from Enron Corpus [12] using precision and recall. We also evaluate our algorithm prediction system using precision and recall as the measurement of evaluation for our system:

$$\text{Recall} = \frac{\text{group found and correct (needs reply)}}{\text{total group correct (rightly predicted)}}$$

$$\text{Precision} = \frac{\text{group found and correct (needs reply)}}{\text{total group found (Total email found)}}$$

ISBN: 978-988-17012-5-1

Precision and recall (compared to gold standard)

Correct predicted group	Total Predicted Group Found	Total Emails	Precision	Recall
309	316	382	98.0%	80.9%

Precision and recall on a per email basis

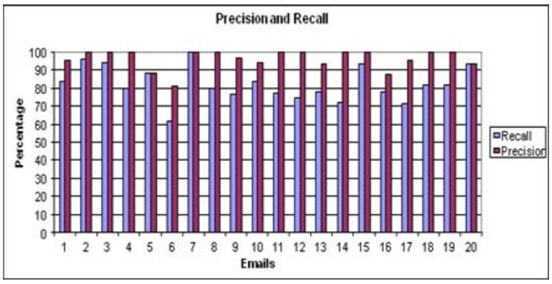

Figure 4: Evaluation Result

Figure 4 also shows our evaluation test results with the accuracy of our precision and recall evaluation on 7000 emails but we show only results of approximately 400 email datasets because of limited space. The email prediction system relies on a simple algorithm but it is very complex to implement.

VI. Conclusions

With our findings, we concluded that without any prior experience to establish an expectation and reply prediction, email users can face problem with deciding which email to respond to especially if one receive hundreds of emails per day coupled with dilemma of how long to wait for a response before deciding that follow-up actions are required. Based on the survey and observation based interviews that was carried out, we then develop and customise an intelligent email reply prediction system based on survey and interviews conducted, and implemented what email users want in an email client that could determine mails that require a response- *email reply prediction system*. So, we are able to build a system that is intelligent enough to manage users email messages on their behalf. This effective and efficient solution will help business organizations, higher institutions and email users to reduce un-necessary time spend on sorting their emails out, reduce cost, reduce email overloads and are better used as effective archiving tools.

REFERENCES

[1]. Steve, W., Candace, S. (1996). *Email overload: exploring personal information management of email.* In the Proceedings of the ACM CHI 96 Human Factors in Computing systems Conference, ACM Press, pp.276-283. [2]. Bradley, J.R., Thad, S. (1996). *The remembrance agent: A continuously running information retrieval system.* In the Proceedings of the First International Conference on Practical Applications of Intelligent Agents and Multi-Agent Technology (PAAM'96), London, pp.486–495.

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

[4]. Salton, G., Wong, A., Yang, CS. (1975). *A vector- space model for automatic indexing*. Communications of the ACM, 18(11), pp.613-620

[5]. Laura, D., Gina, V., Cadiz, J,J. (2003). *Marked for deletion: An analysis of email data*. In the Conference on Human Factors in Computing Systems CHI '03 extended abstracts on Human factors in computing systems **(**CHI 2003), ACM Press, pp.924-925.

[6]. Ducheneaut, N., Belloti, V. (2001). *Email as habitat: An exploration of embedded personal information management*. ACM Interactions, 8(5), pp.30-38.

[7]. Joshua, R.Tyler, John, C.Tang. (2003). When can I expect an email response? A study of rhythms in email usage. In the *Proceedings of the Eighth Conference on European Conference on Computer Supported Cooperative Work*, Kluwer Academic Publisher, Norwell, MA, USA, pp.238-258.

[8]. Mackay, W. (1998). Diversity in the use of electronic mail: A preliminary inquiry. ACM Transactions on Office Information Systems, ACM Press, New York, NY, USA, 6(4), pp.380-397.

[9]. Sproull, L., Kiesler, S. (1991). Connections: New ways of working in the networked organization. MIT Press: Cambridge, MA, USA.

[10]. Mary, S. (1985). *The impact of electronic mail on managerial and organizational communications. In the Proceedings of ACM SIGOIS and IEEECS TCOA Conference on Office Information Systems.* ACM Press, New York, NY, USA. pp.96-109.

[11]. Robert E. Kraut., Paul, A. (1997). *Media use in a global corporation: Electronic mail and organizational knowledge, In Culture of the Internet*. Lawrence Erlbaum, Associates, Mahwah, NJ, USA. pp.323-342.

[12]. Bryan, K., Yiming, Y. (2004). The Enron corpus: A new dataset for email classification research. In European Conference on Machine Learning.