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 16th 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

Motivation and Problem Definition

Related Work

Contextual Intent Tracking for Personal Assistants

Y. Sun, N. J. Yuan, Y. Wang, X. Xie, K. McDonald and R. Zhang
Intelligent Personal Assistants

Morning: Email
Evening: Music
Travel Reminder

Y. Sun, N. J. Yuan, Y. Wang, X. Xie, K. McDonald and R. Zhang
Contextual Intent Tracking for Personal Assistants
What Users Intend to Know/Do

Focused Recommendation/Notification

- Limited display sizes ➼ show limited content
- Push one notification or remind one task

Track Users’ Intent

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

- Intent $\leftrightarrow$ 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 $\leftrightarrow$ driving or using browsers
- to check calendar $\leftrightarrow$ 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 $\leadsto$ time series and panels

### Example of A Panel

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

- Contextual signals $\leadsto$ 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
We formally formulate the intent tracking problem as follows.

**Definition (Intent Tracking)**

- **Given**
  - a set of $M$ users,
  - a tracking granularity $\Delta$,
  - a type of intent $\zeta$,
  - and context $x_u^t$ of user $u$,

- the **intent tracking** problem is to determine
  - whether user $u$ has intent $\zeta$,
  - for every time step $t$ of length $\Delta$
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 $\rightarrow$ GDP)

![Diagram showing nowcast and forecast](image)
Context Explosion

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

![Graphs showing the increase in distinct signals and increase ratio](image)

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

Cross-correlation of contextual signals

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

(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
To avoid estimating a full model, assume $X^u$ driven by a few latent factors (similar to matrix factorization)

$$X^u \approx \Lambda^u F^u$$

- $X^u$ is the panel matrix
- $\Lambda^u$ 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.
- $F^u$ 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.
To exploit the **collaborative** capabilities, assume that these latent factors are shared by all users

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

Same as PARAFAC2 tensor decomposition
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**

\[
\begin{align*}
    x_t^u &= \Lambda^u f_t + \xi_t^u, & t = 1, \ldots, T \\
    f_t &= A^u f_{t-1} + \omega_t^u, & t = 2, \ldots, T
\end{align*}
\]

- \( A^u \in \mathbb{R}^{R \times R} \): transition/system matrix
- \( \xi_t^u \) and \( \omega_t^u \): mutually independent Gaussian r.v. with covariance \( \Psi^u \) and \( Q^u \)

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

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

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

$$G^u = \begin{bmatrix} I & 0 \\ -A^u & I \\ \vdots & \vdots & \ddots & \vdots \\ -A^u & I \\ 0 & 0 \end{bmatrix}, \quad w^u = \begin{bmatrix} A^uf_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

and $\Psi_u = \text{diag}(\Psi^u, T)$, $Q_u = \text{diag}(Q^u, T)$, $\| a \|_Y^2 = a'Ya$. 

Y. Sun, N. J. Yuan, Y. Wang, X. Xie, K. McDonald and R. Zhang  
Contextual Intent Tracking for Personal Assistants
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
First, estimate $\Psi^u$ and $Q^u$ following nowcasting practice

**PCA:** $\Psi^u \approx \text{Diag}(S^u - W^u \Sigma^u W'^u)$

- $S^u = \frac{1}{T} \sum_{t=1}^{T} x^u_t x'^u_t$
- $\Sigma^u \in \mathbb{R}^{R \times R}$ consists of the largest $R$ eigenvalues of $S^u$
- $W^u \in \mathbb{R}^{N^u \times R}$ consists of the corresponding eigenvectors

**VAR:** $Q^u = \frac{1}{T-1} \sum_{t=2}^{T} \tilde{f}_t \tilde{f}'_t - \tilde{A}^u \left( \frac{1}{T-1} \sum_{t=2}^{T} \tilde{f}_t \tilde{f}'_t \right) \tilde{A}'^u$

- $\tilde{f}_t = W'^u x^u_t$
- $\tilde{A}^u = \sum_{t=2}^{T} \tilde{f}_t \tilde{f}'_{t-1} \left( \sum_{t=2}^{T} \tilde{f}_t \tilde{f}'_{t-1} \right)$
Then, use SGD to estimate $A^u$, $\Lambda^u$, and $F$.

Gradients are computed as follows.

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

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

$$\frac{\partial J}{\partial f_T} = 2 \sum_{u=1}^{M} \Lambda'^u (\Lambda^u f_T - x^u_T) + \lambda \sum_{u=1}^{M} \Lambda'^u (\Psi^u)^{-1}(\Lambda^u f_T - x^u_T) - \lambda \sum_{u=1}^{M} (Q^u)^{-1}(A^u f_{T-1} - f_T),$$

$$\frac{\partial J}{\partial f_t} = 2 \sum_{u=1}^{M} \Lambda'^u (\Lambda^u f_t - x^u_t) + \lambda \sum_{u=1}^{M} \Lambda'^u (\Psi^u)^{-1}(\Lambda^u f_t - x^u_t) + \lambda \sum_{u=1}^{M} A'^u (Q^u)^{-1}(A^u f_t - f_{t+1})$$

$$- \lambda \sum_{u=1}^{M} (Q^u)^{-1}(A^u f_{t-1} - f_t) \text{ for } t = 1, \ldots, T - 1.$$
Setup

Log from a commercial personal assistant for evaluation

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

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

- **Task-completion intent**
  - LM (red) vs. P2+K (purple) vs. KP2 (dark blue)
  - F1 scores:
    - Message: .2486, .1192, .0094
    - Music: .1192, .1192, .1192
    - Reserv.: .0094, .0094, .0094
    - Taxi: .0100, .0100, .0100
  - F1 of LM: .2486

- **Information intent**
  - LM (red) vs. P2+K (purple) vs. KP2 (dark blue)
  - F1 scores:
    - News: .0912, .0214, .0130
    - Weather: .0214, .0214, .0214
    - Finance: .0130, .0130, .0130
    - Calendar: .0070, .0070, .0070
  - F1 of LM: .0912

- 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

<table>
<thead>
<tr>
<th>Model</th>
<th>Message</th>
<th>Music</th>
<th>Reservation</th>
<th>Taxi</th>
</tr>
</thead>
<tbody>
<tr>
<td>LM</td>
<td>.6633</td>
<td>.4685</td>
<td>.0104</td>
<td>.0186</td>
</tr>
<tr>
<td>FM</td>
<td>.7381</td>
<td>.5925</td>
<td>.1354</td>
<td>.1398</td>
</tr>
<tr>
<td>K</td>
<td>.9698</td>
<td>.9331</td>
<td>.4896</td>
<td>.4565</td>
</tr>
<tr>
<td>P2+K</td>
<td>.9741</td>
<td>.9547</td>
<td>.4583</td>
<td>.5013</td>
</tr>
<tr>
<td>KP2</td>
<td>.9799</td>
<td>.9764</td>
<td>.5625</td>
<td>.5409</td>
</tr>
</tbody>
</table>

Table: Hit-ratio for information intent

<table>
<thead>
<tr>
<th>Model</th>
<th>Calendar</th>
<th>Weather</th>
<th>Finance</th>
<th>News</th>
</tr>
</thead>
<tbody>
<tr>
<td>LM</td>
<td>.0056</td>
<td>.0457</td>
<td>.0193</td>
<td>.3748</td>
</tr>
<tr>
<td>FM</td>
<td>.0970</td>
<td>.1615</td>
<td>.1273</td>
<td>.6070</td>
</tr>
<tr>
<td>K</td>
<td>.4100</td>
<td>.6884</td>
<td>.4790</td>
<td>.9641</td>
</tr>
<tr>
<td>P2+K</td>
<td>.4127</td>
<td>.7210</td>
<td>.4874</td>
<td>.9853</td>
</tr>
<tr>
<td>KP2</td>
<td>.4183</td>
<td>.7357</td>
<td>.5462</td>
<td>.9857</td>
</tr>
</tbody>
</table>
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.

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

- If-do triggers can be easily deployed and computed
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?
First, predict latent factors for next time step

\[ f_t = A^u f_{t-1} + \omega_t^u \]

Correct predicted latent factors with continuously arrived contextual signals using Kalman filter

Relationship between latent factors and intent is a simple linear function

\[ \text{Intent likelihood} = \alpha^u + \beta^u f_t \]

Parameters \( \alpha^u \) and \( \beta^u \) can be estimated by regression
Appendix B: Effect of Parameters

Figure: Relative performance against $\lambda$

(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

1Picture from: Cliff Mass, Uni. of Washington, 2011
Appendix C: Why Nowcasting?

Washed off a few hours later

---

1 Picture from: Cliff Mass, Uni. of Washington, 2011
Appendix C: Why Nowcasting?

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

¹Picture from: Cliff Mass, Uni. of Washington, 2011
Thunderstorm Nowcasting

Extrapolating more frequently available atmospheric signals

---

GDP Nowcasting

Utilizing more frequently available economic signals

---

\( ^{a} \) Picture from J. Wilson et al. *Nowcasting Thunderstorms: A Status Report*, 1998

\( ^{b} \) 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

<table>
<thead>
<tr>
<th>In meteorology: nowcasting weather</th>
</tr>
</thead>
<tbody>
<tr>
<td>• atmospheric conditions from aircraft</td>
</tr>
<tr>
<td>• water vapor distributions from GPS receivers</td>
</tr>
<tr>
<td>• social media data from Facebook, Twitter, etc.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>In macroeconomics: nowcasting GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>• personal consumption, industrial production</td>
</tr>
<tr>
<td>• surveys, financial variables (e.g., interest rates, CPI)</td>
</tr>
<tr>
<td>• Google trend data</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>In data mining: nowcasting rainfall, illness rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>• search engine query log (e.g., Google trend)</td>
</tr>
<tr>
<td>• posts in social media (e.g., Twitter)</td>
</tr>
</tbody>
</table>
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