Unsubscription: A Simple Way to Ease Overload in Email

Iftah Gamzu
iftah.gamzu@yahoo.com
Amazon, Israel

Liane Lewin-Eytan
lliane@amazon.com
Amazon, Israel

Natalia Silberstein
natalias@oath.com
Yahoo Research, Israel

ABSTRACT

The constant growth of machine-generated mail, which today consists of more than 90% of non-spam mail traffic, is a major contributor to *information overload in email*, where users become overwhelmed with a flood of messages from commercial entities. A large part of this traffic is often junk mail that the user would prefer not to receive. Surprisingly, nearly 95% of this traffic is in fact solicited by the users themselves in the form of subscriptions to mailing services. These subscriptions are many times unintentional. Although unsubscription option from such services is enforced by commercial laws, it is hardly actually used by users.

We perform a large scale study of *unsubscribable* traffic, namely, messages that provide unsubscription option to users. We consider users behavior over such traffic in Yahoo Web mail service, and demonstrate a significant gap between users low interest in this traffic, and their lack of active behavior in decreasing its load. We conjecture that the cause of this gap is the lack of an efficient and easily accessible mechanism that would help users to unsubscribe. We validate our conjecture with an online large scale experiment, where we provide users with a novel mail feature for managing unsubscribable traffic, based on personalized recommendations. The experiment demonstrates the imminent need that exists for such a mechanism.

KEYWORDS

Machine-Generated Mail; Mail Mining; Unsubscription Recommendations

ACM Reference Format:
Iftah Gamzu, Liane Lewin-Eytan, and Natalia Silberstein. 2018. Unsubscription: A Simple Way to Ease Overload in Email. In *WSDM 2018: WSDM 2018: The Eleventh ACM International Conference on Web Search and Data Mining, February 5–9, 2018, Marina Del Rey, CA, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3159652.3159698

1 INTRODUCTION

Machine-generated emails, which comprise today more than 90% of non-spam Web mail traffic, have been the subject of a line of studies during the last few years [4, 16, 18]. These messages are sent at large scale across the user population, and vary in importance: from critical invoices, flight itineraries, and e-tickets, to promotions or newsletters that will remain unread by most users. Machine-generated traffic defines the nature of Web Mail today, and is the source of its growth. Indeed, email volume keeps growing, with an increase of 5% in worldwide emails during the past two years and a prediction of 4.4% increase per year up to 2021 [33, 34].

As a direct consequence, mail users are becoming overwhelmed with a flood of messages generated by organizations and commercial entities. Those are often junk mail that the user would prefer not to receive. Surprisingly, as we show in this paper, nearly 95% of the machine-generated emails are explicitly solicited by the users themselves in the form of subscription to mailing services. Importantly, the user might not even be aware of having subscribed to some of those list in the past. According to a user study from 2015 [27], nearly a quarter of all mailing list subscriptions are unintentional. For example, the result of auto-subscription of users to mailing lists. Other frequent reasons that lead to subscriptions by user interest in getting information about deals and special offers, news updates, and access to interesting articles and content.

A trivial yet important observation is that mailing lists that allow subscriptions also allow unsubscriptions. In fact, providing the user with the possibility to unsubscribe from such services is required by the *CAN-SPAM Act*, which is a law that “establishes requirements for commercial messages, gives recipients the right to have you stop emailing them, and spells out tough penalties for violations" [2]. Moreover, the unsubscribe option has also been defined and formalized in dedicated IETF RFCs [21, 23]. In particular, these RFCs define a *List Unsubscribe* header, which is an optional chunk of text that email publishers and marketers can include in the header portion of the messages they send. It is a recommendation aimed for senders to reduce complaints, improve deliverability (by avoiding spam filters that often take this header into consideration), and improve the experience of subscribers. In most cases, the unsubscribe option is presented to the user as an unsubscription link within the message body, which points to a designated unsubscription page in which unsubscription can be completed. We refer to messages of this type as *unsubscribable.*

Understanding the properties of this traffic is a major step towards devising methods and tools for easing information overload in email [13, 19, 40] caused by irrelevant messages. In the first part of the paper, we address this need by conducting a large-scale study of real mail traffic of Yahoo Web mail service. We focus on unsubscribable traffic and users' behavior over this traffic. Notably, although 85% of the entire mail traffic is unsubscribable, and despite the fact that more than 90% of this traffic remains unread, the unsubscribe option is triggered for less than 2% of all mailing lists. Our study also demonstrates that unsubscriptions are performed by only 3% of the users. One may argue that spam votes are used as proxy to mark and filter irrelevant traffic. Although the spam

This research was done while all authors were associated with Yahoo Research, Israel.

WSDM 2018, February 5–9, 2018, Marina Del Rey, CA, USA

ACM ISBN 978-1-4503-5581-0/18/02…$15.00
https://doi.org/10.1145/3159652.3159698

action has a higher usage, it is still triggered by only 5% of the users with respect to 3% of mailing lists.

These observations clearly show an unexplained phenomenon, demonstrating low user interest in many mailing lists without consistent active behavior to decrease the load created by those lists. This may be the result of a lack of a proper mechanism that would allow the user to efficiently stop such traffic. In contrast to popular mail actions, like read, reply, delete and mark-spam, unsubscribe is not a *mail-native* action. That is, it is not provided by the mail service, but rather by the sender of the message. Although the CAN-SPAM act requires that users are given a "clear and conspicuous explanation of how they can opt out", the unsubscribe link is often rather hard to find, appearing in small fonts in a marginal location within the message. In addition, it usually requires more steps to be performed by the user to complete the unsubscription, turning it to a cumbersome operation that users do not generally initiate.

In the second part of our work, we present a novel mail feature, providing the users with a comprehensive and convenient mechanism for managing unsubscribable traffic. Our solution is based on personalized recommendations for unsubscribing from mailing lists that are likely irrelevant for the user. We validate our solution in an online experiment conducted in Yahoo Web mail service, comprising millions of users. This experiment demonstrates that there is an imminent need for a mechanism that can help users unsubscribe from mailing lists and decrease overload in email.

The contributions of our work are thus threefold: (1) we provide an analysis of unsubscribable traffic, as well as users engagement with such traffic, (2) we develop a machine learning approach for personalized ranking of unsubscription recommendations, and (3) we perform a large-scale online experiment in Yahoo Web mail service, surfacing our unsubscription mechanism to users and validating its effectiveness. The rest of this paper is organized as follows. Section 2 covers related work in the context of mail management and overload in mail. Section 3 characterizes unsubscribable traffic and discuss its identification. Section 4 provides a study of unsubscribable traffic and related behavioral patterns of users. In Section 5, we tackle the problem of computing personalized unsubscription recommendations, while in Section 6, we discuss our online experiment. Finally, conclusions are provided in Section 7.

2 RELATED WORK

Mail has been a fertile ground for research in data science during the last two decades. A lot of work has been done, spanning a wide range of use cases, studies and applications, including mail prioritization, mail classification, mail search, analysis of mail behavioral patterns, and more. Many of these studies are related to the problem of mailbox management under overload in email [13, 19, 40].

Wang et al. [37, 38] develop a prioritization framework for broadcast emails, which are similar in nature to unsubscribable traffic. However, their approach is inherently different than our approach which makes effort to remove irrelevant traffic rather than to prioritize it. Mail prioritization and management, which study solutions for assisting users in processing their inbox, have been investigated in many papers [1, 12, 14, 15, 32, 35, 36]. The work of Aberdeen et al. [1] deals with the notion of message importance, based on the propensity that a user will perform specific actions on the message. Di Castro et al. [15] study different mail actions and devise a learning framework for predicting them. One action attracting much attention is the reply action, which has been the focus of several papers [17, 26, 31, 36, 41]. Other works such as [7, 20] consider mail as a task management resource, and study email behavior over messages that relate to tasks or events.

Classification of mail traffic as a tool to assist users has also been given a lot of attention. Multiple classification methods have been proposed for automatically assigning messages to pre-defined folders or labels [5, 18, 24, 25, 39]. One specifically important classification task is the distinctions between human and machine-generated traffic [4]. This fundamental distinction underlies our work for analyzing the largest type of machine-generated mail.

Mail search is another relevant line of research as an essential mean to efficiently access relevant information and cope with the growing volume of mail traffic. It is the default discovery paradigm for retrieving past mail, instead of exploiting organizational means such as folders, which are hardly used [9]. As such, it has attracted attention during the last few years, both in the context of ranking and of query suggestions [8–11]. The works of [3, 30] also contribute to this theme by providing a general study of mail search, activities that users perform over email, and their search behavior.

Most of the papers mentioned above attempt to alleviate the problem of information overload in email through prioritization, classification, search, and more. Yet, to the best of our knowledge, there is no work that directly considers unsubscribable traffic, or leverages the ability that users have to simply unsubscribe from superfluous mailing lists.

3 IDENTIFYING UNSUBSCRIBABLE TRAFFIC

A key concept in this paper is that of a *mailing list*. A mailing list is a collection of email addresses, also called subscribers, used for the widespread distribution of some information. As the list of subscribers is not publicly available, a natural question regards the identification of the mailing list which underlies a given email message. Although there are formal suggestions for labeling messages with a mailing list identifier [22], in practice, most guidelines are not followed, and thus identifying the underlying mailing list is a non-trivial task. There are proprietary solutions that are highly technical, but simple heuristics also provide good identification guarantees. One primary example is to associate a mailing list with its mass sender [37, 38]. This is also the approach we follow in this paper. Namely, we regard the *subscription entity* as the email address of the sender of an unsubscribable mail.

We identify a message as *unsubscribable* using two ways: we either identify unsubscription links within the message body, or establish that it has a *List Unsubscribe* header [21]. Note that this header can either contain an email address link (e.g., "mailto:..."), which can trigger an unsubscription mail message on behalf of the user, or an http link (e.g., "http://..."), which can transfer the user to a designated unsubscription page of the underlying entity. We refer to the first option as *email-based method* for unsubscription, while the latter option is referred to as *page-based method*.

While identifying the header part is easy, discovering unsubscription links in an email body is a more challenging task that requires applying classification techniques. We thus developed a classification mechanism for identifying such unsubscription links in the messages body. The classification is based on multiple features, like relevant indicative words (such as “unsub" or “optout") in the link itself as well as in the text appearing in its close proximity, link location within the message, and more. Note that it is important to achieve high precision in this task to avoid misclassification of links. Therefore, we tuned our classification mechanism such that it achieves precision of more than 0.98 with recall of 0.88. Due to space limitations, we omit the details of this mechanism.

We analyzed the entire inbound traffic of Yahoo Web mail service, and observed that about 75% of all traffic contains a List Unsubscribe header. Furthermore, more than 10% of inbound traffic that does not contain such a header, contains unsubscription links in the message body. Taking into account that about 90% of today's email traffic is machine-generated [16], and that unsubscribable messages clearly belong to this type of traffic, one can infer that almost 95% of machine-generated traffic is unsubscribable.

Figure 1 presents the overall inbound traffic with respect to the two above-mentioned types of unsubscribable traffic: messages that have a List Unsubscribe header and messages that contain only unsubscribe links in their body (no header). Note that the data was collected over a period of four weeks, from February 23, 2017 to March 22, 2017, and the values are normalized with respect to the maximum daily value of unsubscribable traffic having a List Unsubscribe header. As can be seen, the amount of unsubscribable traffic with no List Unsubscribe header is considerably smaller compared to the traffic that has such a header. This implies that such unsubscribable traffic has only a small impact on all computations, and does not change the general observed trends.

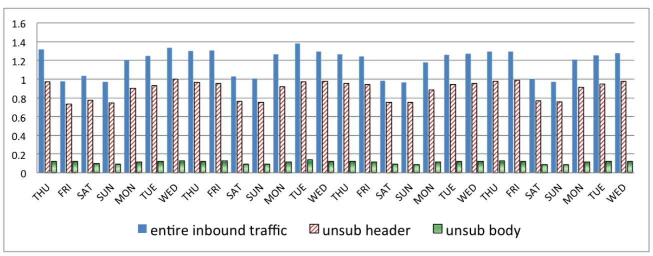


Figure 1: Normalized volume of inbound traffic, unsubscribable traffic with List Unsubscribe header, and unsubscribable traffic with only unsubscribe body links.

4 ANALYSIS OF UNSUBSCRIBABLE TRAFFIC

We provide a large-scale study of unsubscribable traffic and the behavior of users on that traffic.[1] We analyze the inbound traffic of a Yahoo Web service over a period of 28 days, from February 23, 2017 to March 22, 2017. We consider tens of millions of users that were active during this period, namely, performed at least one action over those four weeks (e.g., message read, message delete, search). The inbound traffic of those users originated in above 40M unique senders from about 3M domains.

[1]We note that all processes performed as part of our analysis were conducted under full privacy preservation in accordance with the mailing service privacy policy.

We begin by presenting a high-level analysis of unsubscribable traffic. We first focus on users subscriptions. Figure 2 exhibits the percentage of users as a function of their number of subscriptions, and the volume of unsubscribable messages they received. Note that a user that receives at least one unsubscribable message from a subscription entity is assumed to be subscribed to that entity. As expected, we observe a long-tail distribution in both cases, with 55% of users receiving emails from more than 20 subscription entities, and 50% of users receiving more than 150 emails over the analysis time period. This observation reaffirms that unsubscribable mail is a major contributor to information overload in email across the general user population.

We turn our view to the domains from which unsubscribable traffic originates. Table 1 presents the 5 domains that send unsubscribable messages to the highest number of users, as well as the 5 domains that have the highest number of users clicking on unsubscription links in their messages[2]. If unsubscription was a uniformly-random process then the top domains by number of subscribed users would have also appeared as top domains by unsubscription clicks. As can be seen, this is not true in general. This is rather expected as there are multiple factors that influence users tendency to click on unsubscription links. In what follows, we examine different user properties and actions in mail, and analyze their correlation with subscription and unsubscription behavior. This analysis establishes the basis for our unsubscription recommendations method, which is presented later on.

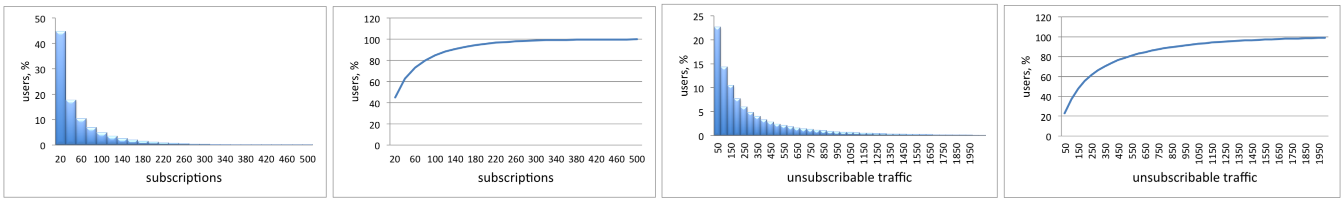
Subscribed users reach	Unsubscription clicks
facebookmail.com	facebookmail.com
linkedin.com	linkedin.com
twitter.com	r.groupon.com
mail.instagram.com	reply.bronto.com
explore.pinterest.com	yahoogroups.com

Table 1: Top 5 domains with respect to (1) the number of subscribed users, and (2) the number of users clicking on unsubscription links.

4.1 Unsubscribable traffic by user groups

We analyze unsubscribable traffic with respect to different user demographic sectors and activity levels. We consider four levels of activity, defined by the number of days that a user has been active during a month. A user that performed at least one action in mail during a day is regarded as active during that day. The four levels of activity are L = "Low", if the user was active between 1 and 3 days inclusive, M = "Medium" if she was active between 4 and 14 days, H = "High" for 15 to 25 days of activity, and U = "Ultra" for 26 days or more. With respect to demographics, we partition the users into seven age groups (namely, younger than 18, 18–24, 25–34, 35–44, 45–54, 55–64, and older than 65), and two gender groups (that is, men and women). Note that the sizes of groups in each partition vary. Still, as each group is large enough (consisting of millions of users), statistic differences between groups are meaningful.

[2]Note that we treat all unsubscription clicks equally since we generally cannot track the unsubscription process that follows them. In particular, we cannot know if they resulted in unsubscription or some other update in the subscription preferences.



(a) (b) (c) (d)

Figure 2: (a,b) PDF and CDF of users as a function of the number of unique subscriptions, (c,d) PDF and CDF of users as a function of their amount of unsubscribable traffic.

Figure 3 exhibits the mean volume of unsubscribable traffic and mean number of subscriptions for each group of users. Note that presented values are the result of normalizing the observed mean values with the corresponding means of all users, where the average user receives 328 unsubscribable emails from 50 senders. Interestingly, we observe that women receive 1.6x more unsubscribable messages from 1.4x more subscription entities, compared to men. Also, older users tend to receive more unsubscribable messages from more subscription entities. In particular, the youngest group of users receives about 4x less emails from about 5x less subscription entities than the oldest group of users. User activity level does not exhibit a clear trend with unsubscribable traffic, although there are clear differences between the different groups. All these observations indicate that different user groups have different subscription and management patterns which are of importance for the task of unsubscription recommendations.

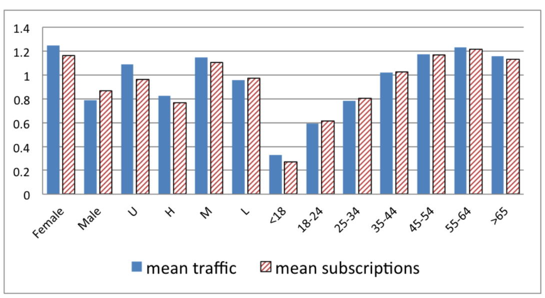


Figure 3: Normalized mean number of unsubscribable messages and subscriptions for different user groups.

To further emphasize this point, we focus on the gender group, and analyze subscriptions with respect to sender domains from few categories. Two examples of domain categories are shopping and travel. We follow Grbovic et al. [18] approach for domains categorization. Figure 4 presents the percentage of women and men that receive unsubscribable traffic from few popular domains in the *shopping* category. As can be seen, almost all domains have a significantly larger portion of subscribed women, indicating a fundamental difference in subscriptions between genders. Interestingly, a similar phenomenon occurs for popular domains in the *travel* category. For popular domains in the *social* and *career* categories, the percentage of subscribed women and men is essentially the same, with few exceptions, like `pinterest.com` and `linkedin.com` which have more tendency towards women and men subscriptions, respectively.

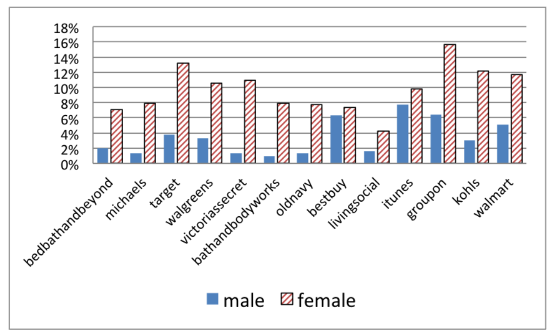


Figure 4: Percentage of women and men that are subscribed to popular shopping sites.

4.2 User actions on unsubscribable traffic

We now turn our attention to the user side. We consider unsubscriptions done by clicking on unsubscription body links as well as native mark-spam actions. Spam actions are of particular interest in our context since they could potentially reflect a similar intent of users that like to stop receiving irrelevant traffic. During the analyzed period of time, only 3% of the users clicked on unsubscription links. Those links were associated with only 2% of all subscription entities. In comparison, on the same time period, about 5% of the users used the spam action over 3% of the subscription entities.

Figure 5 exhibits a comparison between the average daily number of clicks on unsubscription body links and spam votes. Note that the values were normalized by the minimum daily number of unsubscription clicks. One can see that the daily number of spam votes is significantly higher than the number of unsubscription clicks. This difference could be explained by the ease of usage of mark-spam, which is a native mail action inherently provided by the mail service. On the contrary, unsubscription links require the user to open the email and look for an unsubscription link, which is sometimes not easy to find. This hypothesis is validated by our online experiment in Section 6.

We further compare unsubscription clicks and spam votes with respect to the different user demographic sectors and activity levels described before. Figure 6 presents the average number of unsubscription clicks and spam votes for the different user groups. The

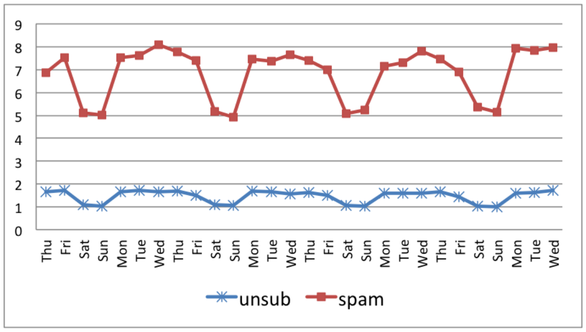


Figure 5: Normalized average daily number of unsubscription clicks and spam votes.

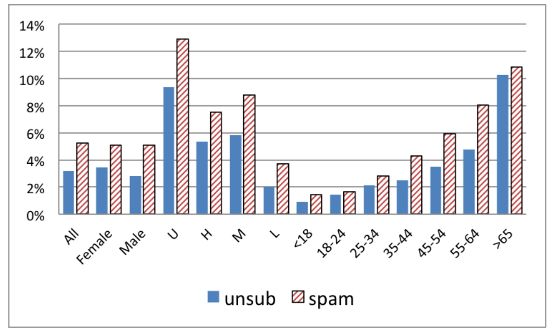


Figure 7: Percentage of users that performed unsubscription click and spam vote for different user groups.

presented values are the result of a normalization by the mean value of unsubscription clicks for the entire user set. The difference between unsubscription clicks and spam votes can be observed for all groups. Figure 7 presents the percentage of users that performed unsubscription clicks and spam votes for each user group. Notice that more users performed spam votes than unsubscription clicks across all user groups. However, the gap here is smaller than before. This implies, in conjunction with Figure 6, that the fact that users perform more spam votes than unsubscription clicks is not only true for the entire user population but also when considering different user groups. This substantiates again that spam actions are more natural in the mail context.

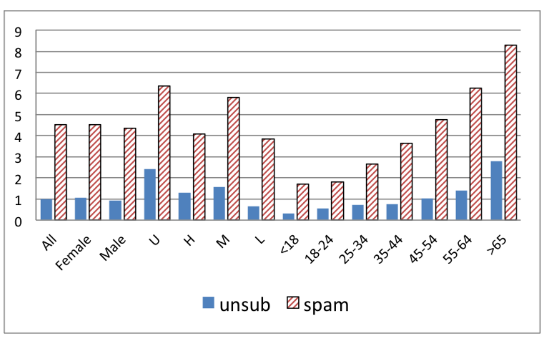


Figure 6: Normalized average number of unsubscription clicks and spam votes for different user groups.

We also consider few additional mail actions, namely, read, delete, and delete-without-read. Note that delete-without-read is simply a delete message action that was not preceded with the message being read [15]. It seems intuitive to assume that irrelevant messages should be less prone to be read and more prone to be deleted-without-read. Table 2 shows a comparison between the above-mentioned actions for unsubscribable traffic and traffic that does not contain unsubscription options. We refer to the latter traffic as "non-unsub". Remark that the table shows the percentage of messages that had each of the actions. As can be observed, only 9% of unsubscribable messages were read, while 13% of the messages were deleted without being read. Comparing this with the percentages for non-unsubscribable traffic, one can infer that unsubscribable traffic is considerably less relevant for users, thus, contributing to overload in email. In addition, those high-level statistics hint that those actions may help to identify subscriptions that are less relevant for a user.

action	unsub	non-unsub
Read	9%	23%
Delete	14%	14%
Delete-without-read	13%	11%

Table 2: Percentage of mail actions on unsubscribable and non-unsubscribable traffic.

To further emphasize the last point, we focus on several popular domains that had a high number of recipients (i.e., above 1M) in the considered time period. We selected those domains from few different categories. We decided to focus on such mass-senders to decrease the variance in our measurements, especially that of unsubscription clicks which are quite scarce. Figure 8 shows the average traffic load as well as the average number of read, delete, spam, and unsubscription actions per user for each of the selected domains. The average traffic load is defined to be the overall number of emails sent from a domain divided by its number of recipients. Note that the values in each curve line are the result of normalization with respect to the minimum (domain) value associated with that line. Also notice that the domains are sorted according to a non-increasing traffic load.

As can be seen, the traffic load and delete curves exhibit a similar trend, while the trend of the read and spam curves is also similar, but different than the former one. This is somewhat surprising since one may expect that a domain that sends more emails to users would get more reads, deletes and spam votes from the users. Perhaps the reason for this discrepancy resides in the category of the domain and its quality. For example, `facebookmail` has higher read ratio than `groupon` and `livingsocial` since its messages are social notifications while the others are of deals and coupons. We believe that this requires further investigation. Having noted that, a more important observation is that all the curves, including the unsubscription clicks, show a decreasing trend in general. This implies that the average load as well as the user actions are positively

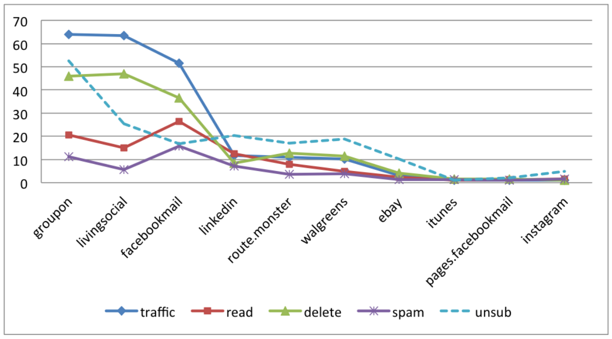


Figure 8: Normalized average traffic load and number of read, delete, spam, and unsubscribe actions. Note that we removed prefixes and suffixes from the domain names.

correlated with unsubscription clicks. Thus, they can be valuable features in predicting the tendency of a user to unsubscribe.

5 UNSUBSCRIPTION RECOMMENDATIONS

We have demonstrated through data analysis that unsubscribable mail is a major contributor to information overload in email across the general user population. Furthermore, we observed that despite the noticeable load that those emails induce, users are not inclined to actively unsubscribe from them. We conjecture that the main reason for this gap is the lack of a simple and efficient mechanism for unsubscription. To bridge this gap, we believe that one should create a native unsubscription mechanism within the mail service.

In this section, we develop a machine learning recommendation approach (backend) that can support such a native unsubscription feature in mail (frontend). This mail feature is presented in the next section. Note that the data-based insights identified in the previous section guide the development of many of the features used by our recommendation approach.

5.1 Problem and data

We formalize the unsubscription recommendation task as a ranking problem. This formalization is motivated by our mail feature from Section 6. This feature presents several unsubscription recommendations to a user. As a consequence, our objective is to provide an ordered list of all the subscription entities of a user according to the propensity of the user to unsubscribe from each of them. Notice that this task does not require to accurately predict the probability of unsubscription for each entity, but rather their relative order.

The data that is used to build and evaluate our models consist of users' inbound mail traffic, their native mail actions, and their clicks on unsubscription links within email bodies. Note that the unsubscription clicks are also the events our models need to predict, i.e., the supervision signal. Our data is collected from millions of users over approximately a month. As mail is a temporally ordered collection, we split the data into train and test by time, and not randomly. Specifically, we use the data of the first three weeks for training and tuning the models, and the remaining week for testing. This is common practice (see, e.g., [41]) to prevent an unrealistic setting in which future information is used to predict past user behavior. We like to emphasize that our test set consists of more than 50K users that clicked on unsubscription links and had more than 50 subscription entities on average. As a result, the prediction task is not trivial.

5.2 Learning and evaluation

We experiment with few machine learning algorithms. Nevertheless, since the focus of our work is on the identification and selection of features that improve the prediction, and not on the optimization of machine learning techniques, we decide to report only the experimental results of a logistic regression (LR) method. This method first predicts the probability that each user would like to unsubscribe from any of her subscription entities. Then, all the entities of each user are sorted in a non-increasing order according to the predictions. We compare this approach with a naïve random ranking method, and a manually-defined baseline heuristic that is inspired by our data analysis. Note that due to the online nature by which our data is generated and its extremely large size, we utilized the highly-scalable Vowpal Wabbit implementation [28, 29], which works well in conjunction with MapReduce architecture.

Features. Our analysis in Section 4 identified several factors that are likely to help predicting unsubscription probability. We believe it is beyond the scope of this paper to provide an exhaustive list of all the features utilized by our models as this number is in the thousands. We describe our general hierarchical design approach and identify some of the more valuable features. Those are also the features that were utilized by our strong baseline heuristic.

At the most intuitive level, we develop a *hierarchical features construction*. Namely, for each subscription entity, we collect features at several user group levels. At the lowest level of this hierarchy, we collect features that are personal for each user. These features capture properties of the traffic between the entity under consideration and the user, and the user behavior on that traffic. For example, we maintain the number of emails that were sent to the user by the entity, the implied daily load of this traffic on the user, the number of times that the user read, deleted, or performed any other action on those emails, the ratio of emails that were deleted-without-read, and more. At the highest level of the hierarchy, we collect similar aggregative features with respect to the entire user population. These features are essentially ones that capture the global properties of the subscription entity. For instance, we maintain the relative ratio of users that unsubscribed from the entity, the portion of emails that were marked spam, and more. In a similar way, we collect features at additional mid-level user groups such as demographic and activity groups. For the purpose of prediction, the features that represent the interaction between the user and subscription entity are a concatenation of all the features of the groups that include the user. For instance, if a user is a 28 years old female who visits the mail service every day then her feature vector will be the concatenation of her personal features with respect to the underlying entity, the entity features associated with the female group, the 25-34 age group, the Ultra-activity group, and finally, the global features of the entity. This approach allows us to capture the intricacies related to the different subscription, management and unsubscription patterns of different users and user groups.

The features that we collect at each level are essentially the same. For example, the feature that accounts for the relative ratio of all users that unsubscribed from an entity (at the highest level of the hierarchy) reduces to a binary feature indicating whether a user unsubscribed from an entity in the past (at the lowest level of the hierarchy). As noted before, we collect features relating to the traffic between the entity and the users, the behavior of the users on that traffic, and additional complex combinations of those features. Examples of traffic-related features are the number of emails that the entity sent to the user group, temporal features such as the days and times that those emails were sent, and content features such as statistics relating to the length of the subject line and content of the emails. Examples of behavioral-related features are the number of various actions that users in the group performed over emails of the entity, along with their temporal aspect. Examples of complex features include the ratio of emails deleted-without-read for the user group, and the ratio of users that already unsubscribed from the entity. Additional features were obtained in similar fashion to recent papers [6, 18, 41]. Overall, our hierarchical approach had several thousands of features.

Baselines. We compare our approach against two baselines:
(1) Random – This method randomly generates a ranking of all the unsubscription entities for each user.
(2) Data-driven – This manually-defined method makes a decision based on few features that have been identified by our data analysis as relatively highly-correlative with unsubscriptions. Specifically, for each user, the method first sorts the entities into the following ordered classes:

- Entities from which the user unsubscribed (or tried to) in the past.
- Entities with mail messages that the user marked as spam.
- Entities for which at least half of the messages were deleted by the user without being read.
- Entities for which any other number of messages were deleted by the user without being read.
- All remaining entities.

Then, the order inside each of those classes is defined by the global unsubscription ratio of the entities.

Experimental results. We analyze the results of our approach compared with the two baseline methods. As we are treating this problem as a ranking problem, we consider standard information retrieval ranking metrics, namely, Mean Average Precision (MAP), Mean Reciprocal Rank (MRR), Area under the ROC curve (AUC), and Normalized Discounted Cumulative Gain (NDCG). The results are presented in the following table.

Method	MAP	MRR	AUC	NDCG
Random	0.1208	0.1264	0.4958	0.2842
Data-driven	0.3562	0.3690	0.8354	0.4985
LR	**0.4426**	**0.4612**	**0.8999**	**0.5731**

As can be seen, our logistic regression-based approach outperforms both baseline methods. For example, it improves over the MRR and the AUC of the data-driven baseline by almost 25% and 8%, respectively. This is very encouraging as we applied a relatively simple learning machinery. It is also important to notice that the data-driven baseline is surprisingly strong, improving greatly over the random method. This proves the effectiveness of our data analysis and feature selection. It also reaffirms the strength of our machine learning solution, which significantly outperformed over this strong baseline.

6 ONLINE EXPERIMENT

We discuss our online experiment for unsubscription recommendations and some of its user engagement metrics. The experiment was conducted on Yahoo Web mail service over a period of one month and involved few millions of users. Our primarily focus is on the experimental setting and results. We do not discuss its internals, and in particular, the backend recommendation service. This service was based on the insights and identification methods that were presented in the previous sections with some light modifications required for matters of productization.

6.1 The setting

The online experiment ran over a period of one month from March 28, 2017 to April 28, 2017, and included few millions of users. In the experiment, users were presented with a pop-up dialog that consisted of at most 5 unsubscription recommendations from their collection of subscriptions. Figure 9 exhibits this dialog. The user could then select to unsubscribe from any number of those recommendations. For each selected recommendation, the user was prompted for approval, before additional steps for final unsubscription were made by the system. The system then sent an unsubscription email on the behalf of the user (*email-based method*) or transfered her to the designated unsubscription page of the underlying entity to complete the action (*page-based method*). The decision of which unsubscription method to use was conditioned by whether email-based unsubscription was supported or not. The dialog was presented to a user at most once a day, and was triggered only if the user performed a *batch-delete* action, which consists of deleting several emails at the same time. The batch-delete action has been identified by internal user studies as an action that indicates that the user is more inclined towards cleaning and managing her inbox.

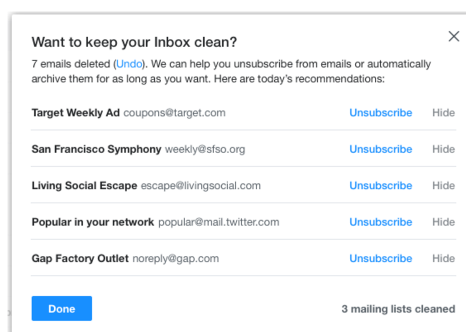


Figure 9: The unsubscription dialog.

The unsubscription recommendations were powered by a personalized recommendations service which was updated every day with respect to new inbound traffic and user actions. The service is based on the insights and techniques that were presented in the previous sections. As a general rule, top recommendations were presented, that is, the subscriptions that the service predicted as the ones that the user most likely wants to stop receiving. There were two practical adjustments applied to this general rule. First, subscriptions whose emails were part of the batch-delete action got some boost to their score. This adjustment most commonly led to the presentation of those subscriptions in the dialog. Second, subscriptions that were already presented to the user in the past got some reduction in their score. This adjustment effectively enabled rotation among recommendations, gradually showing also lower-ranked recommendations (instead of higher ranked recommendations that were not clicked by the user).

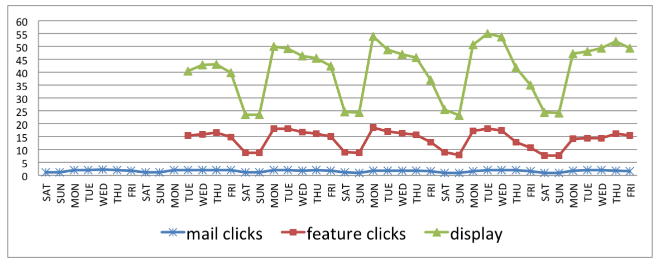
6.2 Online engagement metrics

The primary objective of the experimental feature was to mitigate the process of unsubscribing and reduce overload in email. As a result, we focus our attention on user engagement with respect to the feature. Note that we concentrated on one experimental setting and did not evaluate the influence that different components, such as the backend recommendations service or the frontend interface, have on the engagement metrics. Recall that a user had to go through two steps in the experiment. First, the user had to perform a batch-delete for the unsubscription dialog to be displayed, and second, the user had to select and approve unsubscription recommendations.

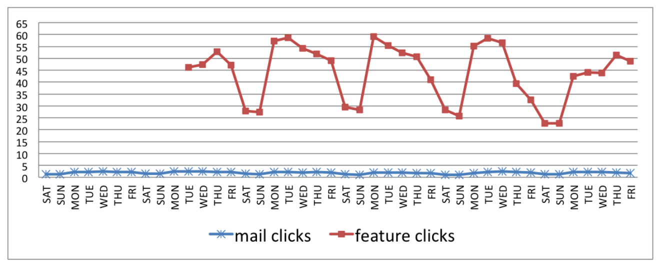
The percentage of users that performed batch-delete at least once during the time period of the experiment was 31.8%. This implies that this action was popular enough, allowing a significant portion of the users to view the unsubscription dialog. Note that batch-delete is a native mail action that also lives outside the confines of our experiment, thus, the presented statistics imply also general measurements of that action, independently of the experiment.

Considering only the users that viewed the dialog, we observe that 49.3% of those users utilized the dialog to unsubscribe at least once during the experiment, and that 34.2% of all dialogs were engaged by users. These engagement metrics are surprisingly high. They suggest few important things. First and foremost, many users have subscriptions they consider irrelevant, but so far, did not find a simple way to manage them. This is validated by the fact that nearly half of the users had at least one subscription they wanted to remove. Note that this statistics also shows that the feature along with the suggested recommendations met their management needs to unsubscribe. In addition, the fact that more than one third of the dialogs led users to unsubscribe emphasizes the usability of this feature and the quality of recommendations. These results disclose the key problem with traditional unsubscription mechanisms which users do not seem to adopt.

To better understand the usability of our new unsubscription feature, we compare unsubscriptions that were done through the feature and those done using the standard approach of clicking unsubscription links within the message body. Figure 10 presents a comparison of the unsubscription behavior of the user population participating in the experiment. The comparison encompasses the experiment time period as well as a period of ten days before its start. As can be seen in Figure 10(a), the number of unique users that utilize the feature is greater by roughly 8x compared to the number of unique users that use the standard approach. When considering the total number of unsubscriptions, as shown in Figure 10(b), the gap is even greater. The number of unsubscriptions performed using the feature is greater by about 23x the number of unsubscriptions by the standard way. These two observations further emphasize the usefulness of the feature. On one hand, it increases the number of users that perform unsubscription by about 8x, while on the other hand, each user now performs about 3x more unsubscriptions on average. This dual-effect is exactly what one can hope for.



(a)



(b)

Figure 10: (a) The number of unique users that unsubscribed using the feature and the standard way, and (b) overall number of unsubscriptions done using the feature and the standard way. Values are normalized by the respective minimum value of standard unsubscription approach.

Two additional observations are worth noting. First, the number of unsubscriptions and the general trend relating to the standard approach both before and during the experiment do not change. This consistency implies that the unsubscription feature can be considered as being complimentary to the standard unsubscription mechanism. Second, the unsubscription numbers and trends seen with respect to the feature approach also stay roughly the same during the time period considered, except some slight decrease during the last week of the experiment. This consistency is encouraging, and is yet another validation of the feature usefulness.

7 CONCLUSIONS

We studied unsubscribable mail, which comprises about 95% of all machine-generated mail traffic, and as such, is the main source of information overload in email. We performed a large-scale analysis of this type of messages, based on data collected from millions of users of Yahoo Web mail service. The analysis reveals a striking gap between the low interest users are showing towards unsubscribable mail and their inactivity in decreasing the load implied by this traffic. Motivated by this gap, we developed an unsubscription recommendation service that was put into operation in an online large-scale experiment. This experiment offered users personalized unsubscription recommendations within a designated native mail feature. This feature provided the users with an efficient and easily accessible mechanism for unsubscribing from superfluous mailing lists. Our experiment resulted in an increase of 8x in the number of users that utilized the new feature, compared to the number of users that performed unsubscriptions using the standard way of clicking unsubscription links within email bodies. On the same time, the overall number of unsubscriptions that were performed using the new feature was greater by about 23x compared to that of the standard unsubscription way. These results demonstrates the imminent need for such a mechanism, which could help bridge the aforementioned gap and ease overload in email.

REFERENCES

[1] D. Aberdeen, O. Pacovsky, and A. Slater. The learning behind gmail priority inbox. In *NIPS 2010 Workshop on Learning on Cores, Clusters and Clouds*. Elsevier, 2010.

[2] C.-S. Act. Federal trade commission, 2003. https://www.ftc.gov/tips-advice/business-center/guidance/can-spam-act-compliance-guide-business.

[3] Q. Ai, S. Dumais, N. Craswell, and D. Liebling. Characterizing email search using large-scale behavioral logs and surveys. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1511–1520. ACM, 2017.

[4] N. Ailon, Z. S. Karnin, E. Liberty, and Y. Maarek. Threading machine generated email. In *Proceedings of the sixth ACM international conference on Web search and data mining*, WSDM '13, pages 405–414. ACM, 2013.

[5] I. Alberts and D. Forest. Email pragmatics and automatic classification: A study in the organizational context. In *Journal of the Association for Information Science and Technology*, volume 63, pages 904–922. John Wiley & Sons, Inc., 2012.

[6] N. Avigdor-Elgrabli, M. Cwalinski, D. D. Castro, I. Gamzu, I. Grabovitch-Zuyev, L. Lewin-Eytan, and Y. Maarek. Structural clustering of machine-generated mail. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management*, pages 217–226, 2016.

[7] V. Bellotti, N. Ducheneaut, M. Howard, and I. Smith. Taking email to task: The design and evaluation of a task management centered email tool. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 345–352. ACM, 2003.

[8] M. Bendersky, X. Wang, D. Metzler, and M. Najork. Learning from user interactions in personal search via attribute parameterization. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, WSDM '17, pages 791–799. ACM, 2017.

[9] D. Carmel, G. Halawi, L. Lewin-Eytan, Y. Maarek, and A. Raviv. Rank by time or by relevance?: Revisiting email search. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, CIKM '15, pages 283–292. ACM, 2015.

[10] D. Carmel, L. Lewin-Eytan, A. Libov, Y. Maarek, and A. Raviv. The demographics of mail search and their application to query suggestion. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17. ACM, 2017.

[11] D. Carmel, L. Lewin-Eytan, A. Libov, Y. Maarek, and A. Raviv. Promoting relevant results in time-ranked mail search. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17. ACM, 2017.

[12] M. E. Cecchinato, A. Sellen, M. Shokouhi, and G. Smyth. Finding email in a multi-account, multi-device world. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 1200–1210, 2016.

[13] L. A. Dabbish and R. E. Kraut. Email overload at work: An analysis of factors associated with email strain. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, CSCW '06, pages 431–440. ACM, 2016.

[14] L. A. Dabbish, R. E. Kraut, S. Fussell, and S. Kiesler. Understanding email use: Predicting action on a message. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 691–700. ACM, 2005.

[15] D. Di Castro, Z. Karnin, L. Lewin-Eytan, and Y. Maarek. You've got mail, and here is what you could do with it!: Analyzing and predicting actions on email messages. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, WSDM '16, pages 307–316. ACM, 2016.

[16] D. Di Castro, L. Lewin-Eytan, Y. Maarek, R. Wolff, and E. Zohar. Enforcing k-anonymity in web mail auditing. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, WSDM '16, pages 327–336. ACM, 2016.

[17] M. Dredze, T. Brooks, J. Carroll, J. Magarick, J. Blitzer, and F. Pereira. Intelligent email: Reply and attachment prediction. In *Proceedings of the 13th International Conference on Intelligent User Interfaces*, IUI '08, pages 321–324, Canary Islands, Spain, 2008.

[18] M. Grbovic, G. Halawi, Z. Karnin, and Y. Maarek. How many folders do you really need? classifying email into a handful of categories. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, pages 869–878. ACM, 2014.

[19] C. Grevet, D. Choi, D. Kumar, and E. Gilbert. Overload is overloaded: Email in the age of gmail. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 793–802. ACM, 2016.

[20] J. Gwizdka. Email task management styles: The cleaners and the keepers. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '04, pages 1235–1238. ACM, 2004.

[21] R. . Internet Engineering Task Force. The use of urls as meta-syntax for core mail list commands and their transport through message header fields, July 1998. https://tools.ietf.org/html/rfc2369.

[22] R. . Internet Engineering Task Force. List-id: A structured field and namespace for the identification of mailing lists, March 2001. https://tools.ietf.org/html/rfc2919.

[23] R. . Internet Engineering Task Force. Signaling one-click functionality for list email headers, January 2017. https://tools.ietf.org/html/rfc8058.

[24] S. Kiritchenko and S. Matwin. Email classification with co-training. In *Center for Advanced Studies on Collaborative Research*, CASCON '11, pages 301–312, 2011.

[25] B. Klimt and Y. Yang. The enron corpus: A new dataset for email classification research. In *Machine Learning: ECML 2004: 15th European Conference on Machine Learning, Proceedings*, pages 217–226, 2004.

[26] F. Kooti, L. M. Aiello, M. Grbovic, K. Lerman, and A. Mantrach. Evolution of conversations in the age of email overload. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, pages 603–613, 2015.

[27] G. Lab. Why do people subscribe to email newsletters?, Mars 2015. https://lab.getapp.com/new-research-getdata-why-do-people-subscribe-to-email-newsletters.

[28] J. Langford, L. Li, and A. Strehl. Vowpal wabbit online learning project, 2007. Technical report, http://hunch.net.

[29] J. Langford, L. Li, and T. Zhang. Sparse online learning via truncated gradient. *Journal of Machine Learning Research*, 10:777–801, 2009.

[30] K. Narang, S. T. Dumais, N. Craswell, D. Liebling, and Q. Ai. Large-scale analysis of email search and organizational strategies. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, CHIIR '17, pages 215–223. ACM, 2017.

[31] B. On, E. Lim, J. Jiang, A. Purandare, and L. Teow. Mining interaction behaviors for email reply order prediction. In *Proceedings of the 2010 International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '10, pages 306–310. IEEE, 2010.

[32] A. Qadir, M. Gamon, P. Pantel, and A. H. Awadallah. Activity modeling in email. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL-HLT '16, pages 1452–1462, 2016.

[33] S. Radicati and J. Levenstein. Email market, 2015-2019, Nov 2015. http://www.radicati.com/wp/wp-content/uploads/2015/07/Email-Market-2015-2019-Executive-Summary.pdf.

[34] S. Radicati and J. Levenstein. Email market, 2017-2021, Nov 2017. http://www.radicati.com/wp/wp-content/uploads/2017/01/Email-Statistics-Report-2017-2021-Executive-Summary.pdf.

[35] M. Sappelli, S. Verberne, and W. Kraaij. Combining textual and non-textual features for e-mail importance estimation. In *Proceedings of the 25th Benelux Conference on Articial Intelligence*, pages 168–174. IBM Press, 2013.

[36] A. Taiwo, Z. Shikun, and K. Rinat. Applying machine learning techniques for e-mail management: solution with intelligent e-mail reply prediction. *Journal of Engineering and Technology Research*, 1(7):143–151, 10 2009.

[37] B. Wang, M. Ester, J. Bu, Y. Zhu, Z. Guan, and D. Cai. Which to view: Personalized prioritization for broadcast emails. In *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, pages 1181–1190, 2016.

[38] B. Wang, M. Ester, Y. Liao, J. Bu, Y. Zhu, Z. Guan, and D. Cai. The million domain challenge: Broadcast email prioritization by cross-domain recommendation. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 1895–1904. ACM, 2016.

[39] J. B. Wendt, M. Bendersky, L. Garcia-Pueyo, V. Josifovski, B. Miklos, and I. Krka. Hierarchical label propagation and discovery for machine generated email. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, WSDM '16, 2016.

[40] S. Whittaker and C. L. Sidner. Email overload: Exploring personal information management of email. In *Conference on Human Factors in Computing Systems: Common Ground, CHI*, pages 276–283, 1996.

[41] L. Yang, S. T. Dumais, P. N. Bennett, and A. H. Awadallah. Characterizing and predicting enterprise email reply behavior. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '17, pages 235–244. ACM, 2017.