

The Yahoo! logo is displayed in a purple, serif font within a white rectangular box in the top right corner of the slide.

E-commerce in Your Inbox

Product Recommendations at Scale

M. Grbovic, V. Radosavljevic, *N. Djuric*, N. Bhamidipati (Yahoo Labs, **Ad Targeting Team**)
J. Savla, V. Bhagwan, D. Sharp (Yahoo, Inc., **Yahoo Mail**)

Introduction

- Distributed embeddings recently gained in popularity
- Tested in a number of applications at Yahoo
 - Search retargeting (WWW 2015)
 - Query categorization (WWW 2015)
 - Query rewriting (SIGIR 2015)
 - Targeting at Tumblr (KDD 2015)
- This talk: Yahoo Mail (KDD 2015)

Introduction

- We can't avoid ads in e-mail accounts
 - Improve user experience (and make more money) through product ads

The screenshot shows the Yahoo! Mail interface. At the top, there's a search bar and buttons for 'Search Mail' and 'Search Web'. Below that, there are navigation icons for mail, calendar, and other services. The main area displays a list of emails. The first email is a product ad from Walmart.com for 'Admiral U.S. Divers Adult LX Mask/Island Dry Snorkel/Trek Fins/Travel Bag, Yellow, Large' priced at \$44.96 - \$67.99. The second email is from Kate Wesson with the subject 'Re: Hi!' and a timestamp of 10:32 PM. The third email, highlighted with a red border, is from Amazon.com with the subject 'Your Amazon.com order of "Maui Revealed: The Ultimate Guidebook"' and a timestamp of 7:14 PM. The fourth email is from Xianzhe Ma with the subject 'Re: Can I borrow your truck?' and a timestamp of 5:42 PM. On the right side, there's a vertical advertisement for 'YAHOO! AUTOS' featuring a woman's profile picture and the name 'Jennifer'.

| | Compose | Delete | Move | Spam | More | View |
|------------------|--------------------------|--------------------|--|--|--|-------------------|
| Inbox (8) | | | | | | |
| Drafts | | | | | | |
| Sent | | | | | | |
| Spam | | | | | | |
| Trash | | | | | | |
| | | Walmart.com Offers | Admiral U.S. Divers Adult LX Mask/Island Dry Snorkel/Trek Fins/Travel Bag, Yellow, Large | | | \$44.96 - \$67.99 |
| | <input type="checkbox"/> | Kate Wesson | ★ | Re: Hi! | What's going on?! I feel like I haven't talked to you in forever. I thought about yc | 10:32 PM |
| | <input type="checkbox"/> | Amazon.com | | Your Amazon.com order of "Maui Revealed: The Ultimate Guidebook" | | 7:14 PM |
| | <input type="checkbox"/> | Xianzhe Ma | | Re: Can I borrow your truck? | I'll leave the keys with my roommate in case I'm not ther | 5:42 PM |

Introduction

- Hundreds of millions of people around the world visit their e-mail inboxes daily
- Ads need to be highly relevant to overcome focus on the e-mail task
- Effective personalization and targeting is essential to tackling this problem
 - Higher revenue, better user experience

Inbound e-mails

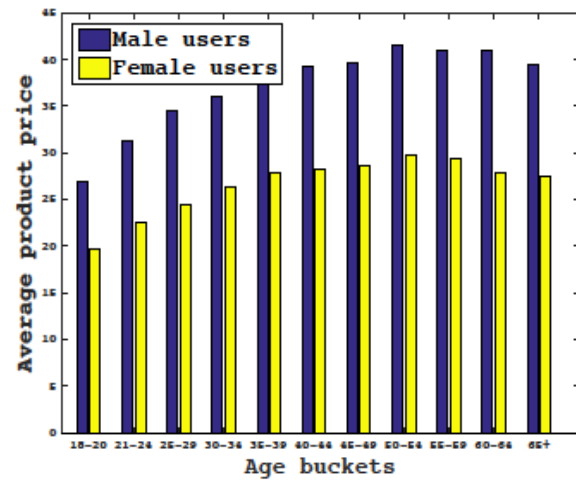
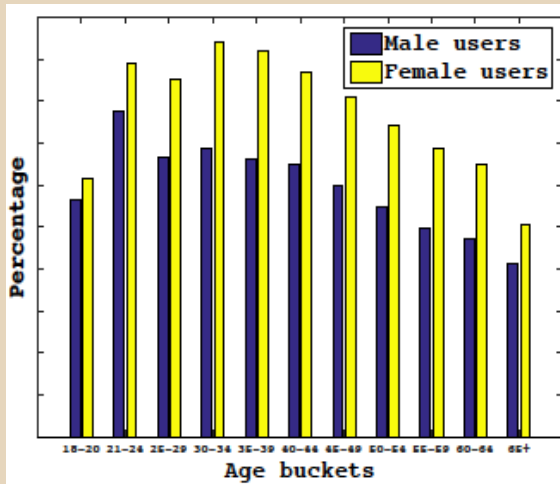
- Still insufficiently explored and exploited area for the purposes of ad targeting
 - Only 10% of inbound volume represents human-generated e-mails
 - For remaining 90% of traffic, more than 22% represents e-mails related to online shopping
- A treasure trove of data
 - Standardized online receipts
 - Data from multiple commercial domains

Data set

- Includes receipts sent to users who opted-in for such research studies
 - March to October 2014
 - Extracted product names and purchase times
 - 280.7M purchases from 172 commercial domains made by 29M users
 - 2.1M unique bought products priced over \$5

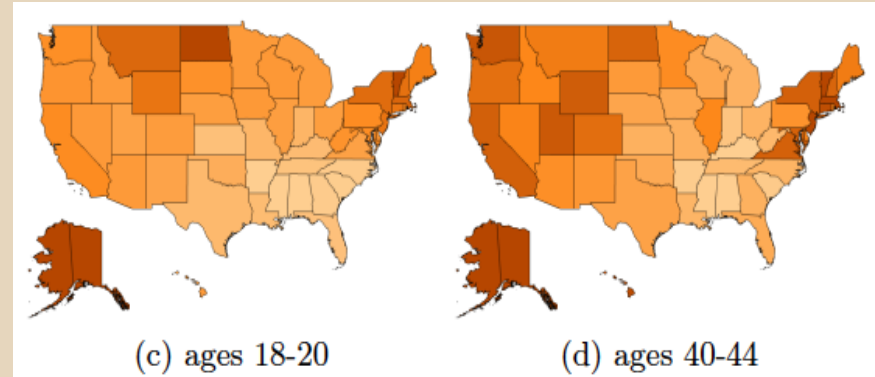
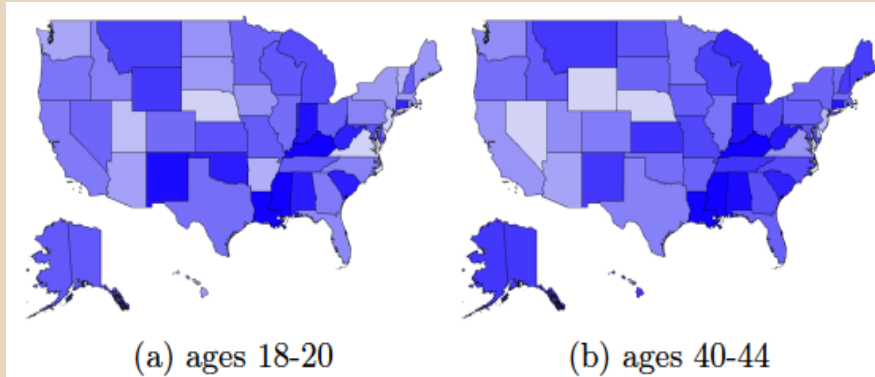
Data analysis

- Purchasing habits for different demographics
 - a. Percentage of female online shoppers is higher
 - b. Male users buy more expensive items



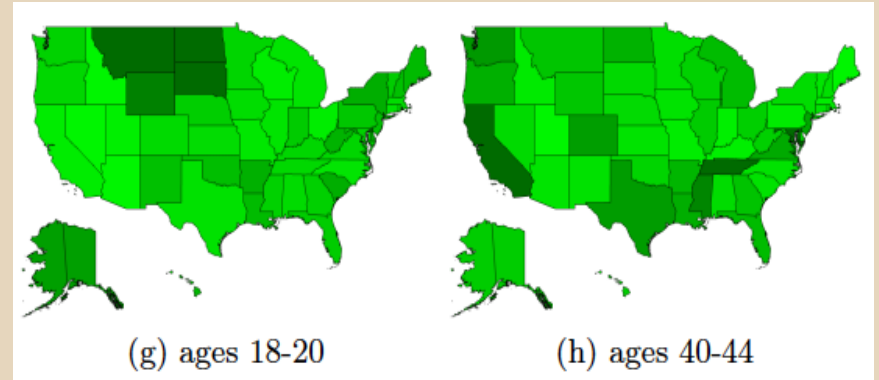
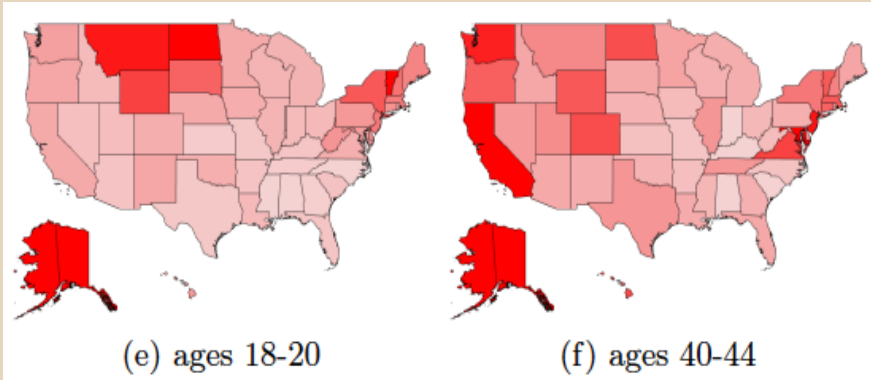
Data analysis

- Purchasing habits for different user cohorts
 - a. Percentage of shoppers among online users
 - b. Average number of purchases per user



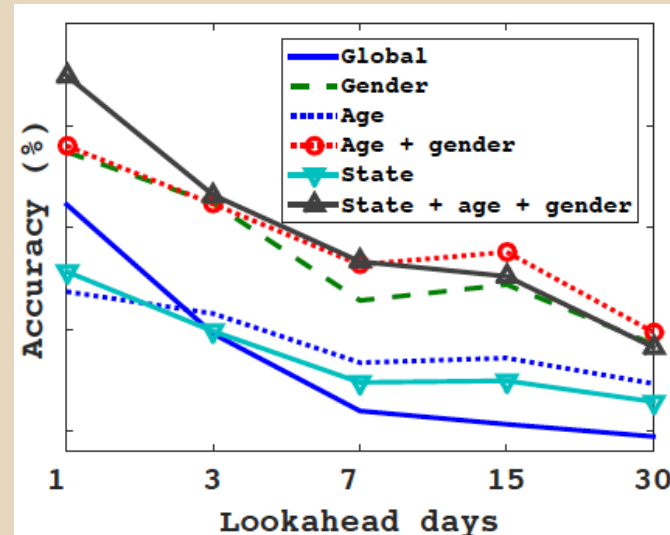
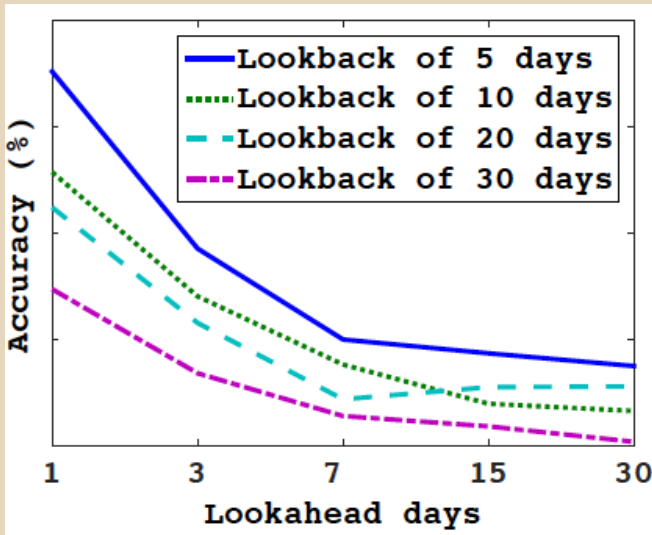
Data analysis

- Purchasing habits for different cohorts
 - a. Average amount spent by a user
 - b. Average price of purchased product



Recommending popular products

- Common and intuitive approach
- Lookback and lookahead parameters

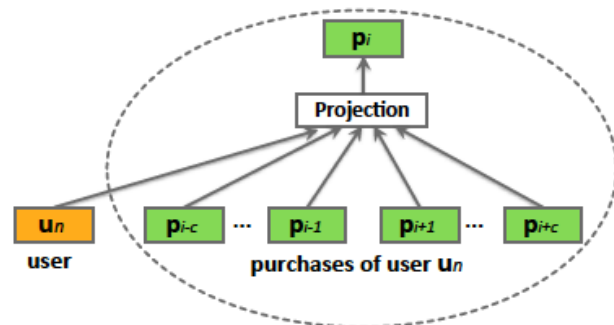
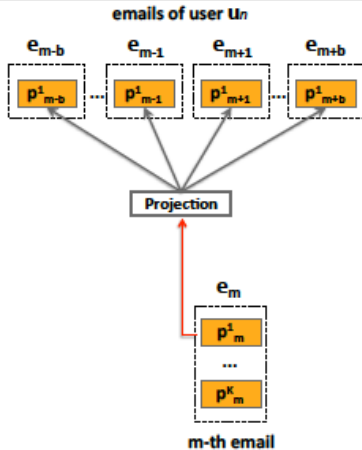
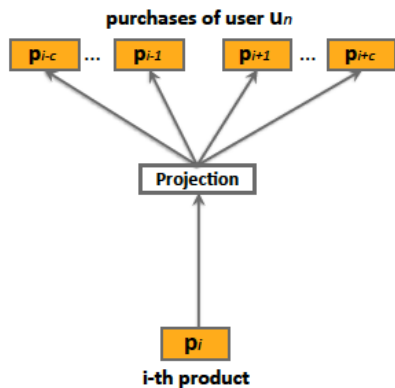


Neural language models

- Neural language models induce low- D , distributed embeddings of words using neural networks
- Recently proposed word2vec gained popularity
 - Applied to sentences, graphs, app prediction, ...
- Can it help in product recommendation?

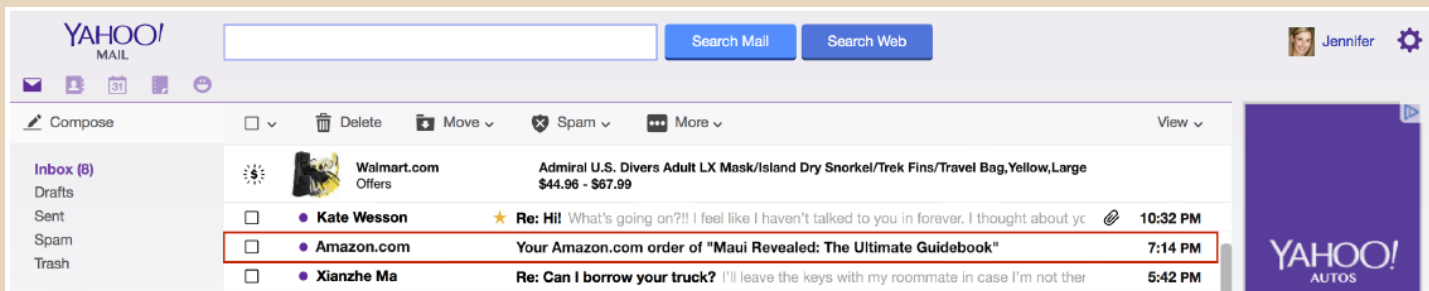
Proposed models

- prod2vec
- bagged-prod2vec
- user2vec



Proposed models

- Efficient product-level purchase prediction algorithm
 - Capable of scaling to millions of users and products
- Embed products to low- D space using neural language model applied to a time series of user purchases
 - Clustering and nearest-neighbor search



Product-to-product models

- prod2vec-top K
 - Use each purchased item to recommend its K neighbors to be shown to user
- prod2vec-cluster
 - Cluster the products, and empirically estimate probability that cluster i follows cluster j
 - Retrieve nearest neighbors from each of the high-probability clusters

Experiments

- The neighbors are highly relevant to the query

| despicable me | first aid for the usmle step 1 | disney frozen lunch napkins |
|-----------------------------------|--|--|
| monsters university | usmle step 1 secrets 3e | disneys frozen party 9 square lunchdinner plates |
| the croods | first aid basic sciences 2e | disneys frozen party 9oz hotcold cups |
| turbo | usmle step 1 qbook | disneys frozen 7x7 square cakedessert plates |
| cloudy with a chance of meatballs | brs physiology | disneys frozen party printed plastic tablecover |
| hotel transylvania | rapid review pathology with student consult | disneys frozen party 7 square cakedessert plates |
| brave | first aid cases for the usmle step 2 | disney frozen 9 oz paper cups |
| the smurfs | highyield neuroanatomy | frozen invitation and thank you card |
| wreckit ralph | lange pharmacology flash cards third edition | disneys frozen party treat bags |

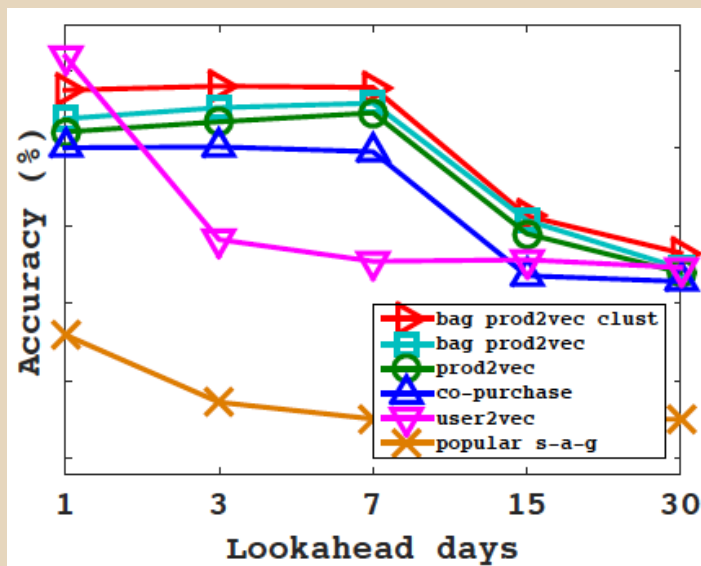
Experiments

- Clustering results in more diverse recommendations
 - Example for product *cressi supernova dry snorkel*

| bagged-prod2vec-topK | bagged-prod2vec-cluster | cluster ID |
|---|---|------------|
| jaws quick spit antifog 1 ounce | cressi neoprene mask strap | 1 |
| cressi neoprene mask strap | cressi frameless mask | |
| cressi frameless mask | akona 2 mm neoprene low cut socks | 2 |
| akona 2 mm neoprene low cut socks | tilos neoprene fin socks | |
| tilos neoprene fin socks | jaws quick spit antifog 1 ounce | 3 |
| cressi scuba diving snorkeling mask snorkel set | aqua sphere kayenne goggle with clear lens black | |
| mares cruise mesh due bag | nikon coolpix aw120 161 mp waterproof camera | 4 |
| us divers island dry snorkel | olympus stylus tg digital camera with 5x optical zoom | |

Recommending predicted products

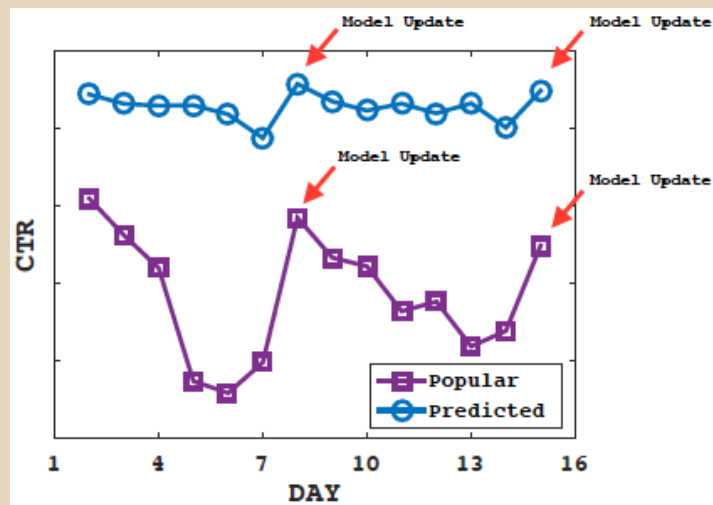
- We fix lookback to 5 days
- Predicted products outperform popular ones



Experiments

- Bucket results

| Metric | Control (5% traffic) | Popular (5% traffic) | Predicted (5% traffic) |
|--------|----------------------|----------------------|------------------------|
| CTR | - | + 8.33% | + 9.81% |
| YR | n/a | - | + 7.63% |



Daily bucket test results

- Implemented in production

Conclusion

- Inbound e-mail data is underutilized
- Significant differences between various user cohorts
- Neural language models can directly be applied to the recommendation problem
 - Don't count, predict!
- Look for our ads during this holiday season!

