# Activity Modeling in Email

**Ashequl Qadir**
School of Computing
University of Utah
Salt Lake City, UT 84112, USA
asheq@cs.utah.edu

**Michael Gamon**
Microsoft Research
One Microsoft Way
Redmond, WA 98052, USA
mgamon@microsoft.com

**Patrick Pantel**
Microsoft Research
One Microsoft Way
Redmond, WA 98052, USA
ppantel@microsoft.com

**Ahmed Hassan Awadallah**
Microsoft Research
One Microsoft Way
Redmond, WA 98052, USA
ahmed.awadallah@microsoft.com

## Abstract

We introduce a latent activity model for workplace emails, positing that communication at work is purposeful and organized by activities. We pose the problem as probabilistic inference in graphical models that jointly capture the interplay between latent activities and the email contexts they govern, such as the recipients, subject and body. The model parameters are learned using maximum likelihood estimation with an expectation maximization algorithm. We present three variants of the model that incorporate the recipients, co-occurrence of the recipients, and email body and subject. We demonstrate the model's effectiveness in an email recipient recommendation task and show that it outperforms a state-of-the-art generative model. Additionally, we show that the activity model can be used to identify email senders who engage in similar activities, resulting in further improvements in recipient recommendation.

## 1 Introduction

Activities are a prominent characteristic of a workplace, typically governed by people's job roles and work responsibilities. Examples of workplace activities can include organizing a conference, purchasing equipment, managing candidate interviews, etc. Activities can be viewed as a collaborative work practice involving a set of people each playing a different role in the activity (Dredze et al., 2006). Although emails are an integral part of workplace communication, current email clients offer little support for the activity oriented use of email (Khoussainov and Kushmerick, 2005). Discussions can get split across long email threads and communications about all activities can get intermixed, making activity management difficult (Balakrishnan et al., 2010).

In this work, we model activities as latent probability distributions personalized to the email sender. We present three variants of the activity model, incorporating: (1) email recipients, (2) email recipient pairs which account for co-occurrence of the email recipients, and (3) email body and subject tokens along with email recipient pairs. Additionally, we experiment with lexical (bag of words), syntactic (nouns and verb phrases), and semantic (things of interest in an email) representations of the body and subject tokens of an email. The parameters of the generative model are learned using an expectation maximization (EM) algorithm.

For evaluation, we formulate a real world task setting for email recipient recommendation, where we assume that all but the last recipient of an email has been entered by the sender, and we test the effectiveness of our activity model in recommending the last recipient. Such a system has practical applications, such as reminding an email sender about a potentially forgotten recipient or recommending the next recipient as the sender enters each recipient.

The main contributions of our research are:

- We introduce a latent activity model for emails



*Proceedings of NAACL-HLT 2016*, pages 1452–1462,
San Diego, California, June 12-17, 2016. 

where the email contexts are governed by workplace activities;

- We present probabilistic modeling that incorporates co-occurring recipients with lexical, syntactic and semantic contexts of an email;
- We identify senders engaging in similar activities using the activity model, and show improvements in recipient recommendation.

## 2 Related Work

Prior research related to our work can be divided into the following three major areas presented below.

### 2.1 Activity in Emails

Prior research has treated emails as a communication tool for workplace activities (Moran, 2005) or a task management resource (Bellotti et al., 2003). Kushmerick and Lau (2005) formalized e-commerce activities as finite-state automata, where transitions among states represent messages sent between participants. Dredze et al. (2006) used user generated activity labels and classified emails into activities using overlapping participants and content similarity. Minkov et al. (2008) modeled user created folders and TO-DO items as activities, and created a heterogeneous graph to perform activity-centric search.

Shen et al. (2006) predicted tasks associated with an incoming email by leveraging email sender, recipients and distinct subject words. They found the body words to not provide additional prediction value. Although they used similar information as we do, they used a combination of generative and discriminative models toward task classification, and did not do recipient recommendation. Our activity model designs are closer to the model introduced by Dredze and Wallach (2008), who presented a Dirichlet process mixture model combined with author and thread information. Our designs differ as we use co-occurring recipients in the generative process, and use various linguistic representations of content.

### 2.2 Generative models for Emails

Latent Dirichlet Allocation (LDA) (Blei et al., 2003) is a frequently used generative topic model. Assuming a Dirichlet prior, LDA models learn probability distributions of words as latent topics in a corpus. In emails, LDA models have been used for learning summary keywords (Dredze et al., 2008), analyzing how topics change over time (Wang and McCallum, 2006), understanding entity relations (Balasubramanyan and Cohen, 2011), analyzing communication networks (Nguyen et al., 2013), for authorship attribution (Seroussi et al., 2012), and discovering topics associated with authors (McCallum et al., 2005).

Other generative models have also been used for analyzing email communication behavior (Navaroli et al., 2012), identifying links between an email sender and a recipient to detect potential anomalous communication (Huang and Zeng, 2006), and resolving personal names in emails (Elsayed et al., 2008). Representing workplace activities of the emails with probabilistic inference in graphical models where observed information is personalized to the email senders is what sets our work apart from previous research in computational models for emails.

### 2.3 Email Recipient Recommendation

For recommending email recipients, interactions among email participants and content similarity are the signals that have been explored most. Carvalho and Cohen (2007) leveraged content similarity by creating tf-idf centroid vectors and determining k-nearest neighbors of a target email. Pal et al. (2007) presented a discriminative author recipient topic model that uses transfer learning. Desai and Dash (2014) used reverse chronologically arranged implicit groups determined from sent emails. Sofershtein and Cohen (2015) created a ranking function combining temporal and textual features.

Among the generative modeling based approaches, Pal and McCallum (2006) learned probability distributions of recipients, and words in body and subject to predict recipients in email cc lists. Dredze et al. (2008) evaluated the impact of summary keywords generated using LDA, for email recipient prediction. More recently, Graus et al. (2014) predicted email recipients by estimating sender and recipient likelihood using a communication graph, email likelihood using content words, and evaluated performance on the Avocado email corpus. In our work, we use latent activity distributions, and identify senders who engage in similar activities. We compare our recipient prediction results against the

generative model of Graus et al. (2014).

## 3 Problem Setting and Data

### 3.1 Activity Modeling in Emails

Our motivation for activity modeling in email stems from the assumption that in the workplace, people primarily use emails as a communication tool for their ongoing activities, and an email's recipient list, content, and other context are governed by a given activity. For example, an employee attending a conference may write emails to the conference organizers regarding registration or scheduling, or emails to a hotel for booking confirmation. The communication may span multiple emails, involving many parties, but all under the same activity.

We model the activities as a latent probabilistic variable over the email recipients and content, personalized to the email sender. Let $D$ be the set of all emails in a corpus containing $N$ emails, generated by $S = \{s_i \mid 1 \leq i \leq S_D\}$ senders, and sent to $R = \{r_i \mid 1 \leq i \leq R_D\}$ recipients. Let $B = \{b_i \mid 1 \leq i \leq B_D\}$, and $T = \{t_i \mid 1 \leq i \leq T_D\}$ represent the body and subject vocabulary of the emails respectively. Let $K$ be the number of latent activities for each sender. We model the activities as probability distributions over email components $S$, $R$, $B$ and $T$.

### 3.2 Corpus Description and Data Sets

For our experiments, we use the Avocado email corpus, available from the LDC catalog.[1] The corpus contains emails from a defunct IT company referred to as "Avocado". For learning the activity model, we extract emails from 7/1/2000 – 5/1/2001 to create a training data set, and from 5/1/2001 – 6/30/2001 for development/tuning. In this work, we did not consider threaded conversations, only retained the first email in a thread and discarded the rest.

As additional filtering steps, we only kept emails written by the Avocado employees, allowing us to confine the scope of the activities within the company. To control sparsity and noise, for each email, we enforced a a minimum of two recipients, and a maximum of ten recipients.[2] In a practical system, we consider it reasonable that a model only activates after some history of email is sent by a user. We therefore removed emails by the senders having fewer than 25 emails in the training data. From email bodies and subjects, we removed stopwords, and words appearing fewer than 5 times or more than 100 times.[3] When a recipient's same email alias was present multiple times, we took it only once, as well as removed a sender's email alias from the recipients list if it was present there. In this work, we focused on recipients from only the "TO" field, and did not include recipients from the "CC" or "BCC" field. Table 1 presents statistics of the data sets.

| Number of | Train | Train + Dev |
|---|---|---|
| Total emails | 18,593 | 22,283 |
| Unique senders | 120 | 129 |
| Unique recipients | 3,157 | 3,658 |
| Unique body words | 31,386 | 36,005 |
| Unique subject words | 6,969 | 7,724 |
| Emails per sender | 154.94 | 172.74 |
| Emails per recipient | 5.89 | 6.09 |

**Table 1:** Data set statistics.

## 4 Activity Models

Our key assumption in modeling the activities in email is that different components of an email contain specific types of information that can help to characterize the activities that drive user behavior. In our generative process of the activity model, for an email $d \in D$, a sender $s \in S$ is first generated from a multinomial distribution with probability vector $\sigma$, then an activity $a$ is generated from a sender personalized multinomial distribution with probability vector $\theta_s$. Let $R_d \subseteq R$, $B_d \subseteq B$ and $T_d \subseteq T$ be the set[4] of recipients, body and subject tokens of $d$ respectively. The generation of the email contexts (recipients and body/subject tokens) varies based on the specific design of each variant of our model. In a first simplistic model, we assume that recipient $r \in R_d$, body token $b \in B_d$ and subject token $t \in T_d$ for an email can be generated from the multinomial distributions with probability vectors $\lambda_{s,a}$, $\phi_{s,a}$, and $\tau_{s,a}$ respectively, that are conditioned $s$ and $a$. Point estimates for $\sigma$ can be directly

[1] https://catalog.ldc.upenn.edu/LDC2015T03

[2] When an email has many recipients, it is often indicative of general announcements or system generated emails, which are less directly relevant to an employee's activities. Some emails in the Avocado dataset have more than 500 recipients.

[3] A heuristic we used to remove words that are too general.

[4] Multiset for body and subject tokens.

calculated from a training corpus, whereas $\theta$, $\lambda$, $\phi$, and $\tau$ are the unknown parameters of the model.

### 4.1 Model 1: Rec

In our first model $Rec$, we assume that the latent activities can be learned as a probability distribution over the recipients alone. The generative process is:

**Model 1: Rec**
For each email document $d \in D$
  sender $s$ ~ Multinomial($\sigma$)
  activity $a$ ~ Multinomial($\theta_s$)
  For $1 \ldots |R_d|$
    recipient $r$ ~ Multinomial($\lambda_{s,a}$)

Figure 1 presents the plate diagram of the model. The joint probability of the $Rec$ model is the product of the conditional distributions:

$$P(s, a, r|\sigma, \theta, \lambda) = P(s|\sigma)P(a|s, \theta) \prod_{r \in R_d} P(r|s, a, \lambda)$$

The probability of a sender $s$, an activity $a$ given $s$, and a recipient $r$ given $s$ and $a$ are defined below[5]:

$$P(s = \hat{s}) = \prod_{i=1}^{S_D} \sigma_i^{I[i=\hat{s}]},\ s.t. \sum_i \sigma_i = 1$$

$$P(a = \hat{a}|s = \hat{s}) = \prod_{i=1}^{K} \theta_{\hat{s},i}^{I[i=\hat{a}]},\ s.t.\ \forall_s \sum_i \theta_{s,i} = 1$$

$$P(r = \hat{r}|s = \hat{s}, a = \hat{a}) = \prod_{i=1}^{R_D} \lambda_{\hat{s},\hat{a},i}^{I[i=\hat{r}]},$$

$$s.t. \quad \forall_{s,a} \quad \sum_i \lambda_{s,a,i} = 1$$

**Inference:** Let $d^n$ be the $n^{th}$ email, where $d^n = \{s^n, R_d^n\}$. We apply Bayes' rule to find the posterior distribution over the activities $P^n(a|d)$, which is directly proportional to the joint distribution $P^n(a, d)$. We can exactly compute this distribution by evaluating the joint distribution for every value of $a$ and the observed document $d^n$.

**Learning:** Point estimates for $\sigma$ can be directly obtained from the training corpus. We estimate the parameters $\theta$ and $\lambda$ by maximizing the (log) probability of observing $D$. We write the $\log(D)$ as:

$$\log P(D) = \sum_{n=1}^{N} \sum_a P^n(a|s, R_d) \log P^n(a, s, R_d)$$

We use the Expectation-Maximization (EM) algorithm to set the parameters. Starting with a random initialization of the parameters (with Gaussian noise), EM iterates between the E-step in which $P^n(a|s, R_d)$ is computed for each email with fixed parameter values computed in the previous M-step, and the M-step in which the parameters are updated with fixed $P^n(a|s, R_d)$ values computed in the E-step. The parameter updates are obtained by taking the derivative of $\log P(D)$ with respect to each parameter, and setting the resultant to 0, providing us with the following parameter updates:

$$\theta_{s^n,i} = \frac{\sum_{n=1}^{N} \sum_a P^n(a|d) I[i = a]}{\sum_{n=1}^{N} \sum_a P^n(a|d)}$$

$$\lambda_{s^n,a,i} = \frac{\sum_{n=1}^{N} \sum_a P^n(a|d) \sum_{r \in R} I[i = r]}{|R_d^n| \sum_{n=1}^{N} \sum_a P^n(a|d)}$$

We run EM until the change in $\log P(D)$ is less than our convergence threshold $10^{-5}$.

### 4.2 Model 2: CoRec

Using co-occurring recipients in generative models for emails has been rarely explored in previous work. Pal and McCallum (2006) modeled co-recipient information as a probability distribution of recipients conditioned on the other recipients, and noted that this information improved their email cc prediction performance. In our $CoRec$ model, we model co-recipients as pairs of recipients generated from a probability distribution conditioned on the sender and the activity. Let $L = \{(r_i, r_j) \mid 1 \leq i \leq R_D, 1 \leq j \leq R_D\}$ having $L_D$ pairs of recipients in the corpus. For an email $d$, $L_d \subseteq L$ is the set of recipient pairs in $d$. The $CoRec$ model first generates a sender $s$ from the probability distribution $\sigma$, then an activity $a$ from a distribution over latent activities $\theta_s$ personalized to $s$, and finally recipient pairs $r_p \in L_d$ from a distribution over recipient pairs $\omega_{s,a}$ conditioned on $s$ and $a$. The generative process is summarized below:

[5] $I$ is an indicator variable

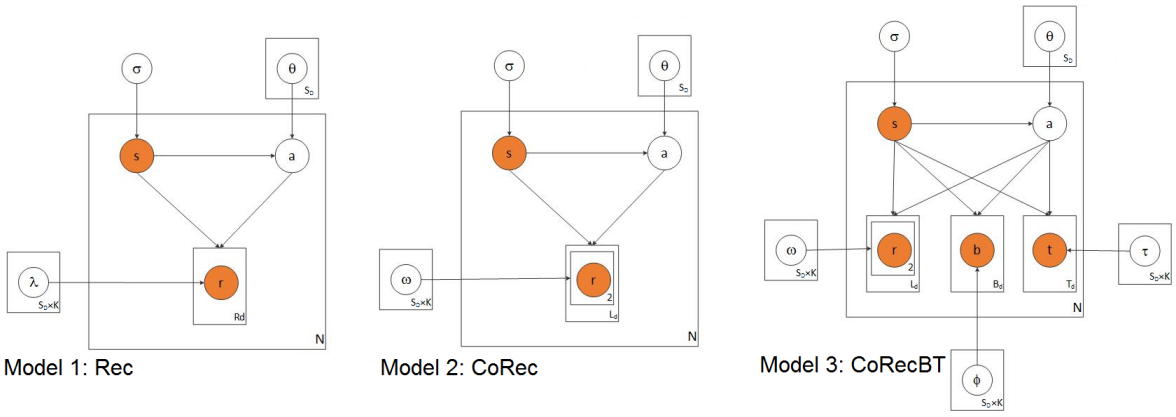


**Figure 1:** Activity model plate diagrams.

**Model 2: CoRec**
For each email document $d$
  sender $s$ ~ Multinomial($\sigma$)
  activity $a$ ~ Multinomial($\theta_s$)
  For $1 \ldots |L_d|$
    recipient pair $r_p$ ~ Multinomial($\omega_{s,a}$)

The joint probability of the $CoRec$ model is:

$$P(s, a, r_p|\sigma, \theta, \omega) = P(s|\sigma)P(a|s, \theta) \prod_{r_p \in L_d} P(r_p|s, a, \omega)$$

This adds over the $Rec$ model the probability of a recipient pair $r_p$ given $s$ and $a$, defined below:

$$P(r_p = \hat{r}_p|s = \hat{s}, a = \hat{a}) = \prod_{i=1}^{L_D} \omega_{\hat{s},\hat{a},i}^{I[i=\hat{r}_p]},$$
$$s.t. \quad \forall_{s,a} \quad \sum_i \omega_{s,a,i} = 1$$

The EM algorithm is applied in the same way as in the $Rec$ model. During the M-step, update for $\theta$ remains the same. The update for $\omega$ is given below:

$$\omega_{s^n,a,i} = \frac{\sum_{n=1}^{N} \sum_a P^n(a|d) \sum_{r_p \in L} I[i = r_p]}{|L_d^n| \sum_{n=1}^{N} \sum_a P^n(a|d)}$$

### 4.3 Model 3: CoRecBT

Finally, in the $CoRecBT$ model, we further incorporate body and subject of emails. The generative process of the model is:

**Model 3: CoRecBT**
For each email document $d$
  sender $s$ ~ Multinomial($\sigma$)
  activity $a$ ~ Multinomial($\theta_s$)
  For $1 \ldots |L_d|$
    recipient pair $r_p$ ~ Multinomial($\omega_{s,a}$)
  For $1 \ldots |B_d|$
    body token $b$ ~ Multinomial($\phi_{s,a}$)
  For $1 \ldots |T_d|$
    subject token $t$ ~ Multinomial($\tau_{s,a}$)

The joint probability of the $CoRecBT$ model:

$$P(s, a, r_p, b, t|\sigma, \theta, \omega, \phi, \tau) = P(s|\sigma)P(a|s, \theta) \prod_{r_p \in L_d} P(r_p|s, a, \omega) \prod_{b \in B_d} P(b|s, a, \phi) \prod_{t \in T_d} P(t|s, a, \tau)$$

where the probability of a body token $b$ and subject token $t$ given $s$ and $a$ defined as:

$$P(b = \hat{b}|s = \hat{s}, a = \hat{a}) = \prod_{i=1}^{B} \phi_{\hat{s},\hat{a},i}^{I[i=\hat{b}]}$$
$$P(t = \hat{t}|s = \hat{s}, a = \hat{a}) = \prod_{i=1}^{T} \tau_{\hat{s},\hat{a},i}^{I[i=\hat{t}]},$$
$$s.t. \quad \forall_{s,a} \sum_i \phi_{s,a,i} = 1, \forall_{s,a} \sum_i \tau_{s,a,i} = 1$$

During the M-step of the EM algorithm, updates for $\theta$ and $\omega$ remain the same as the $CoRec$ model. The updates for $\phi$ and $\tau$ are given below:

$$\phi_{s^n,a,i} = \frac{\sum_{n=1}^{N} \sum_a P^n(a|d) \sum_{b \in B} I[i=b]}{|B_d^n| \sum_{n=1}^{N} \sum_a P^n(a|d)}$$

$$\tau_{s^n,a,i} = \frac{\sum_{n=1}^{N} \sum_a P^n(a|d) \sum_{t \in T} I[i=t]}{|T_d^n| \sum_{n=1}^{N} \sum_a P^n(a|d)}$$

### 4.4 Subject and Body Token Representations

Previous work in modeling email content mostly explored bag of words (e.g., (Graus et al., 2014)) or tf-idf vectors (e.g., (Carvalho and Cohen, 2007)) as the content representation of an email. For modeling activities in emails, we experiment with different linguistic representations of the email content. They are:

- Lexical: as the lexical representation, we use the bag of words (BOW) from email body and subject, after Penn Tree Bank (PTB) style tokenization.
- Syntactic: using heuristics on the output of a PTB constituent parser (Quirk et al., 2012), we identify Nouns (N) and Verb Phrases (VP) in email body and subject.
- Semantic: we identify phrases in emails that represent topics, concept and entities discussed in the emails. We refer to them as *Thing of Interest (TOI)*. To extract these key phrases, we use Web search queries as a source of information. Using queries as a dictionary of possible key phrases is useful but has limited coverage since many topics/concepts are discussed in emails but absent or not widely available in Web search queries. Instead of using queries directly, we use them to construct a training set of key phrases and their content and we train a discriminative model to identify the key phrases. We treat each query as a key phrase and the surrounding text from Web results as context. We use a sample of hundreds of thousands of search queries from the usage logs of a commercial Web search engine. Only queries tagged as English and from the United States locale were retained to remove geographic or linguistic variations. Queries were kept only if they have been submitted by at least 100 different users in one month. For each query, the text from the Web page that is most relevant to the query and that contains the exact query text is collected as the context for the query. Relevance is estimated by the percentage of time the page has received a long dwell time click (greater than 30 seconds) for the query. If no relevant pages exist, the query is ignored. To generate negative examples, random n-grams were extracted from web pages. We experimented with a large number of features including: first word of the phrase, last word of the phrase, n-gram features (n=1 to 3), the word right before/after the phrase, the part-of-speech tag of the first word in the phrase, the part-of-speech tag of the last word in the phrase, n-gram features (n ranges from 1 to 3) over the sequence of part-of-speech tags representing the phrase and the part-of-speech tags of the word right before/after the phrase, phrase length, how many times it appeared in the body/title, and the relative location of the first occurrence of the phrase in the body. We trained a logistic regression classifier using these features and the data described above. The trained classifier is then applied to our email data to identify the *Thing of Interest (TOI)* phrases.

### 4.5 Discussion

**Full Bayesian Treatment:** In the above models, we learn point estimates for the parameters $(\sigma, \theta, \omega, \phi, \tau)$. One can take a Bayesian approach and treat these parameters as variables (for instance, with Dirichlet prior distributions), and perform Bayesian inference. However, exact inference will become intractable and we would need to resort to methods such as variational inference or sampling. We found this extension unnecessary, as we had a sufficient amount of training data to estimate all parameters reliably. In addition, our approach enabled us to learn (and perform inference in) the model with large amounts of data with reasonable computing time.

## 5 Recipient Recommendation

To evaluate the effectiveness of our activity model, we formulate a recipient recommendation task.

**Task Definition:** For a test email document $d$ containing the list of recipients $R_d$, a modified list of recipients $R_d^*$ is created by removing the last recipient $r^* \in R_d$. Given $d$ with $R_d^*$, the task objective is to recommend $r^*$ as the next recipient for $d$.

### 5.1 Our Methods

To recommend a recipient for a test email document $d$ written by sender $s_d$, we first create a candidate recipient list by combining recipients who received an email from $s_d$, and recipients who co-occurred with an observed recipient $r \in R_d^*$ in the training corpus. Sender $s_d$ and any $r \in R_d^*$ are excluded from the candidate list. Next, we determine the probability distribution of the activities in $d$ using:

$$P(a|d) = \frac{P(s, a, d|\sigma, \theta, \omega, \phi, \tau)}{\sum_a P(s, a, d|\sigma, \theta, \omega, \phi, \tau)}$$

Each candidate recipient $r^*$ is then ranked by a score using two different methods defined below. The ranked list is used as our final recommended recipients. The two scoring methods are:

**Reg Method:** In the $Reg$ method, we score using the chain rule[6]:

$$P(r^*|d) \propto \sum_a P(a|d) \prod_{r \in R_d} P(r^*, r|s, a)$$

We smooth the above function using the following linear interpolation:

$$P(r^*, r|a, s) = \alpha_1 \times P(r^*, r|a, s) + (1 - \alpha_1) \times (\alpha_2 \times P(r^*, r) + (1 - \alpha_2) \times P(r_{rare}))$$

Here, $P(r_{rare})$ is the lowest probability of any recipient in the training data. We calculate $\alpha_i$ with a sigmoid logistic function, allowing us to determine when to rely more on the learned probabilities:

$$\alpha_i = \frac{1}{1 + e^{-k(x - x_0)}}$$

For $\alpha_1$, $x$ is the pointwise mutual information (PMI) between $s$ and $r$ in training data, with steepness parameter $k = 50$. For $\alpha_2$, $x$ is the frequency of $r$ in training data, with $k = .5$. Sigmoid's midpoint $x_0$ is the first quartile ($Q1$) of the PMI and recipient frequency distributions respectively. The above values for $k$ have been determined from the shape of the sigmoid curves in the training data.

**Sim Method:** In the $Sim$ method, we explore the idea that the activity model can be used to identify other senders with similar activities as $s_d$, who we refer to as *similar senders*, $S_d^*$. To identify the similar senders, we evaluate senders who maximize the log likelihood of the test document $d$ by calculating $\log P(s, d)$ for all $s \in S$, and identify the top 5 with the highest scores to add to $S_d^*$. The observed sender $s_d$ is not included in $S_d^*$. We then calculate $P_s(r^*|d)$ for each $s \in S_d^*$ using the $Reg$ method, along with a weight $w_s$:

$$w_s = \frac{\log P(s, d)}{\sum_{s \in S_d^*} \log P(s, d)}$$

The final scoring function for the $Sim$ method is:

$$P(r^*|d) = \alpha P_{s_d}(r^*|d) + (1 - \alpha) \sum_{s \in S_d^*} w_s P_s(r^*|d)$$

Here, $\alpha$ is determined with the frequency of $s_d$ in training data, using the sigmoid function with $k = 0.5$ and $x_0$ as the $Q1$ of the frequency distribution.

### 5.2 Baseline Systems

As simple baseline systems to compare with our methods, we use 1) a random recipient baseline; 2) ranked recipients by $P(r = r^*)$; and 3) ranked recipients by $P(r = r^*|s = s_d)$, where the probabilities are calculated from the training data. We evaluate two additional generative baselines using 4) $P(r = r^*|R_d^*)$, and 5) $P(r = r^*|s = s_d, R_d^*)$ by applying Bayes' theorem, and assuming conditional independence among $r \in R_d^*$. For these methods, we used similar interpolation smoothing as before.

We additionally implemented the generative model presented by Graus et al. (2014) for recipient recommendation, which for test email $d$ uses:

$$P(r^*|s_d, d) \propto P(d|r^*, s_d) \times P(s_d|r^*) \times P(r^*)$$

Graus estimated $P(d|r^*, s_d)$ by $P(b|r^*, s_d)$ where $b$ is an observed term in the email. The evaluation task was different from ours as they predicted all recipients of an email. In our evalua-

[6] Scoring function for the $Rec$ model uses $P(r^*|s, a)$

| Method | Precision@1 | Precision@2 | Precision@5 | Precision@10 | MRR |
|---|---|---|---|---|---|
| Baselines | | | | | |
| (1) Random | 0 | 0 | 0 | .10 | .0025 |
| (2) $P(r = r^*)$ | 2.81 | 4.58 | 7.49 | 17.32 | .0736 |
| (3) $P(r = r^*\|s = s_d)$ | 4.47 | 9.72 | 24.18 | 34.69 | .1455 |
| (4) $P(r = r^*\|R_d^*)$ | 17.26 | 25.59 | 39.42 | 53.93 | .2857 |
| (5) $P(r = r^*\|s = s_d, R_d^*)$ | 16.80 | 25.01 | 42.02 | 56.16 | .2871 |
| Graus Methods (Graus et al., 2014) | | | | | |
| (6) GrausB (BOW) | 2.96 | 4.84 | 8.01 | 17.94 | .0769 |
| (7) GrausB (VP-TOI) | 4.63 | 9.00 | 18.25 | 28.86 | .1257 |
| (8) GrausR | 18.88 | 27.2 | 41.97 | 54.39 | .3005 |
| Activity Models (Reg Scoring) | | | | | |
| (9) Rec | 12.27 | 19.81 | 30.53 | 44.46 | .2224 |
| (10) CoRec | 21.63 | 29.07 | 41.45 | 52.16 | .3167 |
| (11) CoRecBT (BOW) | 19.97 | 27.87 | 40.77 | 52.16 | .3037 |
| (12) CoRecBT (NP-VP-TOI) | 20.64 | 28.29 | 40.93 | 51.79 | .3081 |
| (13) CoRecBT (VP-TOI) | 20.59 | 29.17 | 41.39 | 51.95 | .3104 |
| Activity Models (Sim Scoring) | | | | | |
| (14) CoRecBT (NP-VP-TOI) | 22.01 | 30.47 | 44.36 | 56.01 | .3306 |
| (15) CoRecBT (VP-TOI) | **22.26** | **33.63** | **44.57** | **57.05** | **.3336** |

**Table 2:** Recipient recommendation results (BOW = bag-of-words, NP = noun phrase, VP= verb phrase, TOI = thing of interest). Bold indicates statistical significance over all non-shaded results using t-test (p=0.05).

tion task, we recommend the last recipient, allowing us to use the already observed recipients $R_d^*$ for estimating $P(d|r^*, s_d)$. Consequently, we present 3 additional baselines adopting Graus' method: 6) $GrausB(BOW)$ method uses body words, 7) $GrausB(VP - TOI)$ uses the verb phrases and things of interest, and finally 8) $GrausR$ method uses $R_d^*$ for estimating $P(d|r^*, s_d)$. $GrausR$ is equivalent to how we calculate our fifth baseline, $P(r = r^*|s = s_d, R_d^*)$, with the only difference of the smoothing function.

## 6 Experimental Results

### 6.1 Experimental Setup

To evaluate recipient recommendation, we create a test data set by extracting emails from 7/1/2001 – 8/31/2001 from the Avocado data set. First we train our activity model with the training data and determine the optimum number of activities for each method by evaluating recipient recommendation on the development data. The number of activities per model is shown in Table 3. We then combine training and tuning data to create a new training data set in order to minimize the time difference between training and test emails. From the test data, we removed emails that had a sender or recipient never appearing in the training data. Although this limits the scope of the recipient recommendation evaluation task, predicting a recipient for a sender who never appeared in the training data is beyond our current modeling scope and practical settings. The final test set contains 1923 emails with 14.91 emails per sender.

| Model | K |
|---|---|
| Rec | 10 |
| CoRec | 3 |
| CoRecBT (BOW) | 20 |
| CoRecBT (NP-VP-TOI) | 7 |
| CoRecBT (VP-TOI) | 4 |

**Table 3:** No. of activities used for recipient recommendation.

With the ranked lists of recipients generated by each method, we calculate precision@X (X= 1, 2, 5, 10), and MRR (Mean Reciprocal Rank). Precision@X is defined as percentage of emails having the actual recipient in the top X ranked recipients.

### 6.2 Recipient Recommendation Results

Table 2 presents the recipient recommendation results for different methods. The first 5 rows show that the generative baselines from row (4) and (5) performed much better than the simple baselines (row (1)–(3)), yielding up to .2871 $MRR$. Comparatively, the GrausB(BOW) baseline in row (6) that uses body words, did not perform well, which is consistent with the finding by Shen et al. (2006) about


**Figure 2:** Word clouds of activity tokens.

body words not providing additional value in their task classification work. However, the GrausB(VP–TOI) in row (7) shows that using body terms more selectively has the potential for improving performance. Comparatively, the use of observed recipients (GrausR in row (8)) substantially improved recommendation results, yielding the highest $MRR$, precision@1, and precision@2 scores, while the generative baseline in row (5) retained the highest precision@5 and precision@10 scores.

Next, rows (9) to (13) show results for the activity models that use the $Reg$ scoring. First, the $Rec$ model outperformed the simple baselines from rows (1) to (3) as well as the GrausB methods from rows (6) and (7), but did not perform better than the generative baselines from rows (4) and (5) or GrausR. All the $CoRec$ models performed better at recommending a recipient at top of the ranked lists with higher precision@1 and precision@2 scores, which are more practically useful for recommendation purposes, and also resulted in higher $MRR$ scores.

Finally, the rows (14) and (15) show the results with the $Sim$ scoring, and we observe a substantial improvement across the board, with verb phrase and thing of interest as the body context yielding our best results. This model achieved 3.31% additional improvements in MRR, and 3.38% additional improvements in precision@1 over the best baseline results. This demonstrates that the learned activity model can be used to identify senders who are likely to engage in similar activities, improving recipient recommendation performance further.

Figure 2 shows examples of activity tokens from the emails of a sender in our training corpus, using word clouds. This is meant to serve as a case analysis, but it is not straightforward to interpret word clouds. When inspecting, we found that the names of potential customers (Nokia, Siemens, SAP) in the first example are prominent in some of the emails of the sender in the raw data. The recipients in these emails form a small cluster of people who are mainly involved in discussions around a particular event (Mobile Business Forum) where these companies are amongst the sponsors.

The second example, from the same sender, shows a coherent set of recipients. But in this case, the model seemed to have conflated multiple topics (such as the Palm VII device and support issues). We suspect that the cause for this confusion lies in the strong and coherent cluster of recipients which forces divergent topics to coalesce. While the combined signals of co-recipients and topic words improve the overall activity model, in some of the individual cases it leads to one signal improperly dominating the other.

## 7 Conclusion and Future Work

We presented a latent activity model for workplace emails where the activities are modeled as probability distributions over email recipients and other contexts, personalized to the email sender. Our model incorporates co-occurring recipients as part of the generative process, and can be used to identify senders who participate in similar activities, resulting in improved performance in email recipient recommendation. Our experiments suggest that syntactic and semantic knowledge such as verb phrases and thing of interests in emails can model the activities much better than bag-of-words, as demonstrated by the recipient recommendation results. Learning topics and sub-activities under workplace activities is a promising research direction which we will explore in future work.

## References


Aruna D Balakrishnan, Tara Matthews, and Thomas P Moran. 2010. Fitting an activity-centric system into an ecology of workplace tools. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 787–790. ACM.

Ramnath Balasubramanyan and William W Cohen. 2011. Block-lda: Jointly modeling entity-annotated text and entity-entity links. In *SDM*, volume 11, pages 450–461. SIAM.

Victoria Bellotti, Nicolas Ducheneaut, Mark Howard, and Ian Smith. 2003. 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*, pages 345–352. ACM.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.

Vitor R Carvalho and William Cohen. 2007. Recommending recipients in the enron email corpus. *Machine Learning*.

Amish Desai and Subrat Kumar Dash. 2014. Email recipient prediction using reverse chronologically arranged implicit groups. In *Contemporary Computing (IC3), 2014 Seventh International Conference on*, pages 461–466. IEEE.

Mark Dredze and Hanna Wallach. 2008. User models for email activity management. In *Proceedings of the 5th International Workshop on Ubiquitous User Modeling (UbiqUM'08)*, Gran Canaria, Spain, January. online proceedings forthcoming.

Mark Dredze, Tessa Lau, and Nicholas Kushmerick. 2006. Automatically classifying emails into activities. In *Proceedings of the 11th International Conference on Intelligent User Interfaces*, IUI '06, pages 70–77, New York, NY, USA. ACM.

Mark Dredze, Hanna M Wallach, Danny Puller, and Fernando Pereira. 2008. Generating summary keywords for emails using topics. In *Proceedings of the 13th international conference on Intelligent user interfaces*, pages 199–206. ACM.

Tamer Elsayed, Douglas W Oard, and Galileo Namata. 2008. Resolving personal names in email using context expansion. In *ACL*, pages 941–949.

David Graus, David van Dijk, Manos Tsagkias, Wouter Weerkamp, and Maarten de Rijke. 2014. Recipient recommendation in enterprises using communication graphs and email content. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 1079–1082. ACM.

Zan Huang and Daniel Dajun Zeng. 2006. A link prediction approach to anomalous email detection. In *SMC*, pages 1131–1136.

Rinat Khoussainov and Nicholas Kushmerick. 2005. Email task management: An iterative relational learning approach. In *CEAS*.

Nicholas Kushmerick and Tessa Lau. 2005. Automated email activity management: an unsupervised learning approach. In *Proceedings of the 10th international conference on Intelligent user interfaces*, pages 67–74. ACM.

Andrew McCallum, Andres Corrada-Emmanuel, and Xuerui Wang. 2005. Topic and role discovery in social networks. *Computer Science Department Faculty Publication Series*, page 3.

Einat Minkov, Ramnath Balasubramanyan, and William W Cohen. 2008. Activity-centred search in email. In *CEAS*. Citeseer.

Thomas P Moran. 2005. Unified activity management: Explicitly representing activity in work-support systems. In *Proceedings of the European Conference on Computer-Supported Cooperative Work (ECSCW 2005), Workshop on Activity: From Theoretical to a Computational Construct*. Citeseer.

Nicholas Navaroli, Christopher DuBois, and Padhraic Smyth. 2012. Statistical models for exploring individual email communication behavior. In *ACML*, pages 317–332.

Muon Nguyen, Thanh Ho, and Phuc Do. 2013. Social networks analysis based on topic modeling. In *Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), 2013 IEEE RIVF International Conference on*, pages 119–122. IEEE.

Chris Pal and Andrew McCallum. 2006. Cc prediction with graphical models. In *CEAS*.

Chris Pal, Xuerui Wang, and Andrew McCallum. 2007. Transfer learning for enhancing information flow in organizations and social networks. Technical report, DTIC Document.

Chris Quirk, Pallavi Choudhury, Jianfeng Gao, Hisami Suzuki, Kristina Toutanova, Michael Gamon, Wen-tau Yih, Lucy Vanderwende, and Colin Cherry. 2012. Msr splat, a language analysis toolkit. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstration Session*, pages 21–24. Association for Computational Linguistics.

Yanir Seroussi, Fabian Bohnert, and Ingrid Zukerman. 2012. Authorship attribution with author-aware topic models. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics:*

*Short Papers-Volume 2*, pages 264–269. Association for Computational Linguistics.

Jianqiang Shen, Lida Li, Thomas G Dietterich, and Jonathan L Herlocker. 2006. A hybrid learning system for recognizing user tasks from desktop activities and email messages. In *Proceedings of the 11th international conference on Intelligent user interfaces*, pages 86–92. ACM.

Zvi Sofershtein and Sara Cohen. 2015. Predicting email recipients. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, ASONAM '15, pages 761–764, New York, NY, USA. ACM.

Xuerui Wang and Andrew McCallum. 2006. Topics over time: a non-markov continuous-time model of topical trends. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 424–433. ACM.