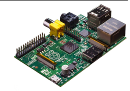
Resource-constrained Graphical Models for App Usage Mining

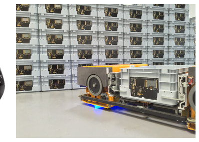
Katharina Morik, Artificial Intelligence,
TU Dortmund University, Germany

technische universität dortmund
SFB 876 - Providing Information by Resource-Constrained Data Analysis

App Usage Mining under Resource Constraints



- Mining app usage data from a population of users
  - Distribution of apps
  - Gender prediction, inspecting the results
- Mining app usage data from an individual user
  - Likeliness of apps over times of a day
    - Spatio-temporal random fields
  - Mining directly on the phone
    - Integer Markov random fields



technische universität dortmund
SFB 876 - Providing Information by Resource-Constrained Data Analysis

App Usage Mining

- Smartphones produce big data.
- Each user generates about 60 gigabytes of data per year. This is big data!
- Data curation is time-consuming.
- Basic Data representation:
  - Did $user_i$ start app_j?

Binary	app_1	...	app_m
usr_1	1		0
usr_n	0		1

technische universität dortmund
SFB 876 - Providing Information by Resource-Constrained Data Analysis

App Usage Mining

- Smartphones produce big data.
- Each user generates about 60 gigabytes of data per year. This is big data!
- Data curation is time-consuming.
- Basic Data representation:
  - Did $user_i$ start app_j?
  - How often did $user_i$ start app_j?
- Probability distributions deliver an overview of the data.

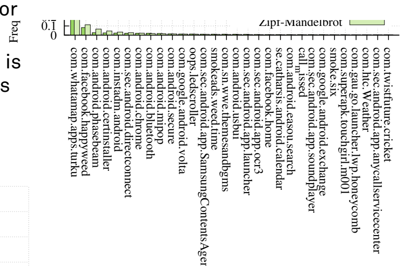
Frequent	app_1	...	app_m
usr_1	#starts		
usr_n			#starts(n,m)

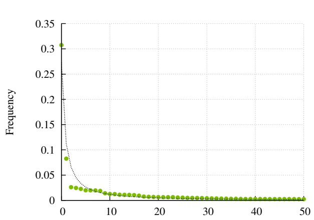
technische universität dortmund
SFB 876 - Providing Information by Resource-Constrained Data Analysis

Probability Distributions of Apps

App usage follows the Zipf or Mandelbrot distribution:

- The probability of an app is inverse proportional to its rank: p(n) = 1/n





technische universität dortmund
SFB 876 - Providing Information by Resource-Constrained Data Analysis

App Usage Mining

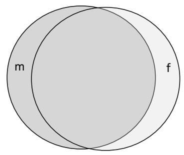
- Smartphones produce big data.
- Each user generates about 60 gigabytes of data per year. This is big data!
- Data curation is time-consuming.
- Basic Data representation:
  - Did $user_i$ start app_j?
  - How often did $user_i$ start app_j?
- Probability distributions deliver an overview of the data.
- Additional attributes of users can be predicted.

Binary	app_1	...	app_m	target
usr_1	0		1	
usr_n	1		0	

Predicting Attributes -- Gender

- Menthal data from Alexander Markowetz https://menthal.org
- 42,482 men and 31,700 women were logged.
- Top 99 apps of women and top 99 apps of men overlap by 82 apps.
- We can predict the gender by app usage data using logistic regression and the L1 norm: accuracy 91.2%.

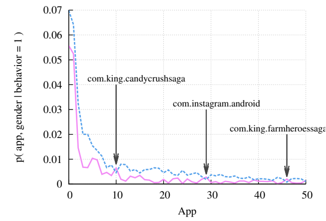
Binary	app_1	...	app_m	gender
usr_1	0		1	f
usr_n	1		0	m



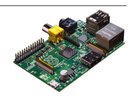
Inspecting the Results -- Gender

- Menthal data from Alexander Markowetz https://menthal.org
- 42,482 men and 31,700 women were logged.
- Ranking apps according to their frequency, compare rankings of men and women.
- Largest differences:
  - ClashOfClans men 80, women 219
  - FarmHeroes men 249, women 78
  - GoogleDocs men 52, women 92

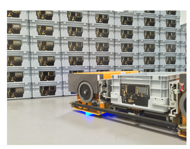
Frequent	app_1	...	app_m
usr_1	#starts		
usr_n			#starts(n,m)



App Usage Mining under Resource Constraints



- Mining app usage data from a population of users
  - Distribution of apps
  - Gender prediction, inspecting the results
- Mining app usage data from an individual user
  - Likeliness of apps over times of a day
    - Spatio-temporal random fields
  - Mining directly on the phone
    - Integer Markov random fields



App Usage Mining

- Smartphones produce big data.
- Each user generates about 60 gigabytes of data per year. This is big data!
- Data curation is time-consuming.
- Individual data:
  - Where did $user_i$ use app_j? Include only places that occur more than t times.

Usr_n	app_1	...	app_m

Usr_i	app_1	...	app_m
$place_1$	#starts		

Usr_1	app_1	...	app_m
$place_1$	#starts		
$place_n$			#starts(n, m)

App Usage Mining

- Smartphones produce big data.
- Each user generates about 60 gigabytes of data per year. This is big data!
- Data curation is time-consuming.
- Individual data:
  - Where did $user_i$ use app_j? Include only places that occur more than t times.
  - When did $user_i$ use app_j? Discretized time of day: 30 minutes

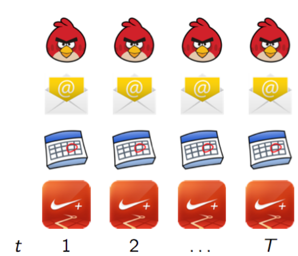
Usr_n	app_1	...	app_m

Usr_i	app_1	...	app_m
0	#starts		

Usr_1	app_1	...	app_m
0	#starts		
1			
...			
48			#starts(n, m)

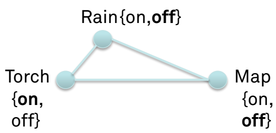
Spatio-temporal Models for App Usage Mining

- Spatial graph with Apps as nodes and edges based on covariance matrix.
- Windows of 30 minutes arranged for a day.
- Likelihood of an app at a particular time of day.
- Data from 8 users, public: http://sfb876.tu-dortmund.de/auto?self=$e675o3goow



Probabilistic Graphical Models

Observation
- x_1:(on,off,off)
- $\phi(x_1)$ = (1,0,0,1,0,1,1,0,0,1,0,0,0,0,1)
- d = 15



Graph G=(V,E)

$\phi(\mathbf{x})$: (/* Nodes*/
1. **on**, /*dom(torch)*/
2. off,
3. on, /*dom(rain)*/
4. **off**,
5. map, /*dom(map)*/
6. **off**,

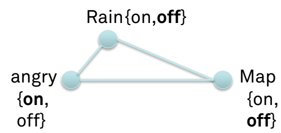
/*Edges*/
1. **on, off**, /* edge torch-rain*/
2. on,on,
3. off,off,
4. **on,off**, /*edge torch-map*/
5. on, on,
6. off, off,
7. on, off, /*edge rain-map*/
8. on, on,
9. **off, off**)

)

Probabilistic Graphical Models

Observation
- x_1:(on,off,off)
- $\phi(x_1)$ = (1,0,0,1,0,1,1,0,0,1,0,0,0,0,1)
- d = 15

Learning
- Likelihood of joint realizations
  $$p_\theta(\vec{x}) = \exp\left(\left\langle \vec{\theta}, \phi(\vec{x}) \right\rangle\right) - A(\vec{x})$$
- Maximum a posteriori prediction of an unobserved node H, given observed nodes O
  $$\vec{x}_H^* = \arg\max\, p_\theta\left(\vec{x}_H \,|\, \vec{x}_O\right)$$



Graph G=(V,E)

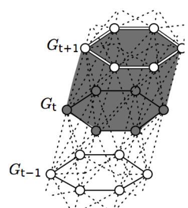
$\phi(\mathbf{x})$: (/* Nodes*/
1. **on**, /*dom(angry)*/
2. off,
3. on, /*dom(rain)*/
4. **off**,
5. map, /*dom(map)*/
6. **off**,

/*Edges*/
1. **on, off**, /* edge angry-rain*/
2. on,on,
3. off,off,
4. **on,off**, /*edge angry-map*/
5. on, on,
6. off, off,
7. on, off, /*edge rain-map*/
8. on, on,
9. **off, off**)

)

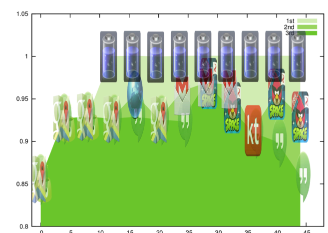
Spatio-temporal Random Fields

- The spatio-temporal graph is trained to predict each node's maximum a posteriori probability with the marginal probabilities.
- Generative model predicting all nodes.
  Maximum a posteriori prediction
  of an unobserved node H, given observed nodes O.
- The learned model answers diverse questions – you only need to choose H.



STRF modeling usage of apps on Android phones

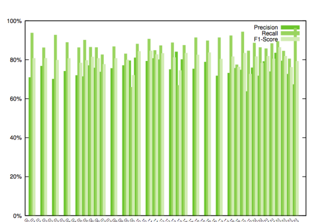
- Probability of an app being switched on
- Ranking according probability
- Show the first three ranks with their probability.

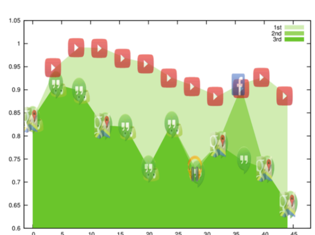


User G keeps battery on, plays Angry Bird, and looks up the map and communicates using hangout.

The Google addict

YouTube, hangout, music, maps

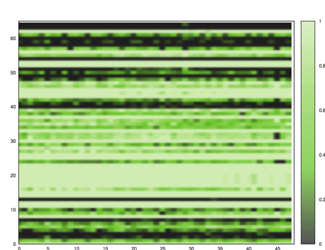




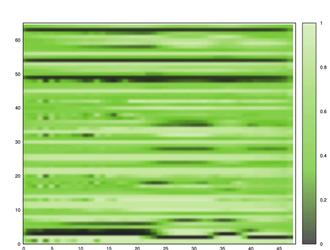
Facebook in the evening

Data record and STRF model of user G

70 apps (excl. OS)



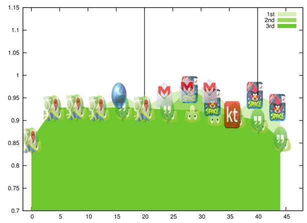
Time of day



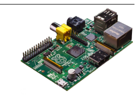
Time of day

Warn the user based on the prediction

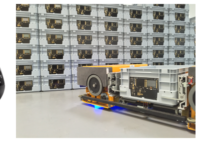
- Given the likelihood of apps, the energy consumption of these apps,
- Calculate when the battery will be empty.
- Here:
  battery was full, when the picture starts,
  already at lunch time, we predict the battery being empty at dinner time,
  if not loaded.



App Usage Mining under Resource Constraints



- Mining app usage data from a population of users
  - Distribution of apps
  - Gender prediction, inspecting the results
- Mining app usage data from an individual user
  - Likeliness of apps over times of a day
    - Spatio-temporal random fields
  - Mining directly on the phone
    - Integer Markov random fields



Graphical models **on** resource-restricted processors

- Spatio-Temporal Random Fields on the phone. but
- Power consumption is restricted!
- "The most obvious technique to conserve power is to reduce the number of cycles it takes to complete a workload."
  (Intel 64, IA-32 architectures optimization reference manual, guidelines for extending battery life)
- Restrict the parameter space of the Markov Random Field $\theta \in \{0,1,\dots,K\} \subset \mathbb{N}$

	Sandy Bridge		ARM 11	
	Real	Int	Real	Int
+	3	**1**	8	**1**
*	5	**3**	8	4-5
/	14	13-15	19	-
Bit shift	-	3	-	2

Clock cycles for arithmetics on different processors: Real vs. integer.

Parameter space transformation

- Graph model tree-structured
- Transform the parameter space:
  $\eta_i(\theta) = \theta_i \ln 2$

$MRF:$

$$p(\vec{x}) = \frac{1}{Z(\theta)} \exp\left(\sum_i \theta_i \phi_i(\vec{x})\right) = \exp\left[\langle \theta, \phi(\vec{x}) \rangle - A(\theta)\right]$$

$IntegerMRF:$

$$p(\vec{x}) = \exp\left[\langle \eta(\theta), \phi(\vec{x}) \rangle\right] = 2^{\left[\langle \theta, \phi(x) \rangle - A(\eta(\theta))\right]} = \frac{2^{\langle \theta, \phi(x) \rangle}}{\sum_{y \in \mathcal{X}} 2^{\langle \theta, \phi(y) \rangle}}$$

Integer belief propagation

- Simply replacing the exp(.) by $2^{(.)}$ is not sufficient
  - Overflows are normally avoided by normalization.
  - Normalization is impossible in integer division.
- Magnitude of messages corresponds to probability
  - Use the length of each message
  - Bit-length is similar to log

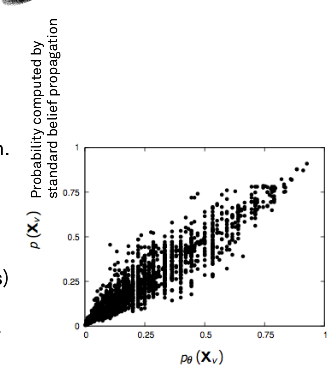
$$m_{v \to u}(y) = \sum_{x \in \mathcal{X}_v} \exp\left(\theta_{vu=xy} + \theta_{v=x}\right) \prod_{w \in N_v - \{u\}} m_{wu}(x)$$

$$\tilde{m}_{v \to u}(y) = \sum_{x \in \mathcal{X}_v} 2^{\left(\theta_{vu=xy} + \theta_{v=x}\right)} \prod_{w \in N_v - \{u\}} \tilde{m}_{wu}(x)$$

$$\beta_{v \to u}(y) = \max_{x \in \mathcal{X}_v} \theta_{vu=xy} + \theta_{v=x} + \sum_{w \in N_v - \{u\}} \beta_{wu}(x)$$

Discretized probability space

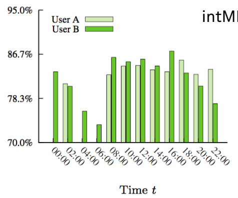
- Belief propagation is now bit-length propagation, i.e. the MAP and marginals are computed using the bit-length.
- The approximation error depends on the number of neighboring nodes and the space of states.
- Some true probabilities (y axis) cannot be expressed by the integer approximation (x axis).

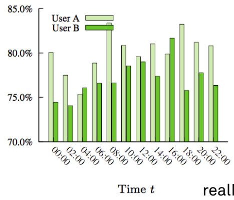


Probability computed by bit-length approximation

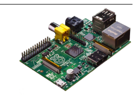
Experiments

- User A
  - 2064 vertices
    (app at time = {on, off})
  - 23368 spatio-temporal edges
  - 38 training instances,
    32 test instances
- User B
  - 2064 vertices
  - 14998 spatio-temporal edges
  - 125 training instances
    119 test instances
- Comparing integer MRF and real MRF
  - Average of 50 trainings
  - Runtime: intMRF 156 (136) s
    realMRF 466 (419) s
  - Predicting apps at t+1,
    given apps at t



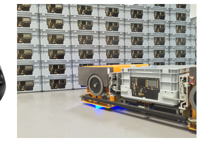


App Usage Mining under Resource Constraints



- Mining app usage data from a population of users (global model)
  - Distribution of apps
  - Gender prediction,
    inspecting the results
- Mining app usage data from an individual user (local model)
  - Likeliness of apps over times of a day
  - Maximum a posteriori prediction
    of an unobserved node H,
    given observed nodes O.
  - Mining directly on the phone.
- Code and local data available at

http://sfb876.tu-dortmund.de/index.html



References

- Nico Piatkowski, Sangkyun Lee, Katharina Morik (2016) Integer Undirected Graphical Models for Resource-Constrained Systems. In: Neurocomputing, 173:1, 9 – 23
  (Now including loopy graphical structures and both, MRF and CRF)
- Nico Piatkowski, Katharina Morik (2016) Stochastic Discrete Clenshaw-Curtis Quadrature. In: 33rd Int. Conference on Machine Learning, Proceedings JMLR
- Jochen Streicher, Nico Piatkowski, Katharina Morik, Olaf Spinczyk (2013) Open Smartphone Data for Mobility and Utilization Analysis in Ubiquitous Environments. In: Atzmüller, Scholz (eds) 4th Int. Workshop on Mining Ubiquitous and Social Environments (MUSE)
- Nico Piatkowski, Sangkyun Lee, Katharina Morik (2013) Spatio-temporal Random Fields: Compressible Representation and Distributed Estimation. In: Machine Learning Journal, 93:1, 115 – 140.
- Thomas Liebig, Nico Piatkowski, Christian Bockermann, Katharina Morik (2014) Predictive Trip Planning – Smart Routing in Smart Cities. In: Mining Urban Data Workshop at 17th Intern. Conf. on Extending Database Technology
- Nico Piatkowski, Sangkyun Lee, Katharina Morik (2014) The Integer Approximation of Undirected Graphical Models. In: 3rd International Conference on Pattern Recognition Applications and Methods, 296 – 304.
  (First approach: restricted to tree structured models, direct estimate of parameters).