Contextual Intent Tracking for Personal Assistants

Yu Sun †, Nicholas Jing Yuan ∗, Yingzi Wang ⋆‡, Xing Xie ‡, Kieran McDonald ∗, Rui Zhang †

† University of Melbourne ‡ Microsoft Research ∗ Microsoft Corporation
⋆ University of Science and Technology of China
{sun.y, rui.zhang}@unimelb.edu.au
{nicholas.yuan, t-yingwa, xing.xie, kieran.mcdonald}@microsoft.com

August 16^{th} 2016 @ KDD

Outline

1. Motivation and Problem Definition
2. KP2 Model
3. Experiments
4. Conclusion and Future Work

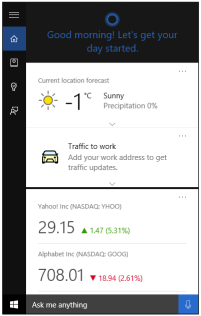
Outline

1. Motivation and Problem Definition
   - Motivation
   - Problem Definition
   - Related Work
2. KP2 Model
   - Data Analysis
   - Model Formulation
   - Optimization Algorithm
3. Experiments
   - Setup and Results
   - If-Do Triggers
4. Conclusion and Future Work

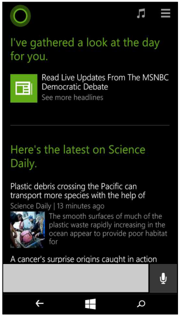
Intelligent Personal Assistants

Microsoft Cortana

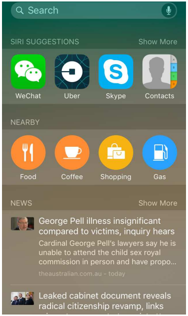
Windows 10



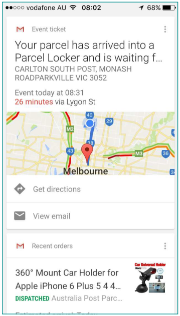
Win Phone



Apple's Siri



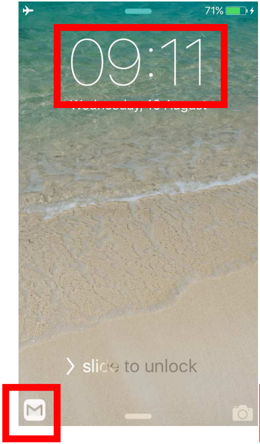
Google Now



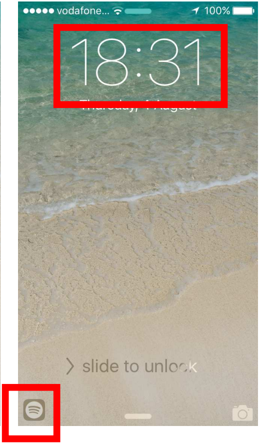
- IPAs proactively recommend various information

Intelligent Personal Assistants

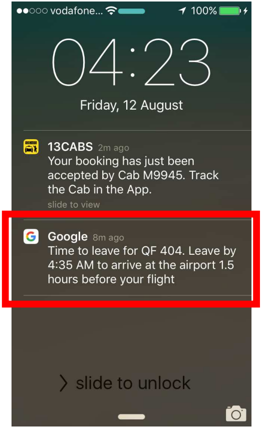
Morning: Email



Evening: Music



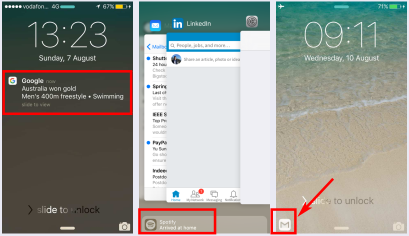
Travel Reminder



What Users Intend to Know/Do

Focused Recommendation/Notification

- Limited display sizes
  $\rightsquigarrow$ show limited content
- Push one notification or remind one task



$\Longrightarrow$

Track Users' Intent

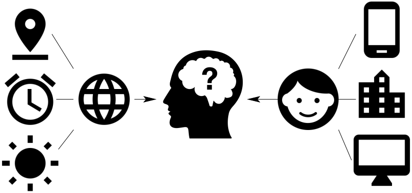
- What users intend **to know**: information intent
- What users intend **to do**: task-completion intent

Intent and Context

- Intent $\leftrightsquigarrow$ Context

Contextual Signals

- external: physical environment, e.g, location, time
- internal: users' activities, e.g., apps, venues



Intent and Context Examples

- to listen to music $\leftsquigarrow$ driving or using browsers
- to check calendar $\leftsquigarrow$ Sunday evenings or at office

- Exploiting contextual signals to track users' intent

Context and Intent Examples

- E.g., contextual signals and intent data shown below
- Data $\rightsquigarrow$ time series and panels

Example of A Panel

Time step	11 a.m.	12 p.m.	1 p.m.	2 p.m.	Now
Chrome	2345	784	0	435	23
Lync	0	1053	0	0	-
Starbucks	0	1251	766	0	0
Fitness First	0	0	0	143	1334
Dist-to-Home	3.45	5.34	10.3	15.7	-
Day-of-Week	5	5	5	5	5
Taxi intent	1	0	0	1	?

- Contextual signals $\rightsquigarrow$ intent

Outline

1 Motivation and Problem Definition
- Motivation
- Problem Definition
- Related Work

2 KP2 Model
- Data Analysis
- Model Formulation
- Optimization Algorithm

3 Experiments
- Setup and Results
- If-Do Triggers

4 Conclusion and Future Work

Intent Tracking

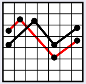
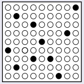
We formally formulate the intent tracking problem as follows.

Definition (Intent Tracking)

- Given
  - a set of M users,
  - a tracking granularity Δ,
  - a type of intent ζ,
  - and context $\mathbf{x}_t^u$ of user u,
- the **intent tracking** problem is to determine
  - whether user u has intent ζ,
  - for every time step t of length Δ

Characteristics of Intent Tracking

Properties

-  i) Real-time tracking
-  ii) Complex correlation
-  iii) Efficient computation
-  iv) Personalized service
-  v) Data sparsity

Explanations

- i) predict intent continuously
- ii) co-occurring and sequential correlation
- iii) computing on mobile devices
- iv) personal friends and assistants
- v) common for recommendation

Outline

1. Motivation and Problem Definition
   - Motivation
   - Problem Definition
   - Related Work
2. KP2 Model
   - Data Analysis
   - Model Formulation
   - Optimization Algorithm
3. Experiments
   - Setup and Results
   - If-Do Triggers
4. Conclusion and Future Work

Related Work

Traditional recommendation models cannot be applied

- assume intent (e.g., to find movies, books) already there
- recommend new items based on similarities

Time-aware recommendation models also cannot be applied

- do not consider other contextual signals besides time
- not suitable for swiftly changing context/intent

Context-aware recommendation models do not work either

- do not take sequential correlation into account
- consider only external context (e.g., time, location)

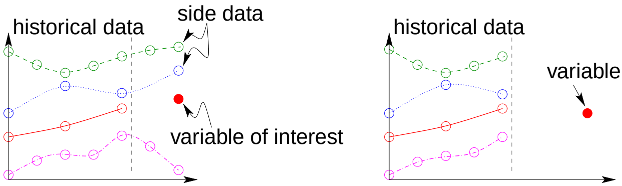
Nowcasting for Intent Tracking

- Take a nowcasting approach to intent tracking

Nowcasting

- First meteorology, then macroeconomics
- Definition: prediction of **current** or very **near future**

- Nowcast v.s. forecast: side data
  - contemporaneous with
  - more frequently available (e.g., industrial output → GDP)



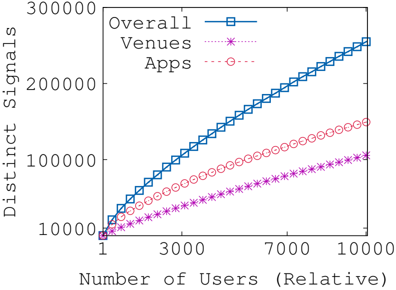
(a) nowcast (b) forecast

Outline

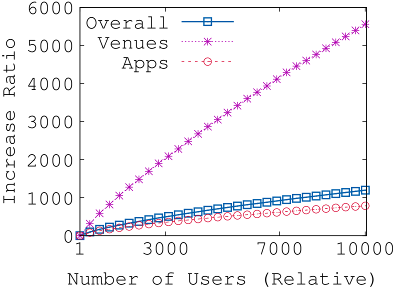
1. Motivation and Problem Definition
   - Motivation
   - Problem Definition
   - Related Work
2. KP2 Model
   - Data Analysis
   - Model Formulation
   - Optimization Algorithm
3. Experiments
   - Setup and Results
   - If-Do Triggers
4. Conclusion and Future Work

Context Explosion

- First analyze the data (personal assistant usage log)
- A booming number of distinct contextual signals



(c) Number of signals



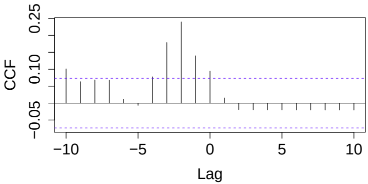
(d) Increase rate

- Num. of distinct signals increases almost linearly
- Users visit different venues (POIs), use similar apps

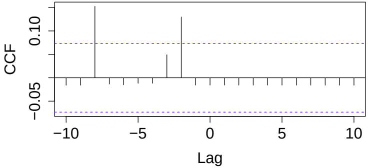
Sequential Correlation

Cross-correlation of contextual signals

$$\mathcal{R}(\tau) = \frac{\sum_t [(\mathbf{x}(t) - \mu_{\mathbf{x}})(\mathbf{y}(t-\tau) - \mu_{\mathbf{y}})]}{\sqrt{\sum_t (\mathbf{x}(t) - \mu_{\mathbf{x}})^2}\sqrt{\sum_t (\mathbf{y}(t-\tau) - \mu_{\mathbf{y}})^2}}$$



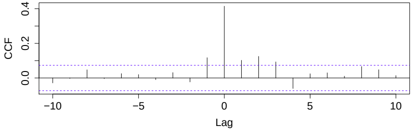
(e) Visit an educational place v.s. Use document editor apps



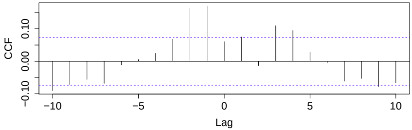
(f) Browse social media apps v.s. Visit a shopping mall

Sequential Correlation

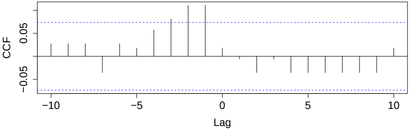
- Cross correlation between context and intent



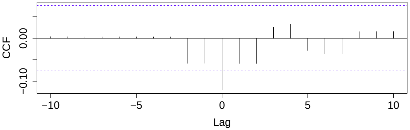
(g) Use browsers v.s. Listen to music



(h) Stay at restaurants v.s. Take taxis



(i) Play video games v.s. Send messages



(j) Distance to office v.s. Check calendar

- Design the KP2 model with these observations

Outline

1. Motivation and Problem Definition
   - Motivation
   - Problem Definition
   - Related Work
2. KP2 Model
   - Data Analysis
   - Model Formulation
   - Optimization Algorithm
3. Experiments
   - Setup and Results
   - If-Do Triggers
4. Conclusion and Future Work

Obtaining Compact Representation

- To avoid estimating a full model, assume $\mathbf{X}^u$ driven by **a few** latent factors (similar to matrix factorization)

$$\mathbf{X}^u \approx \Lambda^u \mathbf{F}^u$$

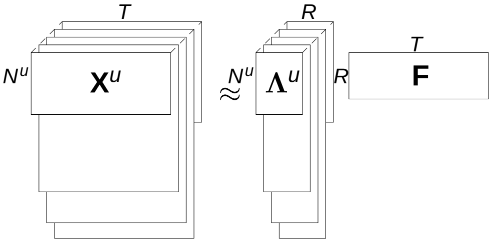
  - $\mathbf{X}^u$ is the panel matrix
  - Λ^u is the loading matrix – roughly every row corresponds to a user's usage (venue visit) habits of a certain app. E.g. use of Google Map has 30% probability when eating, 80% when going to work, 5% when sitting in office, etc.
  - $\mathbf{F}^u$ is the latent factors – roughly every row corresponds to the signal of a category of behavior over time. E.g., the likelihood of eating over time.

Exploit Collaborative Capability

- To exploit the **collaborative** capabilities, assume that these latent factors are shared by all users

$$\mathbf{X}^u \approx \Lambda^u \mathbf{F} \quad \text{for all} \;\; u = 1, 2, \ldots, M \,.$$

- Same as PARAFAC2 tensor decomposition



Linear Dynamic System

- To model the dynamics and sequential correlation, assume latent factors follow **linear** transition
- And latent factors and contextual signals follow the following **linear dynamics system**

$$\begin{cases} \mathbf{x}_t^u = \Lambda^u \mathbf{f}_t + \xi_t^u \,, & t = 1, \ldots, T \\ \mathbf{f}_t = \mathbf{A}^u \mathbf{f}_{t-1} + \omega_t^u \,, & t = 2, \ldots, T \end{cases}$$

  - $\mathbf{A}^u \in \mathbb{R}^{R \times R}$: transition/system matrix
  - ξ_t^u and ω_t^u: mutually independent Gaussian r.v. with covariance Ψ^u and $\mathbf{Q}^u$
- Estimate latent factors and other parameters, and utilize this LDS for efficient intent tracking

PARAFAC2 with Kalman Filter Regularization

- Kalman filter: estimating internal state of LDS
- Jointly optimize PARAFAC2 and LDS
- KP2: use Kalman filter as a regularizer of PARAFAC2

$$\min_{\mathbf{F},\boldsymbol{\Lambda}^u,\mathbf{A}^u} \sum_{u=1}^{M} \|\mathbf{X}^u - \boldsymbol{\Lambda}^u \mathbf{F}\|_F^2 + \frac{\lambda}{2}\left(\|\mathbf{H}^u\mathbf{f} - \mathbf{x}^u\|^2_{\boldsymbol{\Psi}_u^{-1}} + \|\mathbf{G}^u\mathbf{f} - \mathbf{w}^u\|^2_{\mathbf{Q}_u^{-1}}\right)$$

where $\mathbf{H}^u = \mathrm{diag}(\boldsymbol{\Lambda}^u, T)$, $\mathbf{f} = \mathrm{vec}(\mathbf{F})$, $\mathbf{x}^u = \mathrm{vec}(\mathbf{X}^u)$,

$$\mathbf{G}^u = \begin{bmatrix} \mathbf{I} & \mathbf{0} & & \\ -\mathbf{A}^u & \mathbf{I} & \ddots & \\ & \ddots & \ddots & \mathbf{0} \\ & & -\mathbf{A}^u & \mathbf{I} \end{bmatrix}, \quad \mathbf{w}^u = \begin{bmatrix} \mathbf{A}^u\mathbf{f}_0 \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix},$$

and $\boldsymbol{\Psi}_u = \mathrm{diag}(\boldsymbol{\Psi}^u, T)$, $\mathbf{Q}_u = \mathrm{diag}(\mathbf{Q}^u, T)$, $\|\mathbf{a}\|^2_{\mathbf{Y}} = \mathbf{a}'\mathbf{Y}\mathbf{a}$.

Outline

1. Motivation and Problem Definition
   - Motivation
   - Problem Definition
   - Related Work
2. KP2 Model
   - Data Analysis
   - Model Formulation
   - Optimization Algorithm
3. Experiments
   - Setup and Results
   - If-Do Triggers
4. Conclusion and Future Work

Estimating Noise Covariance Matrices

- First, estimate $\mathbf{\Psi}^u$ and $\mathbf{Q}^u$ following nowcasting practice

- PCA: $\mathbf{\Psi}^u \approx \mathrm{Diag}(\mathbf{S}^u - \mathbf{W}^u \mathbf{\Sigma}^u \mathbf{W}^{u\prime})$
  - $\mathbf{S}^u = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t^u \mathbf{x}_t^{u\prime}$
  - $\mathbf{\Sigma}^u \in \mathbb{R}^{R \times R}$ consists of the largest R eigenvalues of $\mathbf{S}^u$
  - $\mathbf{W}^u \in \mathbb{R}^{N^u \times R}$ consists of the corresponding eigenvectors

- VAR: $\mathbf{Q}^u = \frac{1}{T-1}\sum_{t=2}^{T} \tilde{\mathbf{f}}_t^u \tilde{\mathbf{f}}_t^{u\prime} - \tilde{\mathbf{A}}^u \left(\frac{1}{T-1}\sum_{t=2}^{T} \tilde{\mathbf{f}}_{t-1}^u \tilde{\mathbf{f}}_{t-1}^{u\prime} \right) \tilde{\mathbf{A}}^{u\prime}$
  - $\tilde{\mathbf{f}}_t^u = \mathbf{W}^{u\prime} \mathbf{x}_t^u$
  - $\tilde{\mathbf{A}}^u = \sum_{t=2}^{T} \tilde{\mathbf{f}}_t^u \tilde{\mathbf{f}}_{t-1}^{u\prime} \left(\sum_{t=2}^{T} \tilde{\mathbf{f}}_{t-1}^u \tilde{\mathbf{f}}_{t-1}^{u\prime} \right)$

Gradient Computation

- Then, use SGD to estimate $\mathbf{A}^u$, $\mathbf{\Lambda}^u$, and $\mathbf{F}$
- Gradients are computed as follows.

$$\frac{\partial J}{\partial \mathbf{A}^u} = \lambda \sum_{t=2}^{T} (\mathbf{Q}^u)^{-1} (\mathbf{A}^u \mathbf{f}_{t-1} - \mathbf{f}_t) \mathbf{f}_{t-1}' ,$$

$$\frac{\partial J}{\partial \mathbf{\Lambda}^u} = 2 \sum_{t=1}^{T} (\mathbf{\Lambda}^u \mathbf{f}_t - \mathbf{x}_t^u) \mathbf{f}_t' + \lambda \sum_{t=1}^{T} (\mathbf{\Psi}^u)^{-1} (\mathbf{\Lambda}^u \mathbf{f}_t - \mathbf{x}_t^u) \mathbf{f}_t' ,$$

$$\frac{\partial J}{\partial \mathbf{f}_T} = 2 \sum_{u=1}^{M} \mathbf{\Lambda}^{u\prime} (\mathbf{\Lambda}^u \mathbf{f}_T - \mathbf{x}_T^u) + \lambda \sum_{u=1}^{M} \mathbf{\Lambda}^{u\prime} (\mathbf{\Psi}^u)^{-1} (\mathbf{\Lambda}^u \mathbf{f}_T - \mathbf{x}_T^u) - \lambda \sum_{u=1}^{M} (\mathbf{Q}^u)^{-1} (\mathbf{A}^u \mathbf{f}_{T-1} - \mathbf{f}_T) ,$$

$$\frac{\partial J}{\partial \mathbf{f}_t} = 2 \sum_{u=1}^{M} \mathbf{\Lambda}^{u\prime} (\mathbf{\Lambda}^u \mathbf{f}_t - \mathbf{x}_t^u) + \lambda \sum_{u=1}^{M} \mathbf{\Lambda}^{u\prime} (\mathbf{\Psi}^u)^{-1} (\mathbf{\Lambda}^u \mathbf{f}_t - \mathbf{x}_t^u) + \lambda \sum_{u=1}^{M} \mathbf{A}^{u\prime} (\mathbf{Q}^u)^{-1} (\mathbf{A}^u \mathbf{f}_t - \mathbf{f}_{t+1})$$
$$- \lambda \sum_{u=1}^{M} (\mathbf{Q}^u)^{-1} (\mathbf{A}^u \mathbf{f}_{t-1} - \mathbf{f}_t) \quad \text{for} \quad t = 1, \ldots, T-1 .$$

- Use the learned LDS to track each user's real-time intent

Outline

1. Motivation and Problem Definition
   - Motivation
   - Problem Definition
   - Related Work
2. KP2 Model
   - Data Analysis
   - Model Formulation
   - Optimization Algorithm
3. Experiments
   - Setup and Results
   - If-Do Triggers
4. Conclusion and Future Work

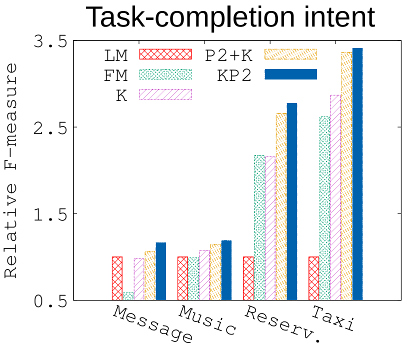
Setup

- Log from a commercial personal assistant for evaluation

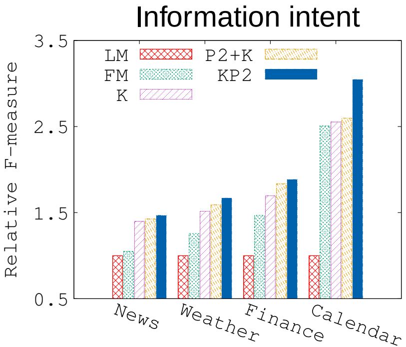
Types of Intent	Types of Log Data	Time	Intent
Task-completion intent	App-launch log	11/02/2015-11/30/2015	send messages, listen to music, make reservations, get taxis
Information intent	Proactive-card log	08/15/2015-09/10/2015	news, weather, finance, calendar

- Contextual signals: used apps, visited venues, distances to home & office, time of day & day of week
- Other configurations
  - $\Delta = 1$, i.e., one hour as the tracking granularity
  - First 3 weeks for training, and then 1 week for testing
  - Mini-batch SGD with a batch size of 30
  - An initial learning rate of 10^{-4} with bold driver adaption

Comparison across Models on F-measure



F1 of LM .2486 .1192 .0094 .0100



F1 of LM .0912 .0214 .0130 .0070

- LM (LambdaMart) is a baseline method
- KP2 (dark blue bin) outperforms several strong methods

Comparison across Models on Hit-ratio

- Hit-ratio: percentage of users who have at least one accurate prediction in one week
- KP2 also outperforms compared methods

Table : Hit-ratio for task-completion intent

Model	Message	Music	Reservation	Taxi
LM	.6633	.4685	.0104	.0186
FM	.7381	.5925	.1354	.1398
K	.9698	.9331	.4896	.4565
P2+K	.9741	.9547	.4583	.5013
KP2	**.9799**	**.9764**	**.5625**	**.5409**

Table : Hit-ratio for information intent

Model	Calendar	Weather	Finance	News
LM	.0056	.0457	.0193	.3748
FM	.0970	.1615	.1273	.6070
K	.4100	.6884	.4790	.9641
P2+K	.4127	.7210	.4874	.9853
KP2	**.4183**	**.7357**	**.5462**	**.9857**

Outline

1. Motivation and Problem Definition
   - Motivation
   - Problem Definition
   - Related Work
2. KP2 Model
   - Data Analysis
   - Model Formulation
   - Optimization Algorithm
3. Experiments
   - Setup and Results
   - If-Do Triggers
4. Conclusion and Future Work

If-Do Triggers Generation

- Investigate the rationale of KP2
- Uncompress latent factors with a decision tree, and generate if-do triggers for some randomly sample users.

	Intent	Triggers
	Message	Between 5:30 p.m. and 7:30 p.m., weekday, arriving at a food and drink venue
	Music	Later than 6:30 p.m., using browsers
	Taxi	Later than 8:30 p.m., weekday, distance to office > 8km, leaving a supermarket
	Reservation	Earlier than 6:30 p.m., Sunday, playing computer games for a long time
	News	Between 6:00 a.m. and 10:00 a.m., Friday, or weekends, distance to office > 10km

- If-do triggers can be easily deployed and computed

Conclusion

Intent Tracking

- Tracking users' intent is important for intelligent personal assistants
  - understand what users intend to know/do
  - provide effective proactive experiences

KP2 Model

- Following the nowcasting framework, the proposed KP2 model outperforms many state-of-the-art methods for contextual intent tracking.

Future Work

We Plan

- to investigate non-linear transition between latent factors with extended Kalman filter or particle filters,
- to investigate more on the underlying rationale of the KP2 model, and explain the model.

Demo Video

An introductory video can be found at
`https://www.youtube.com/watch?v=WaZ0EL3E7XY`

Questions?

Appendix A: Using LDS for Intent Tracking

- First, predict latent factors for next time step

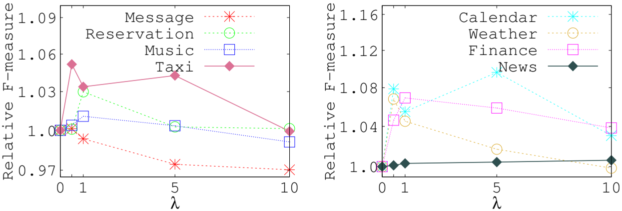
$$\mathbf{f}_t = \mathbf{A}^u \mathbf{f}_{t-1} + \boldsymbol{\omega}_t^u$$

- Correct predicted latent factors with continuously arrived contextual signals using Kalman filter
- Relationship between latent factors and intent is a simple linear function

$$\textit{Intent likelyhood} = \alpha^u + \boldsymbol{\beta}^{u\prime} \mathbf{f}_t$$

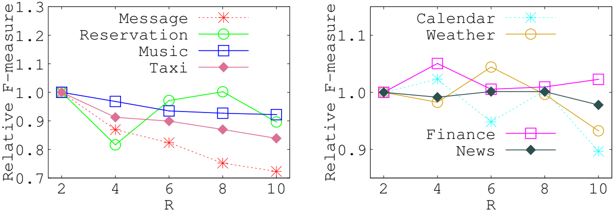
- Parameters α^u and $\boldsymbol{\beta}^u$ can be estimated by regression

Appendix B: Effect of Parameters



(k) Task-completion intent (l) Information Intent

Figure : Relative performance against λ

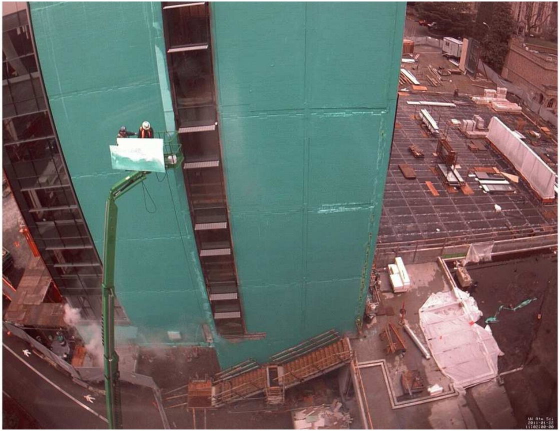


(a) Task-completion intent (b) Information Intent

Figure : Relative performance against R

Appendix C: Why Nowcasting?

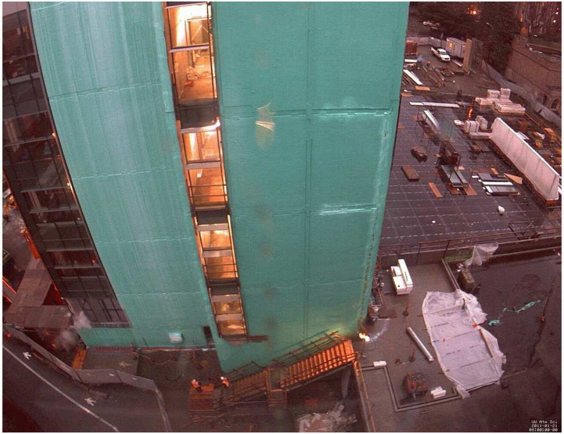
UW Molecular Eng. Bldg: Waterproofing went on one day[1]



[1] Picture from: Cliff Mass, Uni. of Washington, 2011

Appendix C: Why Nowcasting?

Washed off a few hours later[1]



[1] Picture from: Cliff Mass, Uni. of Washington, 2011

Appendix C: Why Nowcasting?

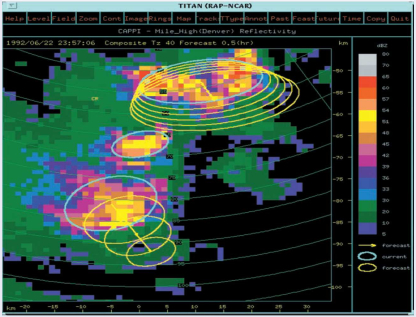
Reapplied the next day. Waste lots of money every year.[1]

[1]Picture from: Cliff Mass, Uni. of Washington, 2011

Appx C: Nowc. in Meteorology and Macroeconomics

Thunderstorm Nowcasting

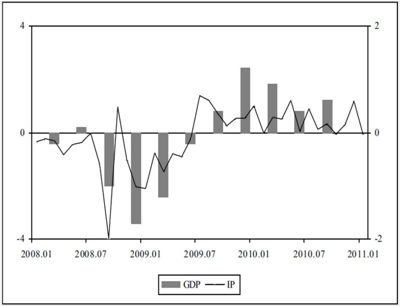
Extrapolating more frequently available atmospheric signals[a]



[a] Picture from J. Wilson et al. *Nowcasting Thunderstorms: A Status Report, 1998*

GDP Nowcasting

Utilizing more frequently available economic signals[b]



[b] Diagram from M. Camacho et al. *Short-Term Forecasting for Empirical Economists: A Survey of the Recently Proposed Algorithms, 2013*

Appendix C: Side-Data Used in Nowcasting

In meteorology: nowcasting weather

- atmospheric conditions from aircraft
- water vapor distributions from GPS receivers
- social media data from Facebook, Twitter, etc.

In macroeconomics: nowcasting GDP

- personal consumption, industrial production
- surveys, financial variables (e.g., interest rates, CPI)
- Google trend data

In data mining: nowcasting rainfall, illness rates

- search engine query log (e.g., Google trend)
- posts in social media (e.g., Twitter)

Appx C: Existing Nowcasting Methods Cannot Apply

Thunderstorm: linear regression with exponential smoothing

- variable of interest quite different from intent

GDP nowcasting: dynamic factor model

- granularity much larger than hours
- macroeconomic variables are non-personalized

Rainfall nowcasting: Bootstrapped LASSO + regression

- cannot address the personalized scenario
- hard to obtain textual features for personalized intent