Calendar-Aware Proactive Email Recommendation

QIAN ZHAO. **UNIVERSITY OF MINNESOTA***

PAUL N. BENNETT, ADAM FOURNEY, ANNE LOOMIS THOMPSON, SHANE WILLIAMS, ADAM D. TROY, SUSAN T. DUMAIS. **MICROSOFT**

* Work done while having an internship at Microsoft.

An advisor-student meeting

An advisor-student meeting

Anna: Hi Professor, thanks for meeting with me.

Professor: So, what are we going to talk about today?

An advisor-student meeting

Anna: Hi Professor, thanks for meeting with me.

Professor: So, what are we going to talk about today?

Anna: I sent you an email with a research proposal.

Professor: Okay, let me pull up my emails … emails from Anna (searching) …

An advisor-student meeting

Anna: Hi Professor, thanks for meeting with me.

Professor: So, what are we going to talk about today?

Anna: I sent you an email with a research proposal.

Professor: Okay, let me pull up my emails … emails from Anna (searching) …

Professor: When did you send it?

Anna: Yesterday.

An advisor-student meeting

Anna: Hi Professor, thanks for meeting with me.

Professor: So, what are we going to talk about today?

Anna: I sent you an email with a research proposal.

Professor: Okay, let me pull up my emails … emails from Anna (searching) …

Professor: When did you send it?

Anna: Yesterday.

Advisor: Okay, looks like there are a few since we met last time. One moment … The one with "research proposal" in the title?

An advisor-student meeting

Anna: Hi Professor, thanks for meeting with me.

Professor: So, what are we going to talk about today?

Anna: I sent you an email with a research proposal.

Professor: Okay, let me pull up my emails … emails from Anna (searching) …

Professor: When did you send it?

Anna: Yesterday.

Advisor: Okay, looks like there are a few since we met last time. One moment. The one with "research proposal" in the title?

Anna: Yes!

Can the email client further help?

Anna: Hi Professor, thanks for meeting with me.

Professor: So, what are we going to talk about today?

Anna: I sent you an email with a research proposal.

Professor: Okay, let me pull up my emails … ~~emails from Anna (searching) …~~

~~Professor: When did you send it?~~

~~Anna: Yesterday.~~

~~Advisor: Okay, looks like there are a few since we met last time. One moment.~~ The one with "research proposal" in the title?

Anna: Yes!

Useful Emails for Your Meeting **Meet with Anna at 2:00 pm today**
From Anna, Yesterday 4:00 pm Subject: Research Proposal Hi Professor, attached is my research proposal…
From <sender>, <time> Subject: <subject> <Email Preview>
From <sender>, <time> Subject: <subject> <Email Preview>
…
…

CAPERS: a Calendar-Aware Proactive Email Recommender System

Proactively select and display potentially useful emails

Based on available calendar information

To help with email re-finding

RQs about CAPERS

Why could it potentially be useful? Is it an important problem to solve?

RQs about CAPERS

Why could it potentially be useful? Is it an important problem to solve?

How to design the interface of CAPERS and collect user feedback?

RQs about CAPERS

Why could it potentially be useful? Is it an important problem to solve?

How to design the interface of CAPERS and collect user feedback?

What meetings to focus on at any time?

RQs about CAPERS

Why could it potentially be useful? Is it an important problem to solve?

How to design the interface of CAPERS and collect user feedback?

What meetings to focus on at any time?

What factors affect the usefulness of emails for a meeting and how?

RQs about CAPERS

Why could it potentially be useful? Is it an important problem to solve?

How to design the interface of CAPERS and collect user feedback?

What meetings to focus on at any time?

What factors affect the usefulness of emails for a meeting and how?

What models and algorithms to use and how to deal with system-level cold-start?

Methods

Survey

Task-based experiment (with a pilot study)

Field experiment

Survey

Questions asking about

- meeting preparation habits
- how emails are used for meeting preparation

592 participants

Do people access emails for meetings?

Do people access emails for meetings? Yes!

Frequency of Email Access for Meetings

- “How often do you prepare for meetings by reading, replying or forwarding relevant emails?” (with single-selection options: "always", "frequently", "regularly", "occasionally" or "never")
- **68.4%** regularly or frequently or always

The Last Email Access

- “What was your last email access about?” (options given for meeting related and non-meeting related behaviors)
- **32%** meeting-related

CAPERS Design

How to design the interface of CAPERS and collect user feedback?

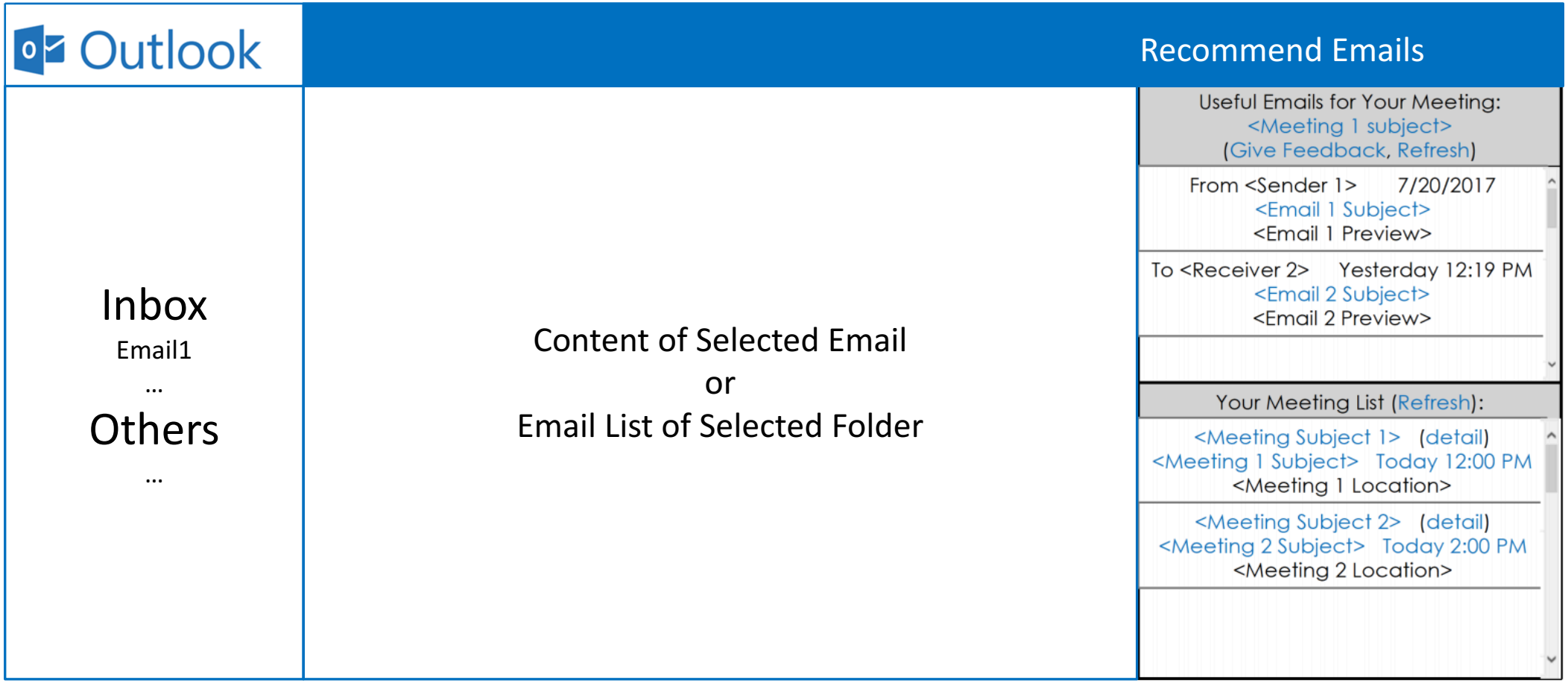
- Iterative process taking account of the survey and pilot study feedback

What meetings to focus on at any time? *

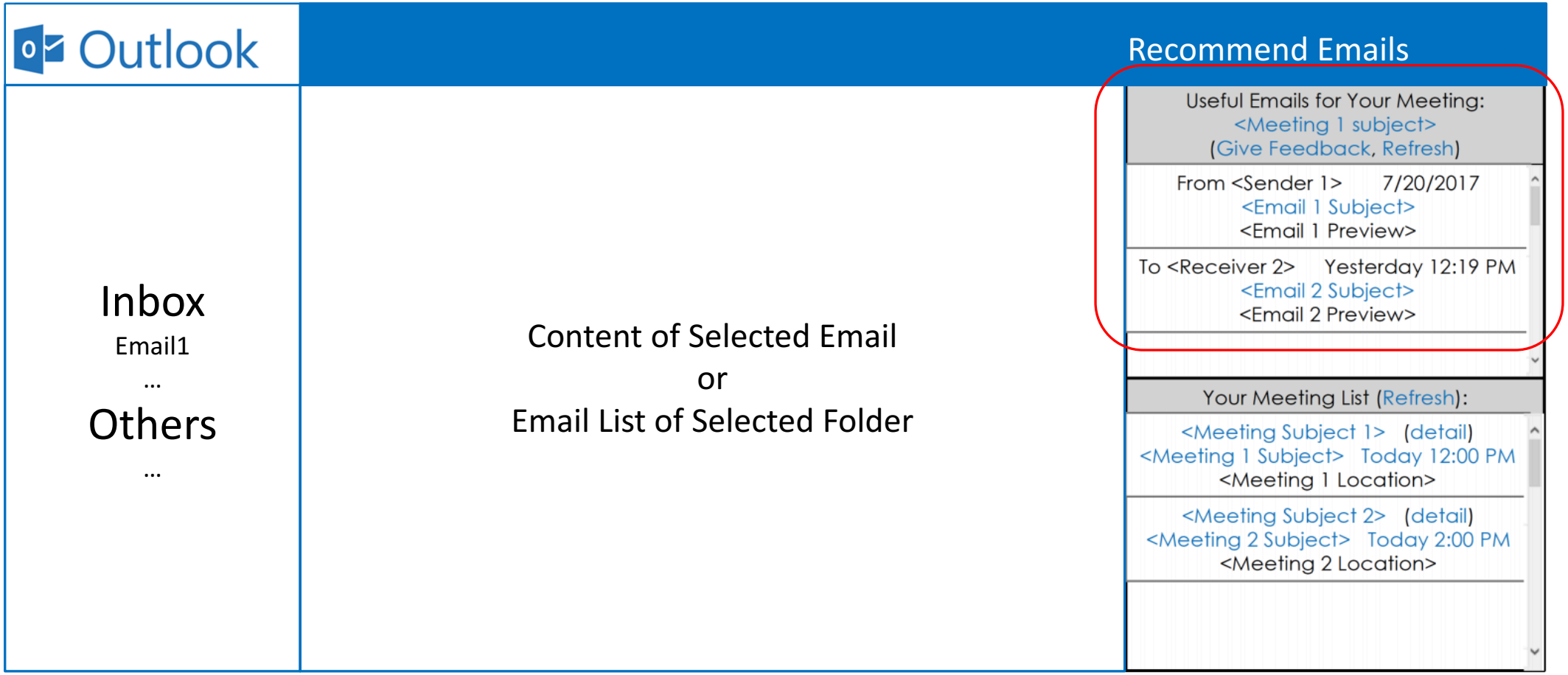
- not based on predictive models
- give users the control to select

* See the paper for details on how attendee roles and meeting sizes etc. affect the likeliness of preparation.

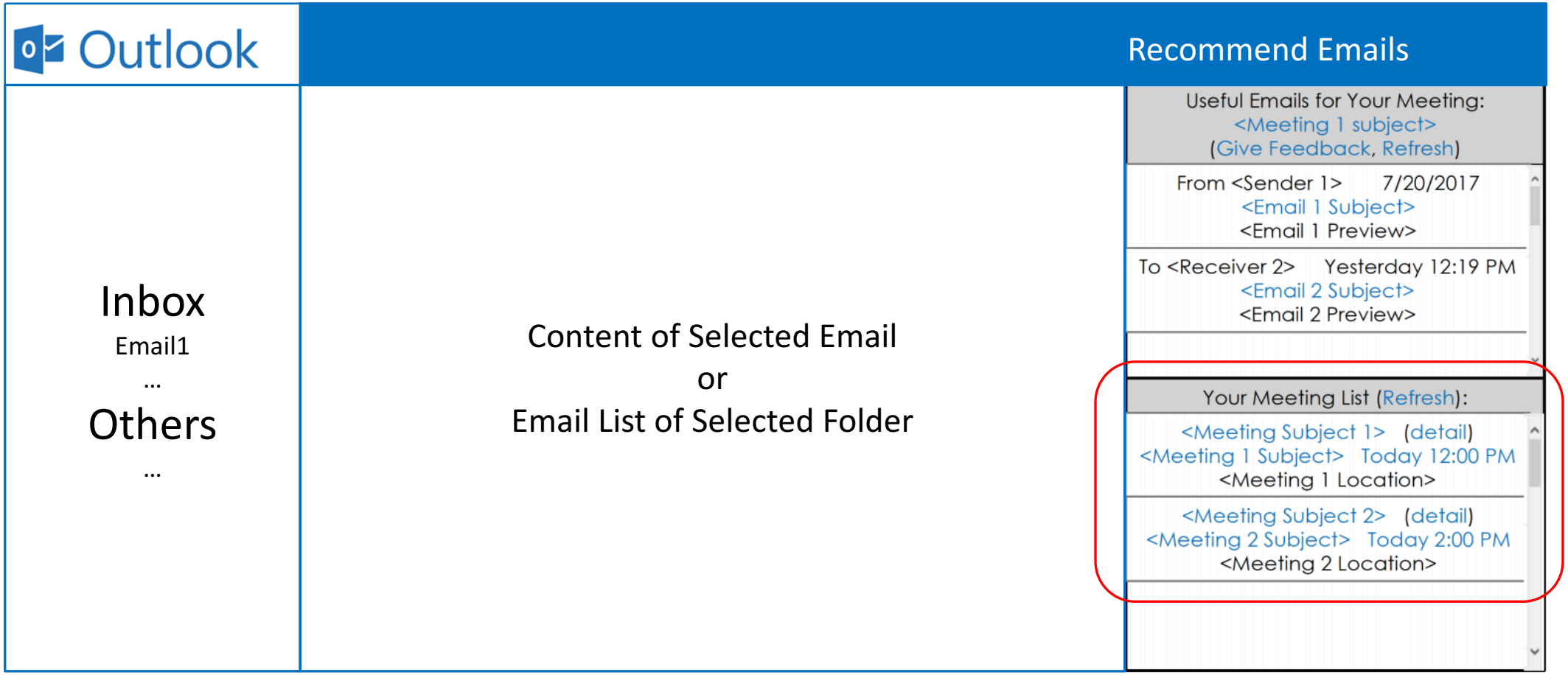
CAPERS Design Mockup (Outlook Add-in)



CAPERS Design Mockup (Outlook Add-in)



CAPERS Design Mockup (Outlook Add-in)



What models and algorithms to use and how to deal with system-level cold-start?

Design user studies and first-run experience to collect feedback from users

Interactive real-time machine learning

User Study Design

First-run task-based experiment (within-subject)

- Step 1: “Select a meeting that you are most likely to prepare for”
- Step 2: “For each of the displayed emails, please label them as Not Useful, Maybe Useful or Useful for you to prepare for the selected meeting”
- (predictive models are updated in real-time while labeling)
- Step 3: “We have built 6 recommenders, please tell us your preferences on them…”
- …

User Study Design

Field experiment (between-subjects)

- Random and persistent assignment after the first-run task-based experiment (5 recommenders)
- Measure activities for a month for each user

Four Categories of Usefulness-Impacting Factors

Factor Category	Feature Name
Amount of Content	*EmailLength*
	HavingAttachment
Email Recency	*RecencyScore*
Content Match	*ContentMatchScore*
	SubjectMatchScore
People Match	*PeopleMatchScore*
	FromAttendee
	FromOrganizer

The Six Recommenders

Non-learned

- Time: most recent received or sent (simple baseline, not used in the field experiment)
- Search: pick three key-words from meeting subject and search (mimicking human search)

Learned from data (through Samantha*)

- Static Linear: linearly fuse four categories of factors, fixed *bootstrapped* weights (adapted LinearUCB)
- Online Linear: on top of “Static Linear”, online updating of the weights (adapted LinearUCB)
- Static Hero: on top of “Static Linear”, pick the top two and sort the rest by time (inspired by Camel et al. WWW’17)
- Online FM: same set of features as ”Online Linear”, factorization machine model and online updating (SGD)

* https://github.com/grouplens/samantha

The Pilot Study

20 participants

Go through the first-run task-based experiment and label additional 10 emails for another meeting

334 labeled data points (used for model bootstrapping)

Results

Feature Name	Coefficient (Std.)
EmailLength	1.72 (0.326) ***
RecencyScore	1.10 (0.333) ***
SubjectMatchScore	3.53 (0.638) ***
PeopleMatchScore	1.58 (0.421) ***
RecencyScore:LowContentMatch	-0.319 (0.382)
RecencyScore:LowPeopleMatch	-0.167 (0.384)
EmailLength:LowContentMatch	-1.21 (0.379) **
EmailLength:LowPeopleMatch	-1.01 (0.388) **

SubjectMatchScore is the most important.

Significance codes: ** p<0.01, *** p<0.001 (from mixed-effect ordinal regression model).

Results

Feature Name	Coefficient (Std.)
EmailLength	1.72 (0.326) ***
RecencyScore	1.10 (0.333) ***
SubjectMatchScore	3.53 (0.638) ***
PeopleMatchScore	1.58 (0.421) ***
RecencyScore:LowContentMatch	-0.319 (0.382)
RecencyScore:LowPeopleMatch	-0.167 (0.384)
EmailLength:LowContentMatch	-1.21 (0.379) **
EmailLength:LowPeopleMatch	-1.01 (0.388) **

Interaction between *EmailLength* and *PeopleMatch, ContentMatch*

Significance codes: ** p<0.01, *** p<0.001 (from mixed-effect ordinal regression model).

Results

Recommender	Mean (std.), median of *numRequests*
Search	3.80 (1.39), 2.0
Static Linear	5.05 (0.880), 3.0
Online FM	3.14 (0.716), 3.0
Online Linear	**11.1 (2.78), 7.0** >Search* >Static Linear** >Online FM***

~2x engagement for Online Linear vs. Search, Static Linear, Online FM

Significance codes: * p<0.05, ** p<0.01, *** p<0.001 (from negative binomial regression model).

Accuracy or *online updating*?

Recommender	nDCG@5	nDCG@10	nDCG@15
Static Linear	0.837	0.807	0.804
Static Hero	0.832	0.810	0.785
Online Linear	0.835	0.805	0.802
Online FM	0.793	0.768	0.753

nDCG (Online FM) < nDCG (Online Linear), but
nDCG (Online Linear, Static Linear, Static Hero) not substantially different

Takeaways

Calendar-aware email recommendation is scientifically interesting and practically important.

Takeaways

Calendar-aware email recommendation is scientifically interesting and practically important.

Recency, amount of information, people match and content match are all important factors.

Takeaways

Calendar-aware email recommendation is scientifically interesting and practically important.

Recency, amount of information, people match and content match are all important factors.

Continuously online updating is effective for system-level cold-start and its algorithms for more complex models need more work (e.g., online updating of GBDT, GB-CENT WWW'17).

Takeaways

Calendar-aware email recommendation is scientifically interesting and practically important.

Recency, amount of information, people match and content match are all important factors.

Continuously online updating is effective for system-level cold-start and its algorithms for more complex models need more work (e.g., online updating of GBDT, GB-CENT WWW'17).

Recency can be captured well similar to other factors (not as dominant as in email search).

Takeaways

Calendar-aware email recommendation is scientifically interesting and practically important.

Recency, amount of information, people match and content match are all important factors.

Continuously online updating is effective for system-level cold-start and its algorithms for more complex models need more work (e.g., online updating of GBDT, GB-CENT WWW'17).

Recency can be captured well similar to other factors (not as dominant as in email search).

CAPERS reflects a thread of future work and novel applications based on contextual intelligence.

Acknowledgements

Microsoft (for supporting the work with an internship)

GroupLens lab at the University of Minnesota (for continuous support of the work)

Bloomberg (for supporting my travel here)

Thank you! Q&A?

“Calendar-Aware Proactive Email Recommendation”

Qian Zhao. **University of Minnesota*** (zhaox331@umn.edu)

Paul N. Bennett, Adam Fourney, Anne Loomis Thompson, Shane Williams, Adam D. Troy, Susan T. Dumais. **Microsoft**.

* Work done while having an internship at Microsoft. Qian Zhao is joining Bloomberg soon.