360° User Profile Learning

from Multiple Social Networks

for Wellness and Urban Mobility Applications

by Aleksandr Farseev

Supervised by Prof. Chua Tat-Seng

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Agenda

Brief summary of the talk...

1 Motivation and Data

1. Definition and Motivation
2. NUS-MSS Dataset
3. NUS-SENSE Dataset
4. Framework Overview

2 Individual User Profiling

1. Multi-modal data representation
2. Multi-Source Ensemble Learning for Demographic Profiling
3. Learning From Temporal Data for Personality Profiling
4. Sensor data representation
5. Multi-Task Learning for Wellness Profiling

3 Group User Profiling

1. Multi-modal data representation
2. Clustering on Multi-Layer Graphs for Group Profiling
3. Cross-Domain Recommendation as Implicit Evaluation Approach
4. Real-World Application: bBridge Analytics Platform

4 Conclusion

1. Conclusion
2. Future Work

Novelty ☆

Answer ✓

> And “smartness” is not a measure of how advanced or complex the technology being adopted is, but how well a society uses technology to solve its problems and address existential challenges. Citizens are ultimately at the heart of our Smart Nation vision…

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Smart Nation: 5 Key Domains

Where Wellness and Mobility applications are the backbone

Smart Nation SINGAPORE

Mobility

1. Transport
2. Home & environment
3. Business productivity
4. Health and enabled aging
5. Public sector services

Mobility

Wellness

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

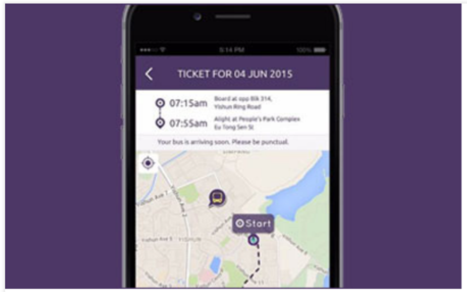
Smart Nation: Dominance of Wellness and Mobility Applications

They indeed are related to our everyday living

Mobility

Smart Nation
SINGAPORE

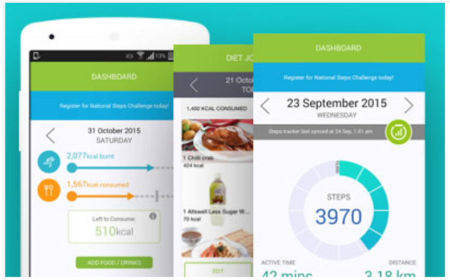
Wellness



Beeline

Suggest routes, and book a direct ride to y
transport, Beeline crowdsources requests a
up with new options for commuters. Sugge
to your destination.

GET IT ON Google Play | Download on the App Store



Healthy 365

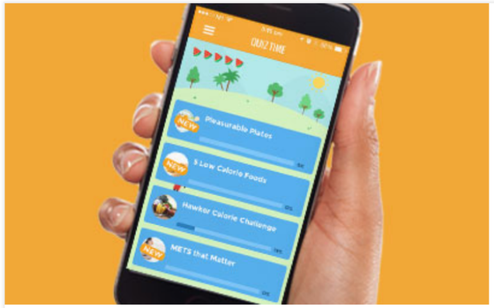
Health & diet tracking app for a healthier life
calculate the corresponding calories burned
tracking mobile app also helps to track your
calories consumed.

GET IT ON Google Play | Download on the App Store

OneHMap

Evolution of our city's landscape! An open g
photographs of Singapore to help trace the
the geographical history of Singapore by ge
photographs of Singapore, and add to the s
of yesteryears.

GET IT ON Google Play | Download on the App Store



HealthHub

Personalised health records at your fingert
your fingertips through this one-stop portal
related content, deals, rewards and e-serv

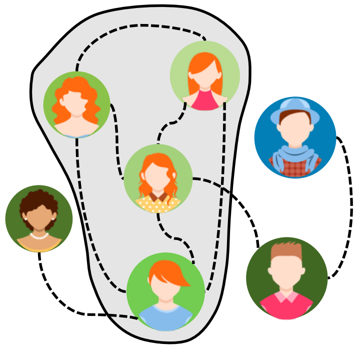
GET IT ON Google Play | Download on the App Store

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

User Profiling for Wellness and Mobility Domains

On relation between Wellness Mobility and 360 ° User Profiling

User Profile



User Communities

Mobility

Individual User Attributes

Wellness

Smart Nation
SINGAPORE

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Multiple social networks describe users from multiple views

Some facts about social networks…

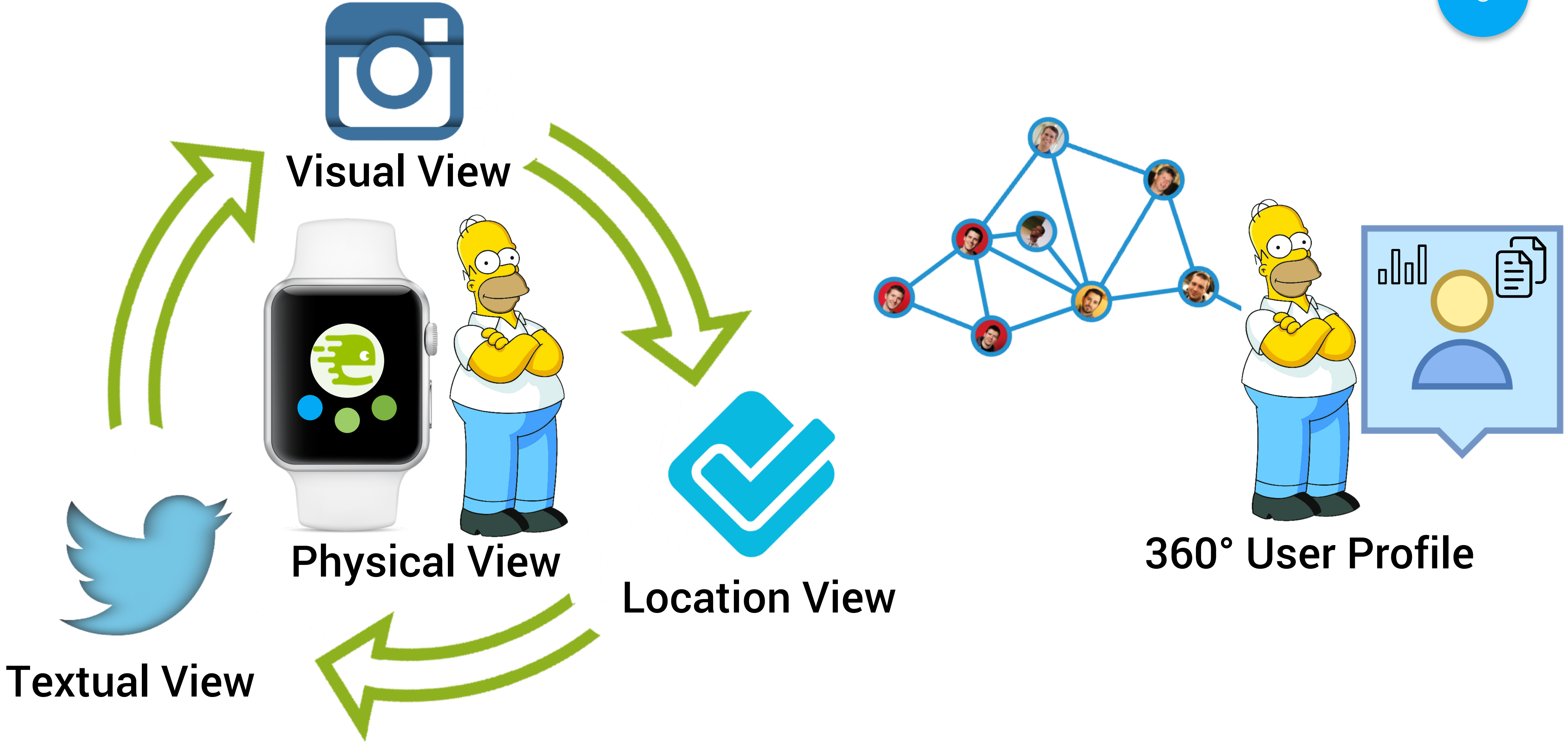
More than 50% of online-active adults use more than three social networks in their daily life*

*According Paw Research Internet Project's Social Media Update 2014 (www.pewinternet.org/fact-sheets/social-networking-fact-sheet/)

E-Mail: farseev@u.nus.edu l Website: http://farseev.com

Different data modalities describe users from multiple views

Indeed, they are:



E-Mail: farseev@u.nus.edu | Website: http://farseev.com

User profiling in our works

Those attributes that we infer

360° User Profile

Group Profile

Individual Profile

User Communities

Wellness profile

Identity profile

Diabetes

Asthma

Obesity

Age

Gender

Personality

Profiling from ML Point of View

Single Source Learning

Classification

N known samples

Source 1

Features $\mathbf{X} \in \mathbb{R}^{N \times K}$,
Ground Truth $\mathbf{y} \in \mathbb{R}^{N \times 1}$

Learn $\boldsymbol{f}(\mathbf{X} \,|\boldsymbol{\Theta})$

M unknown samples

Features $\widehat{\mathbf{X}} \in \mathbb{R}^{M \times K}$,

Predict $\hat{\mathbf{y}} = \boldsymbol{f}(\widehat{\mathbf{X}}|\boldsymbol{\Theta})$

Predictions $\hat{y} \in \mathbb{R}^{M \times 1}$,

Clustering

N known samples

Source 1

Features $\mathbf{X} \in \mathbb{R}^{N \times K}$,

Learn $\boldsymbol{f}(\mathbf{X})$

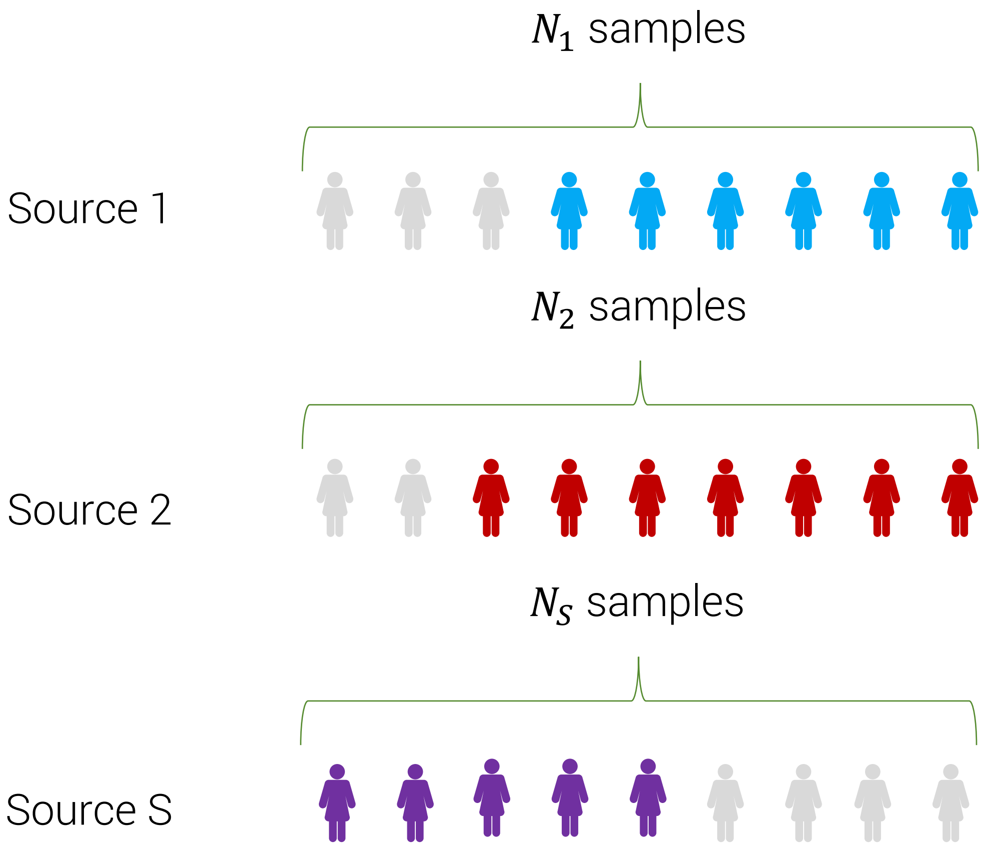
N known samples

Assign $\hat{\boldsymbol{y}} = \boldsymbol{f}(\mathbf{X})$

Cluster assignments $\hat{\mathbf{y}} \in \mathbb{R}^{N \times 1}$,

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Multi-Source Incomplete Learning



How to?

Learn $f(\mathbf{X}\,|\,\Theta)$

Learn $f(\mathbf{X})$

by

1. Incorporating all information from incomplete sources

2. Considering the differences between sources

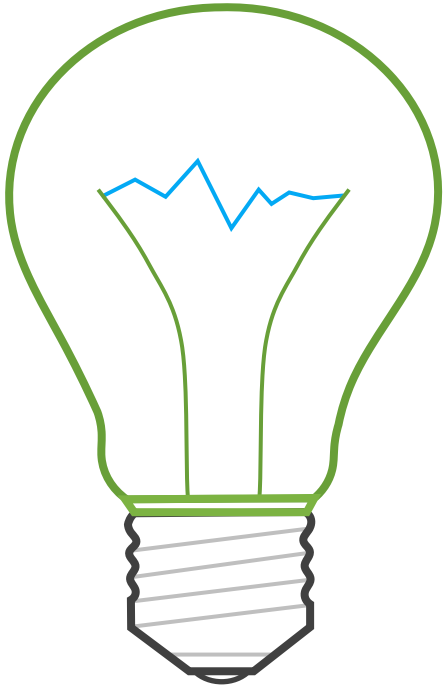
E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Key Contributions of the Thesis

From what it all started…

1 Multi-View Individual User Attribute Inference Approach

2 Multi-View User Community Detection Approach



3 Multi-Network Data Gathering Methodology

4 Two Benchmark Datasets

Key Contributions

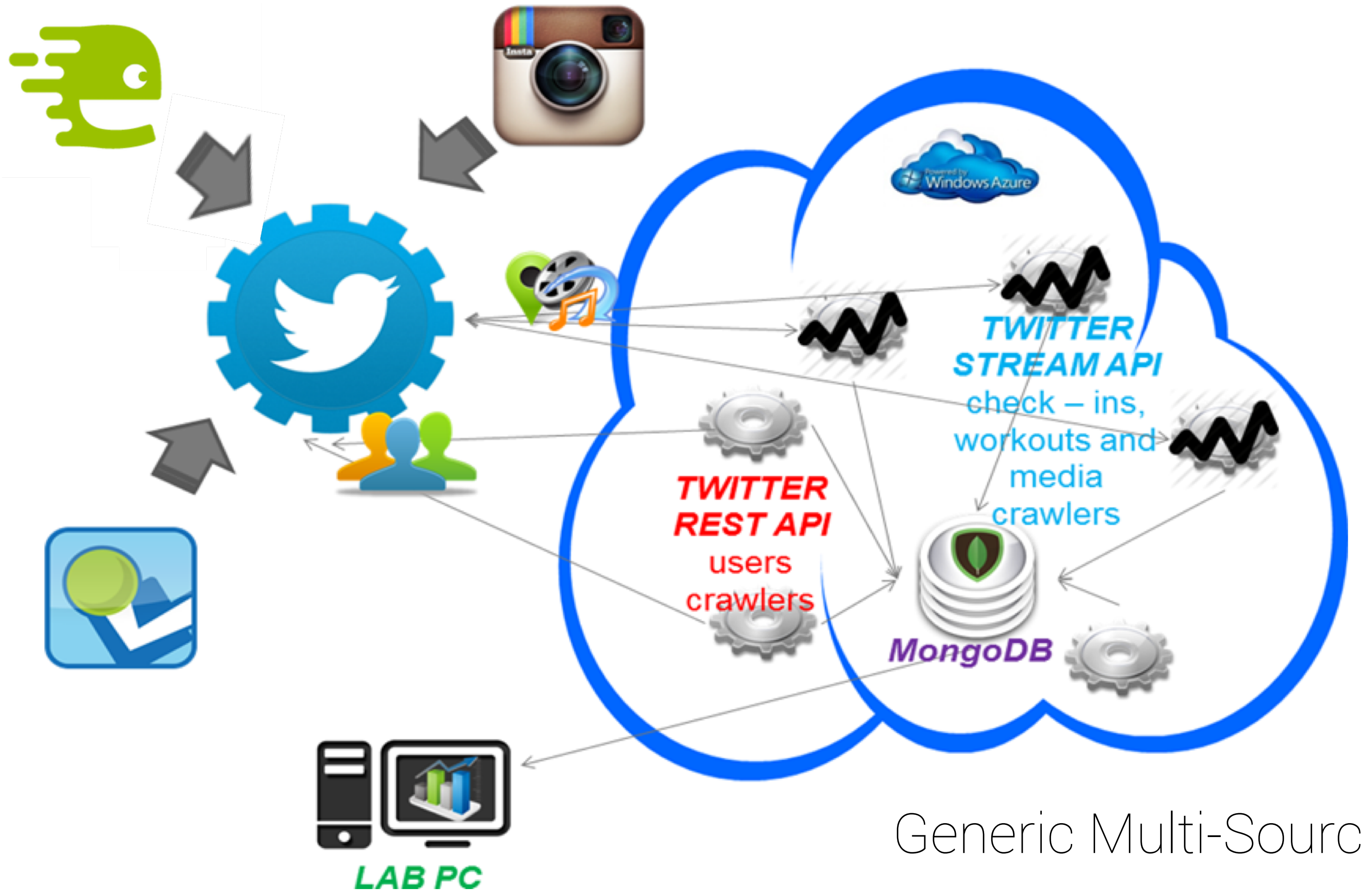
Data for User Profiling

*A. Farseev, N. Liqiang, M. Akbari, and T.-S. Chua. **Harvesting multiple sources for user profile learning: a Big data study.** ACM International Conference on Multimedia Retrieval (ICMR). China. June 23-26, 2015.

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Data Gathering And Simultaneous Cross-Network Account Mapping

About finding the same users in different social networks…



Twitter plays a role of a "sink" for multi-modal data from other social networks.

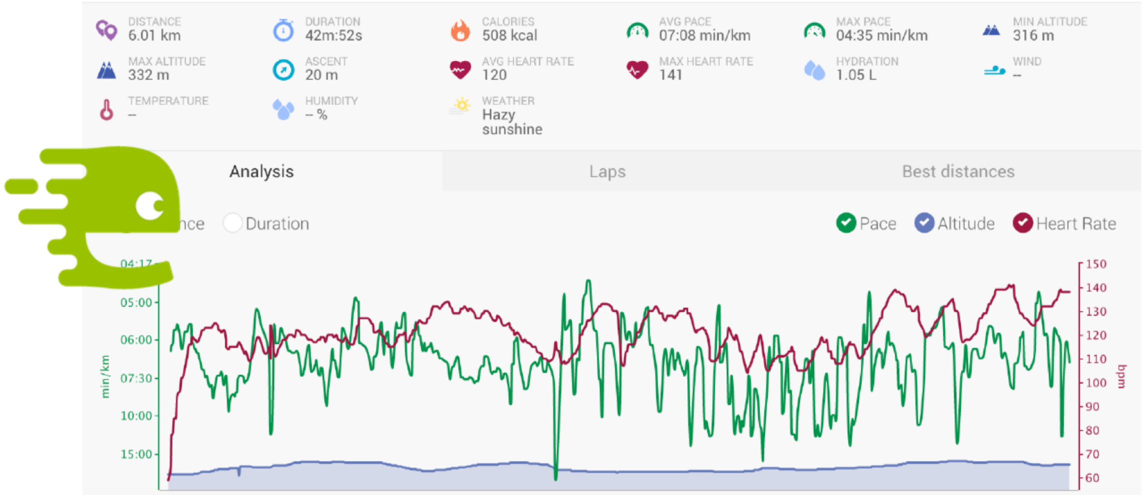
Cross-network ambiguity is resolved after collection of the first cross-network post.

Generic Multi-Source Data Gathering Approach

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Cross-Network Account Mapping: Example

How to grab Alex's personal data…



Cross-network post

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Data Representation: Summary

All data types together

TEXT Features:

Linguistic features: LIWC; Latent Topics

Heuristic features: Writing behavior

Location Features:

Location Semantics: Venue Category Distribution

Mobility Features: Areas of Interest (AOI)

Image Features

Image Concept Distribution (Image Net)

Sensor Features

Exercise statistics + sport types + spectrum

Our released large multi-source multi-modal datasets

NUS-SENSE
http://nussense.azurewebsites.net

Location	#users	#tweets	#check-ins	#images	#check-ins
Worldwide	5,375	16,763,310	19,743	48,137	140,926

NUS-MSS
http://nusmss.azurewebsites.net

Location	#users	#tweets	#check-ins	#images
Singapore	7,023	11,732,489	366,268	263,530
London	5,503	2,973,162	127,276	65,088
New York	7,957	5,263,630	304,493	230,752

Data was voluntarily publicly released by Twitter users and collected via official Twitter API
Datasets are released in a form of features thus user privacy is not affected.

Two Large Multi-Source Social Media & Sensor Datasets

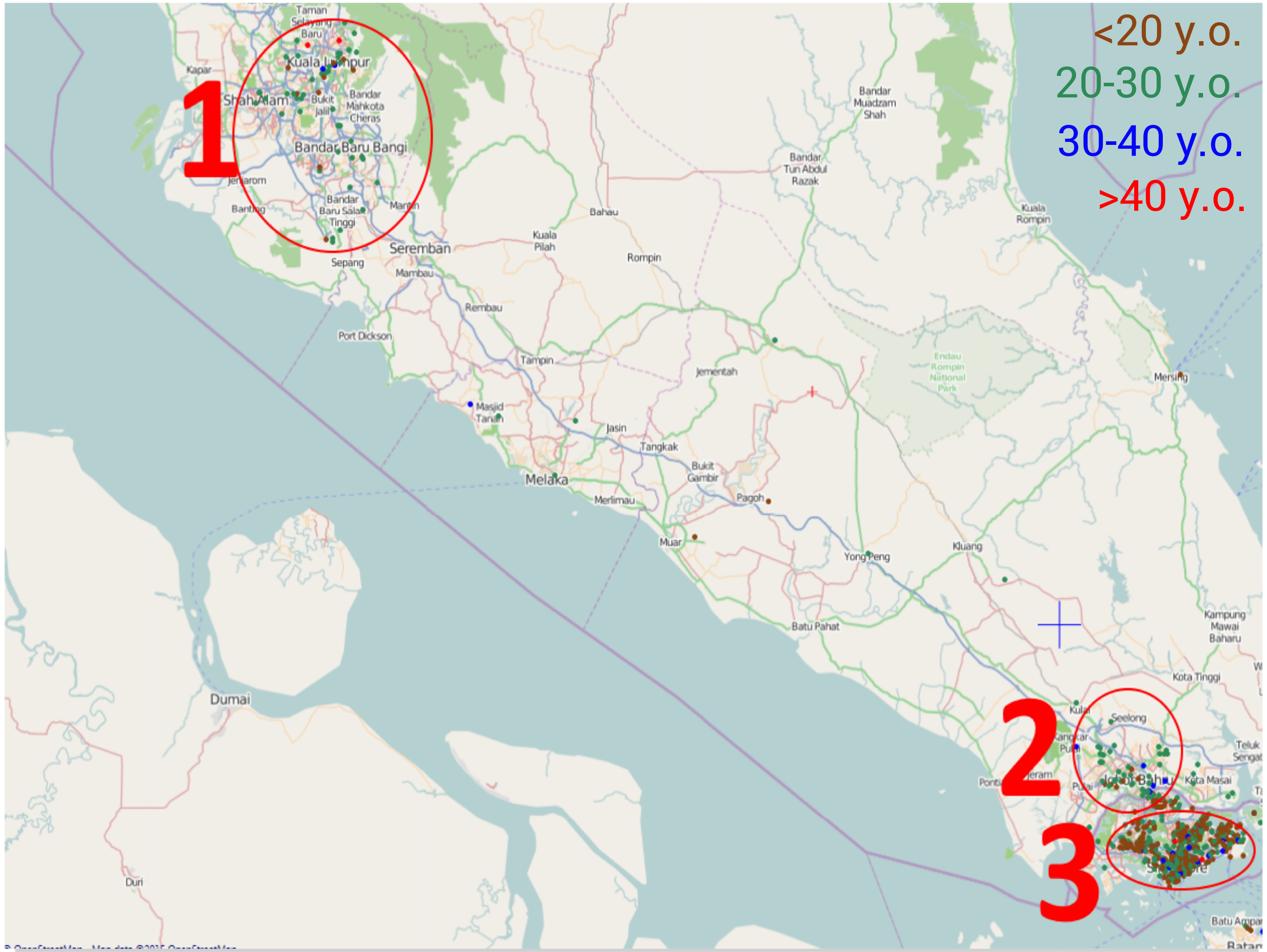
Individual Multi-View Learning

Part I: Demographic Profiling

*A. Farseev, N. Liqiang, M. Akbari, and T.-S. Chua. **Harvesting multiple sources for user profile learning: a Big data study.** ACM International Conference on Multimedia Retrieval (ICMR). China. June 23-26, 2015.

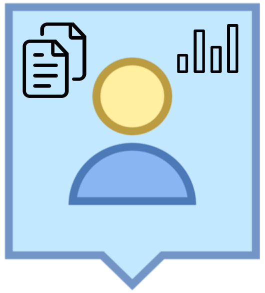
E-Mail: farseev@u.nus.edu | Website: http://farseev.com

On importance of demographics and its relation to mobility



On cross-domain importance of basic demographic attributes

What we can do if we know Homer’s age?



Age: 40
Gender: Male

Marketing
Trade are analysis
Demography and interest - based marketing

Tent to stay at home, visit local pubs and shopping mall daily.

Wellness
Health group prediction
Lifestyle recommendation

Medium overweight, potential hypertonia and diabetes.

Advertisement
Demography and interest - based personalized advertisement

Advertise new Beer brand and new car models.

Assistance
Activity recommendation,
Venue recommendation,
Etc.

Morning excursive with medium intensity.

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Idea I

From what it all started…

Input One

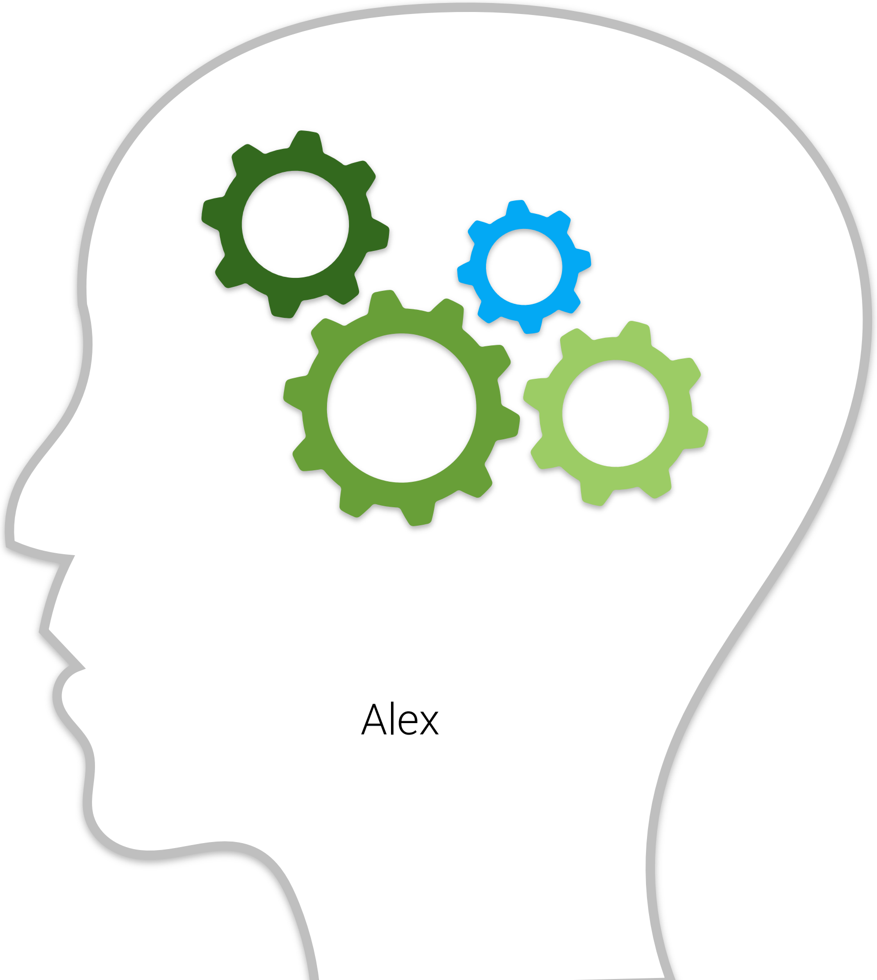
1

Multi-Source Multi-Modal Data Describes Users From Multiple Views

Second Input

2

Demographic Profiling is possible Based On Textual Data



Idea

Perform Demographic Profiling Based On Multiple Social Networks

Research Questions

Question One

1

Is it possible to boost supervised machine learning for individual user profiling performance by incorporating multi-modal data from multiple social networks?

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Related Research Directions: Individual User Profiling

Highest Gender Prediction
curacy: 0.8592

Conventional urces
Manually-c d small-scale d voice reco d-written te .

SMALL SCALE

De phic Profiling
PAN 15', 16', 17'
Mostly n odal data processing from Microblogs, B tc.)

MONO-MODAL

Other Profiling D
Mostly
m ata
ng.

MONO-SOURCE

urce Profiling
Utilize urces, but usually sources to be co which is unrealistic.

REQ. COMP. DATA

Sensor data utilization
Mostly mon solve han gnition, which directly related.

MONO-MODAL

Physical Attrib erence
Mostly odal.
Limi ber of ons to the field..

MONO-MODAL

Usually, this problem of individual user profiling is treated as a supervised learning task and performed based on mono-source mono-modal datasets. Research on multi-source data integration is sparse and limited by unrealistic assumptions. Wellness attributes inference was not yet comprehensively studied.

Contributions…

The First Work On Multi-Source Individual User Profiling via Late Fusion

01

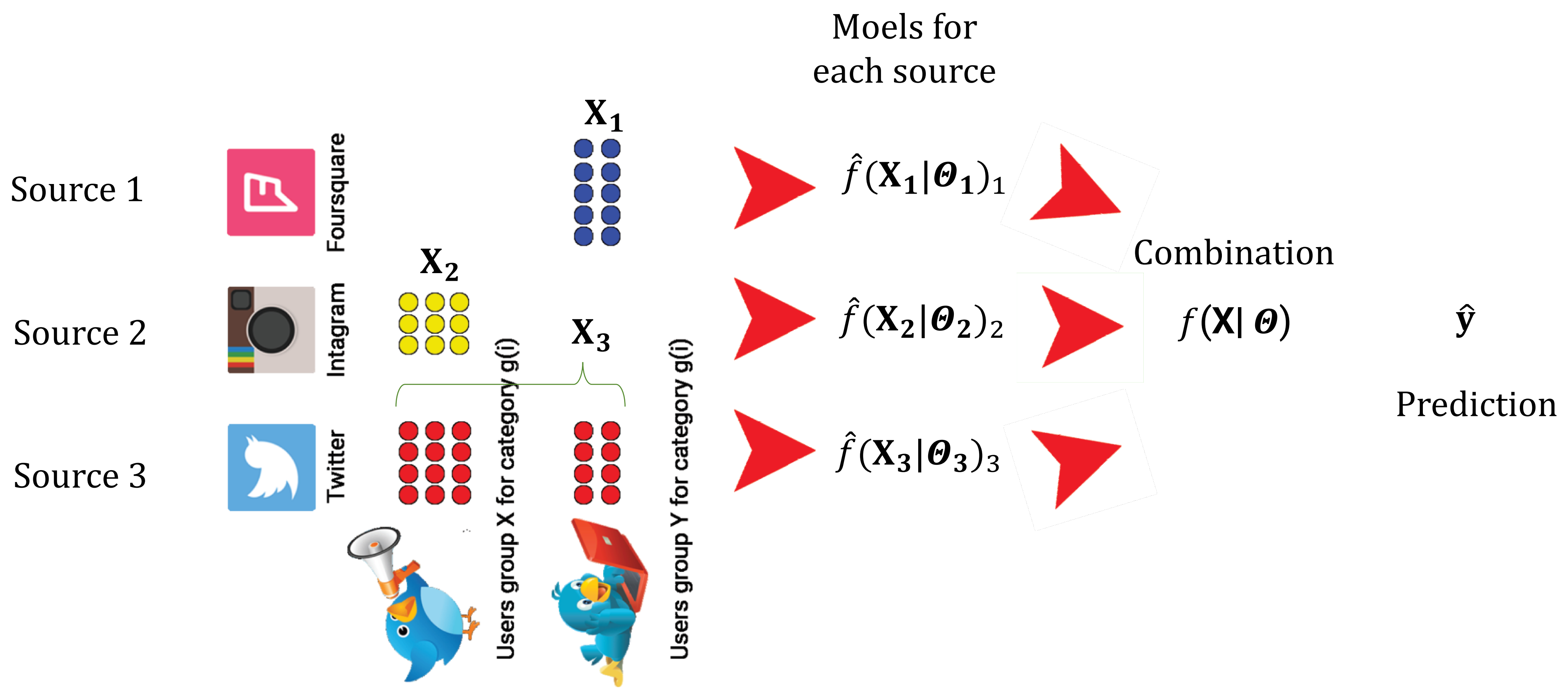
Methodology for Multi-Source Data Gathering via Cross-posting for Arbitrary number of Social Networks

02

*A. Farseev, N. Liqiang, M. Akbari, and T.-S. Chua. **Harvesting multiple sources for user profile learning: a Big data study.** ACM International Conference on Multimedia Retrieval (ICMR). China. June 23-26, 2015.

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Intuition behind late-fused multi-source learning



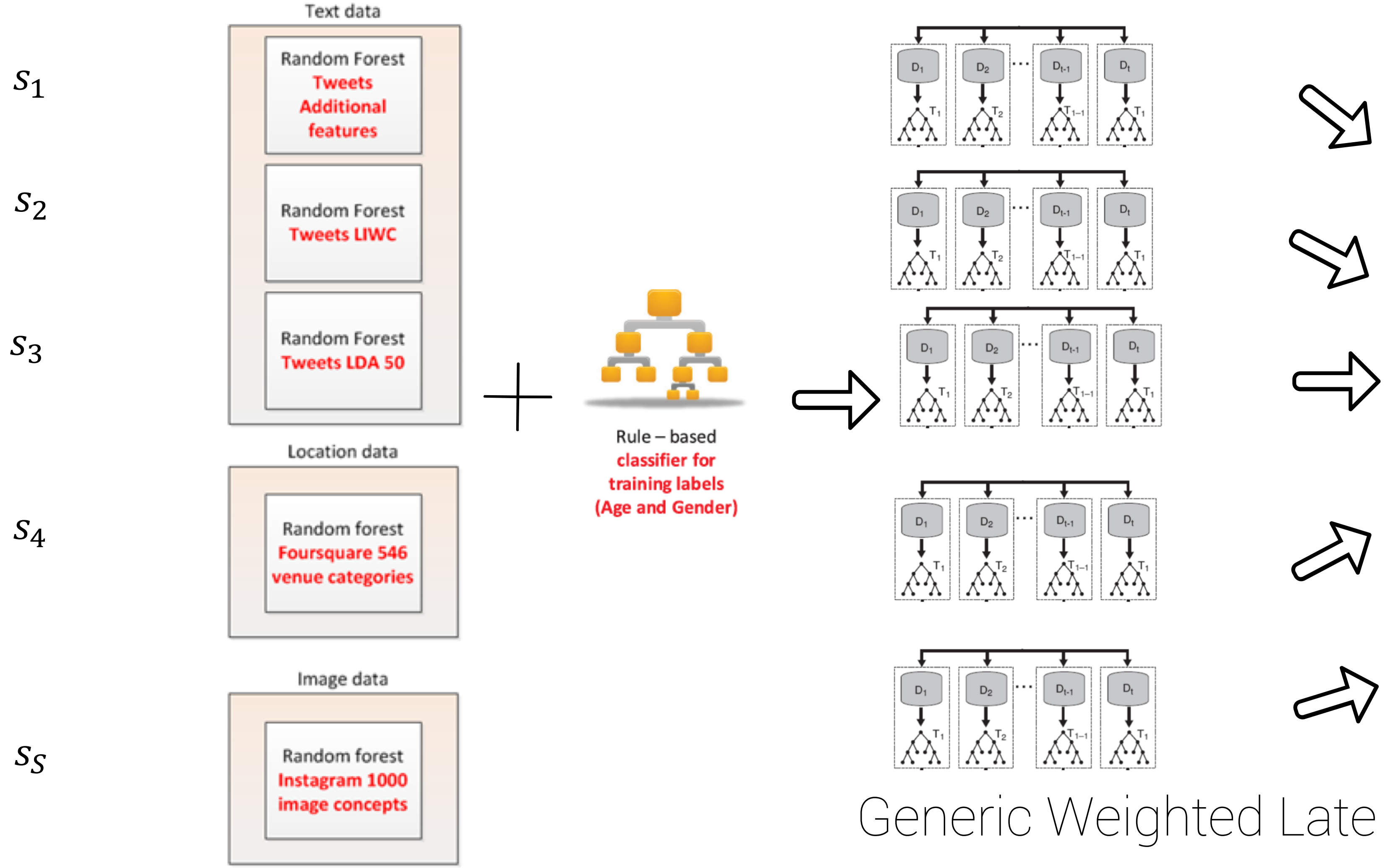
E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Age and Gender Prediction

Running Random Forests With Random Restart

Sources 1..S

Random Forests for each S



Weighted voting

$$f_i(\mathbf{X}) = \frac{1}{S}\sum_{s=1}^{S} \hat{f}_i(\mathbf{X}_s)_s \times w_s$$

$\hat{f}_i(\mathbf{X}_s)_s$ - s-th model prediction
w_S - s-th view weight obtained by Stochastic Hill Climbing

Generic Weighted Late Fusion Approach

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

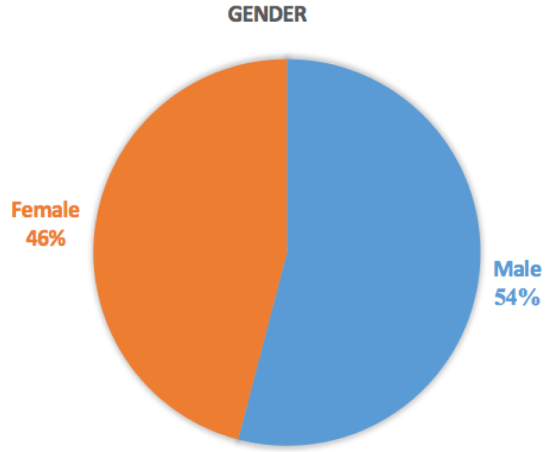
Multi-View User Attribute Inference via Late Fusion: Details

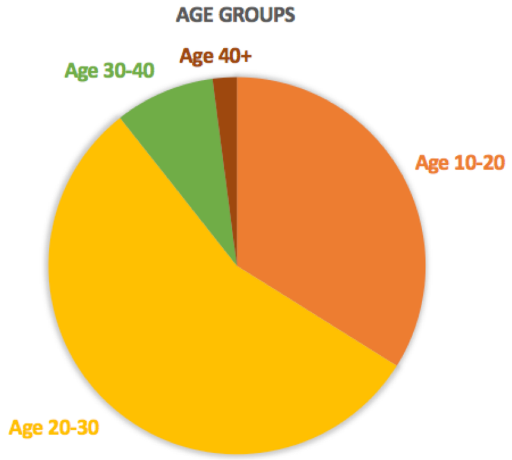
for Individual User Profiling

- $S > 1$ data sources
- D_s number of features extracted from each data source
- N_s number of exclusively labeled data samples for each data source
- $D = \sum_{s=1}^{S} D_s$ total number of features in all data sources
- $N = \sum_{s=1}^{S} N_s$ total number of samples from all data sources
- $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_S\} \in \mathbb{R}^{N \times D}$ block-wise missing feature matrix
- For each $\mathbf{X}_s$ we train a family of decision tree classifiers $h_i(\mathbf{X}_s)_s$, where each $h_i(\mathbf{X}_s)_{s,k} \equiv h(\mathbf{X}_s | \Theta_{i,s,k})$ with the model parameters $\Theta_{i,s,k}$ randomly chosen from s-th source model random vector $\Theta_{i,s}$
- To obtain s-th source classification output $\hat{f}_i(\mathbf{X}_s)_s$, s-th source model performs majority voting among $\{h_i(\mathbf{X}_s)_s\}$
- The final prediction is obtained by weighted voting among source models $f_i(\mathbf{X}) = \frac{1}{S}\sum_{s=1}^{S} \hat{f}_i(\mathbf{X}_s)_s \times w_s$, where w_S is s−th source weight

Age and Gender Ground Truth (NUS-MSS)

Attribute	Train (Age was Estimated from Education Path)	Test (Real Age Mentions)
Gender		
Male	2536	129
Female	2155	93
Age Groups		
10-20	360	181
20-30	589	28
30-40	91	8
40+	22	5





Note: Age ground truth is small
Solution: estimated age ground truth from users' Education and Occupation history

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Age and Gender Prediction: Results

About The Power Of Multiple Sources…

Data Source Combinations

Method	Gender	Age
Single-Source		
RF Location Cat. (Foursquare)	0.649	0.306
RF LWIC Text(Twitter)	0.716	0.407
RF Heuristic Text(Twitter)	0.685	0.463
RF LDA 50 Text(Twitter)	0.788	0.357
RF Image Concepts(Instagram)	0.784	0.366
Multi-Source combinations		
RF LDA + LIWC(Late Fusion)	0.784	0.426
RF LDA + Heuristic(Late Fusion)	0.815	0.480
RF Heuristic + LIWC (Late Fusion)	0.730	0.421
RF All Text (Late Fusion)	0.815	0.425
RF Media + Location (Late Fusion)	0.802	0.352
RF Text + Media (Late Fusion)	0.824	0.483
RF Text + Location (Late Fusion)	0.743	0.401
All sources together		
RF Early fusion for all features	0.707	0.370
RF Multi-source (Late Fusion)	0.878	0.509

Baselines

Method	Gender	Age
SVM Location Cat. (Foursquare)	0.581	0.251
SVM LWIC Text(Twitter)	0.590	0.254
SVM Heuristic Text(Twitter)	0.589	0.290
SVM LDA 50 Text(Twitter)	0.595	0.260
SVM Image Concepts(Instagram)	0.581	0.254
NB Location Cat. (Foursquare)	0.575	0.185
NB LWIC Text(Twitter)	0.640	0.392
NB Heuristic Text(Twitter)	0.599	**0.394**
NB LDA 50 Text(Twitter)	**0.653**	0.343
NB Image Concepts(Instagram)	0.631	0.233
Statistical Analysis		
Weighted Cohen's Kappa	$\kappa_w = 0.745, p < 0.01$	$\kappa_w = 0.297, p < 0.01$

4 Age Groups: <20; 20-30; 30-40; >40
2 Genders: Male; Female

*A. Farseev, N. Liqiang, M. Akbari, and T.-S. Chua. **Harvesting multiple sources for user profile learning: a Big data study.** ACM International Conference on Multimedia Retrieval (ICMR). China. June 23-26, 2015.

“

Health is a state of body. Wellness is a state of being…

- James Stanford

Individual Multi-View User Profiling

Part II: Wellness Profiling

*A. Farseev, A., & Chua, T. S. (2017). Tweetfit: Fusing multiple social media and sensor data for wellness profile learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. AAAI.

E-Mail: farseev@u.nus.edu l Website: http://farseev.com

Weight Problems Consequences

It is not just about looking not fit…

Weight Problems Consequences

- All-causes of death (mortality)
- **High blood pressure (Hypertension)**
- High / Low HDL cholesterol
- **Type 2 diabetes**
- **Coronary heart disease**
- Stroke
- Gallbladder disease
- **Osteoarthritis**
- **Some cancers**
- **Mental illness such as clinical depression**
- Body pain

*Health effect of overweight and obesity. *Center of disease control and prevention*. http://www.cdc.gov/healthyweight/effects/

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Idea II

Going further towards realizing the ideal of 360° User Profile Learning

Input One

1

Multi-Source Individual User Profiling is possible via multi-source data late fusion

Input Two

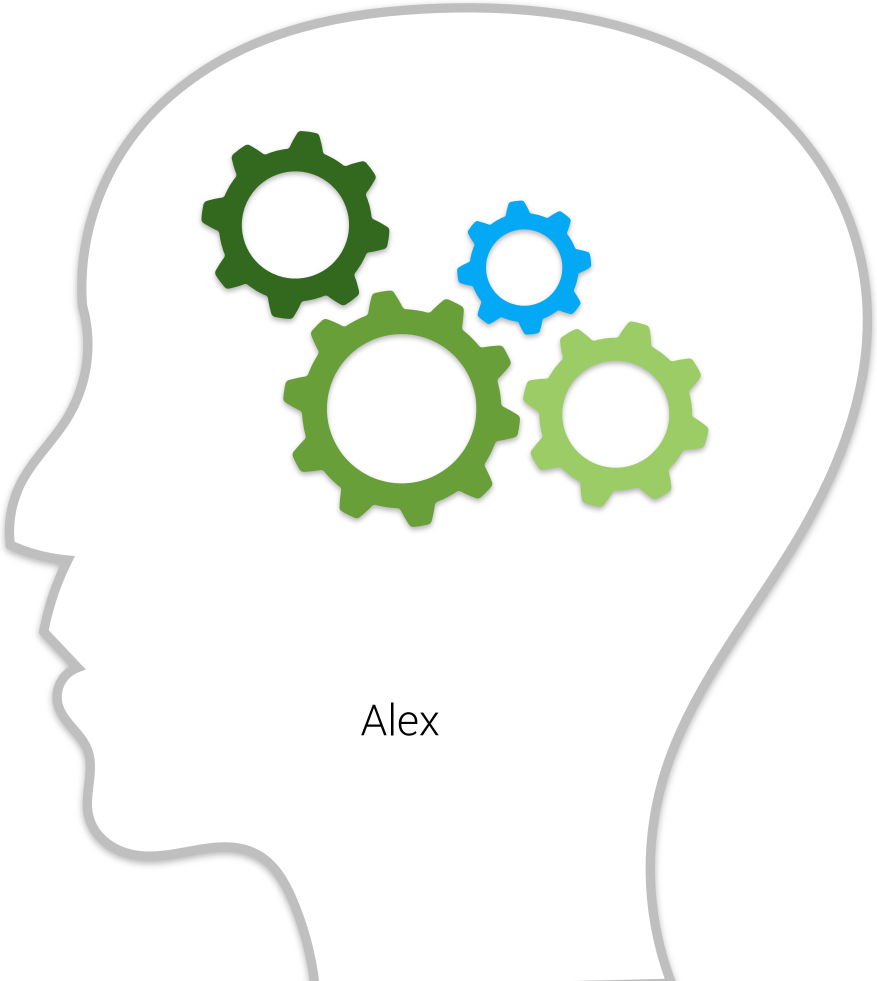
2

Wearable devices are widely popular. Sensor Data is available.

Input Three

3

Sensor data is tightly related and correlated with human overall health.



Idea

Fuse Sensor Data And Data From Multiple Social Networks For Wellness Profile Learning

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Research Questions

1

Question One

Is it possible to improve the performance of BMI category and “BMI Trend” inference by fusing multiple social media and sensor data?

2

Question Two

What is the contribution of sensor data towards BMI category and “BMI Trend” inference?

3

Question Three

Is it possible to improve the performance of BMI category and “BMI Trend” inference by incorporating inter-category relatedness into the learning process?

Contributions

Generic Model
For Supervised Joint Learning
From Multi-Source
Multi-Modal Incomplete Data

01

First Social-Sensor Dataset
NUS-SENSE

02

*Farseev, A., & Chua, T. S. (2017). Tweet can be Fit: Integrating Data from Wearable Sensors and Multiple Social Networks for Wellness Profile Learning. ACM Transactions on Information Systems (TOIS).

Unite Social Media And Wearable Sensors For Physical Attributes Inference

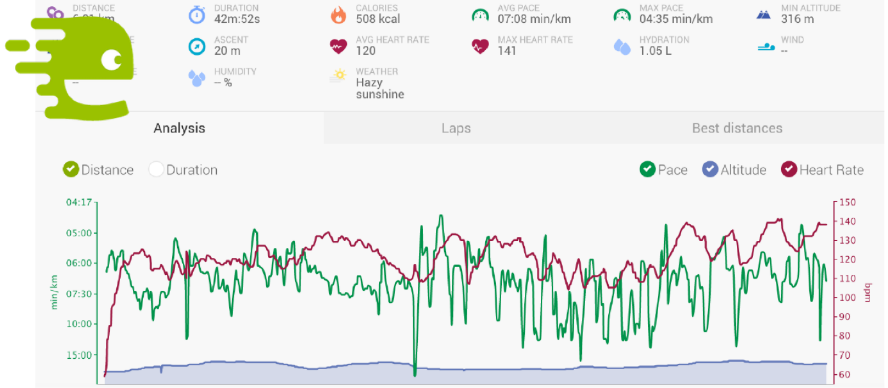
Just tweet to be fit….

FOURSQUARE

twitter

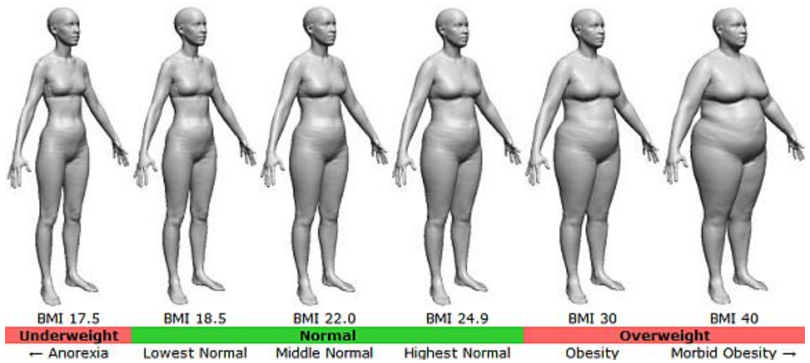
Instagram

+



Weight Fluctuation Trend (BMI Trend)

+



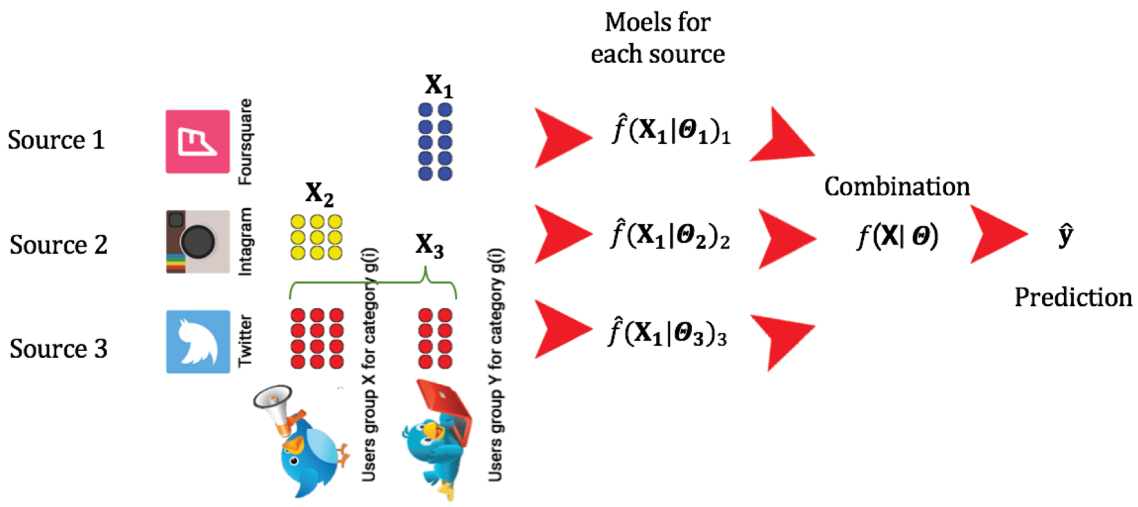
$$BMI=\frac{height}{weight^2}$$

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

What was the problem of the multi-source ensemble?

Previous framework advantages and limitations.

Limitation:



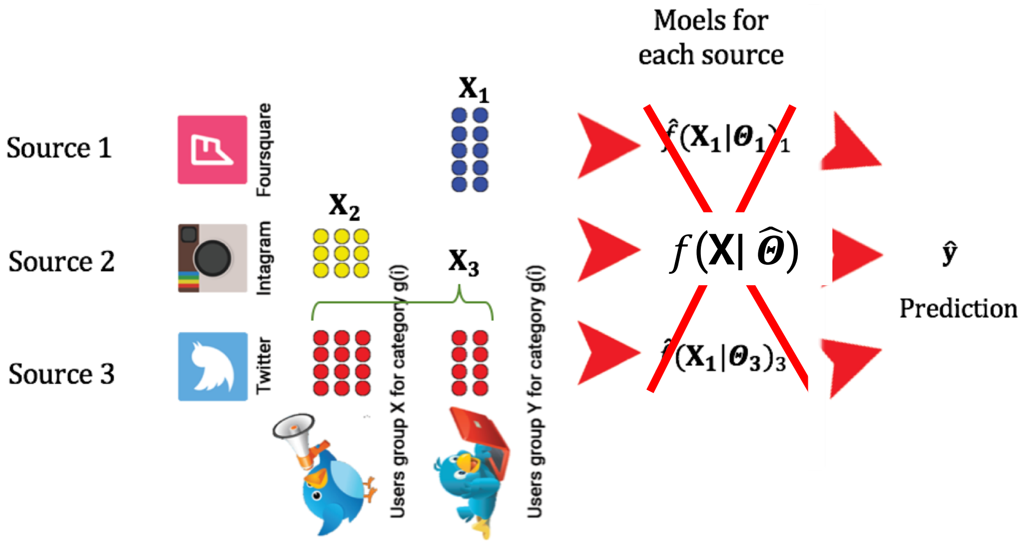
Models were trained independently:

$$\hat{f}(\mathbf{X_2}|\boldsymbol{\Theta_3})_3 \;\; \hat{f}(\mathbf{X_2}|\boldsymbol{\Theta_2})_2 \;\; \hat{f}(\mathbf{X_3}|\boldsymbol{\Theta_1})_1$$

Limitation: Information about source relations was not considered.

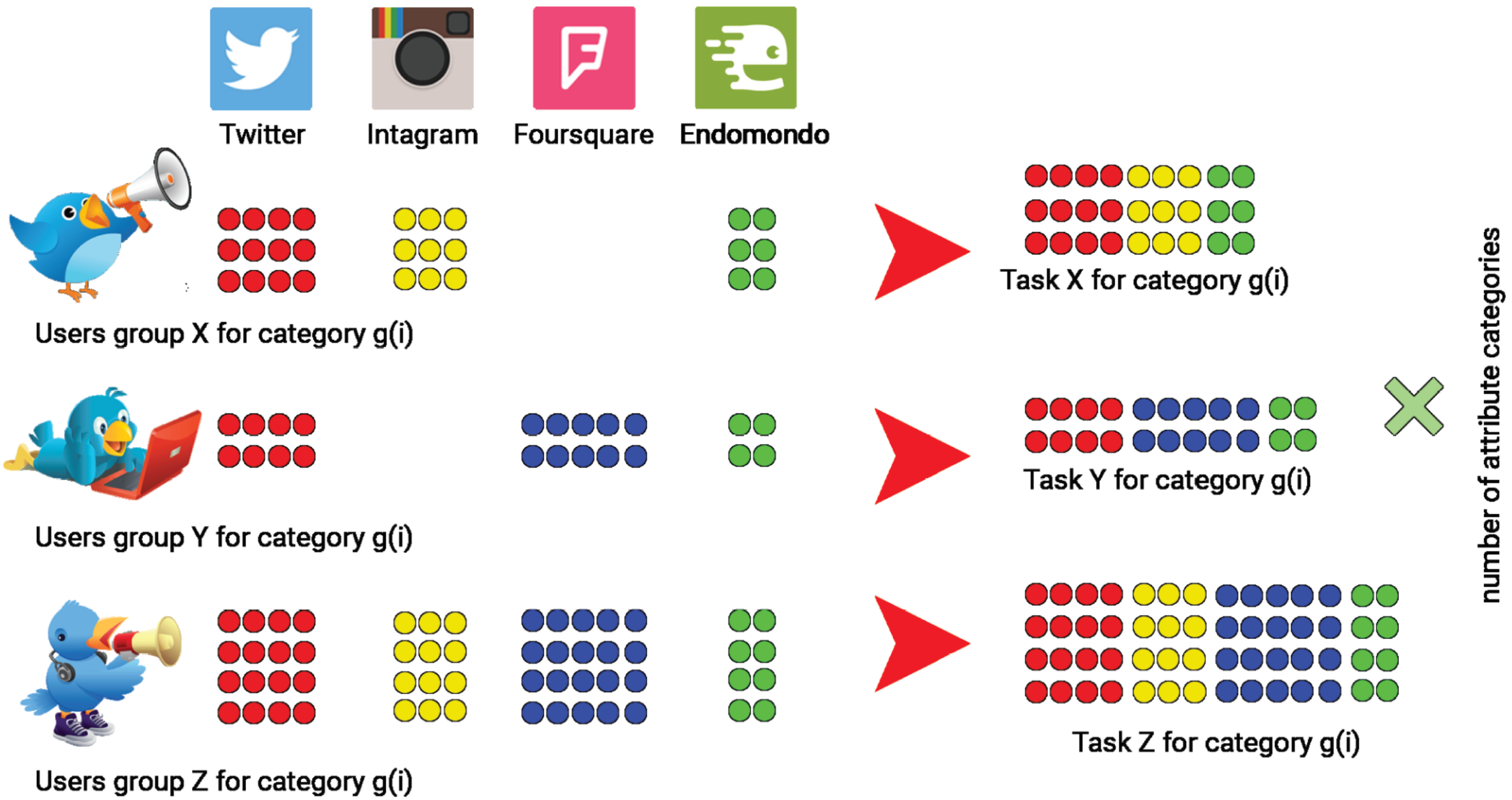
Solution:

Learn $f(\mathbf{X}|\widehat{\boldsymbol{\Theta}})$
from all data sources simultaneously.



E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Multi-Source Multi-Task Learning



Doing Predictions via Multi-Source Multi-Task Learning

Notations

Notation	Description
N	Number of exclusively labeled data samples
$S(\geq 2)$	Number of data sources (data modalities)
$G(\geq 1)$	Number of inference attribute categories (for BMI category, $G = 8$; for "BMI Trend", $G = 1$)
g	Inference attribute category (class). For example, "Obese" or "Normal" in case of BMI category attribute.
T	Number of multi-task learning Tasks
t	A multi-task learning Task
D_t	Dimensionality (feature vector dimension) of the task t
D_{max}	Maximum possible dimensionality of a task
N_t	Number of data samples of the task t
$\hat{T}$	Number of different existing combinations of sources
$f_t(\mathbf{x}_j^t; \mathbf{w}^t)$	Linear prediction model for the jth data sample of task t
$\mathbf{w}^t \in \mathbb{R}^{D_t}$	Model parameter vector of task t
$\mathbf{W}$	All model parameters, denoted as linear mapping block matrix
$\Gamma(\mathbf{W})$	Objective function
$\Psi(\mathbf{X}, \mathbf{W}, \mathbf{Y})$	Loss function
$\Upsilon(\mathbf{W})$	Sparsity regularizer
$\Omega(\mathbf{W})$	Inter-category smoothness regularizer
$\rho(s, f)$	Index function that denotes all the model parameters of the fth feature from the sth source
$\xi(t, g)$	Index function that picks up the model parameter ($\mathbf{w}_{g+1}^t$), which corresponds to the attribute category $g + 1$ (adjacent to g)

Generic Multi-View Hybrid Fusion Approach ☆

$$\Gamma(\mathbf{W}) = \arg\min_{\mathbf{W}} \Psi(\mathbf{X}, \mathbf{W}, \mathbf{Y}) + \lambda\Upsilon(\mathbf{W}) + \mu\Omega(\mathbf{W}),$$

$$\Psi(\mathbf{X}, \mathbf{W}, \mathbf{Y}) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N_t}\sum_{i=1}^{N_t}\log(1 + e^{-y_i^t f_t(\mathbf{x}_i^t; \mathbf{w}^t)})$$

$$\Upsilon(\mathbf{W}) = \sum_{s=1}^{S}\sum_{f=1}^{F_s}\left\|\mathbf{w}_{\rho(s,f)}\right\|$$

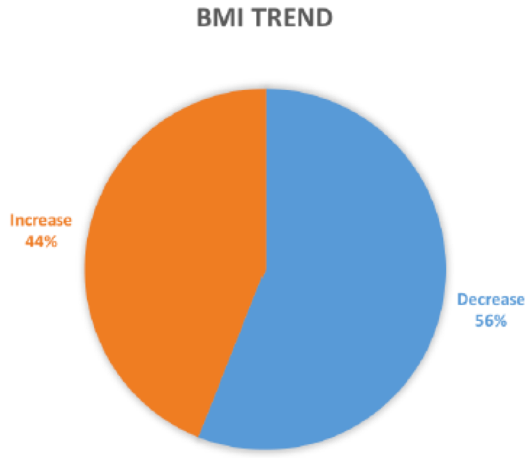
Inter-Category Smoothness Regularization ☆

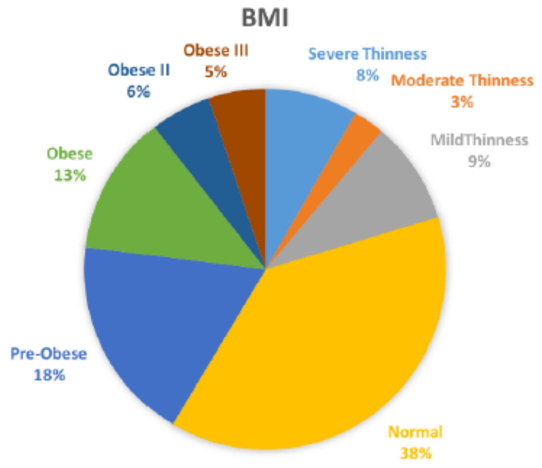
$$\Omega(\mathbf{W}) = \sum_{t=1}^{\hat{T}}\sum_{g\in C_{D_t}} \kappa_{g,\xi(t,g)}\left\|\mathbf{w}_g^t - \mathbf{w}_{\xi(t,g)}^t\right\|^2$$

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

BMI and BMI Trend Ground Truth (NUS-SENSE)

Attribute	Train	Test
BMI Trend		
Decrease	67	16
Increase	53	11
BMI		
Severe Thinness	71	16
Moderate Thinness	24	6
Mild Thinness	80	18
Normal	331	76
Pre Obese	157	36
Obese I	105	25
Obese II	47	11
Obese III	45	9





Note: data for some categories is not large.
Solution: applied SMOTE oversampling

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

BMI Category and BMI Trend Prediction: Results (1)

Data Source Combinations

Data Source Combination	BMI category prediction R_{Mac}/P_{Mac}	BMI category prediction $F_{1,Mac}$
Visual	0.049/0.188	0.077
Venue Semantics & Mobility	0.194/0.107	0.137
Sensors	0.153/0.158	0.155
Textual	0.229/0.146	**0.178**
Visual + Sensors	0.174/0.201	0.186
Visual + Text	0.126/0.245	0.166
Visual + Venue Semantics & Mobility	0.161/0.154	0.157
Text + Venue Semantics & Mobility	0.160/0.204	0.179
Sensors + Venue Semantics & Mobility	0.163/0.233	**0.191**
Sensors + Text	0.148/0.270	**0.191**
Visual + Text + Venue Semantics & Mobility	0.126/0.233	0.163
Sensors + Text + Visual	0.137/0.207	0.164
Sensors + Text + Venue Semantics & Mobility	0.182/0.236	0.205
Sensors + Venue Semantics & Mobility + Visual	0.180/0.283	**0.221**
All Data Sources	0.214/0.292	**0.246**

Other Baselines

Method	BMI category R_{Mac}/P_{Mac}	BMI category $F_{1,Mac}$	"BMI Trend" R_{Mac}/P_{Mac}	"BMI Trend" $F_{1,Mac}$
MSESHC [47]	0.141/0.145	0.142	0.634/0.655	0.644
Random Forest	0.135/0.226	0.169	0.333/0.863	0.480
iMSF [160]	0.171/0.174	0.172	0.649/0.649	0.649
$aMTFL_2$ [85]	0.162/0.215	0.184	0.700/0.722	0.710
TweetFit	0.222/0.202	0.211	0.705/0.732	**0.718**
M^{2}WP	0.221/0.229	**0.225**	Ω is not applicable	

8 BMI Categories:
Thinness I, II, III; Normal; Obese I, II, III, IV
2 BMI Trends: Increase; Decrease

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Source Importance Analysis: Feature level

Different feature types

Feature type	S. Th.	M. Th.	Md. Th.	Nrm.	P-Ob.	Ob. I	Ob. II	Ob. III
Latent topics	2	5	0	0	1	2	4	1
Lexicon	4	3	3	1	0	2	1	0
Writing style	2	1	1	1	0	2	1	3
Image con.	25	5	4	1	3	7	9	4
Venue sem.	19	5	1	2	11	5	11	2
Mob. & Tmp.	3	4	3	1	2	4	4	2
Work. sem.	1	0	1	0	1	2	2	2
Freq. domain	15	8	19	10	18	17	25	13
Work. stat.	1	2	2	1	1	2	2	3

1. Text features are less useful as compared to others (consistent with cross-source experiment).

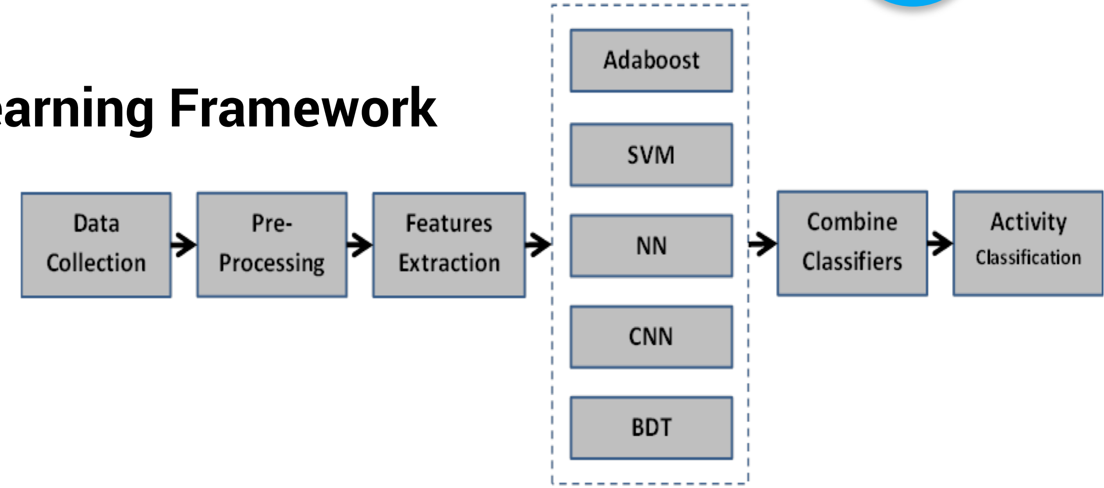
2. Image features are more helpful in distinguishing weight problems (abnormal BMI categories).

3. Venue categories (semantics) are more powerful for the whole BMI scale as compared to geographical mobility patterns.

4. Temporal workout features, are the most useful and absolutely necessary, while the type of exercise as well as exercise statistics play auxiliary roles.

On Power of Sensor Data for Other Applications…

Ensemble Learning Framework



1. **Simple sensor features (statistics) are powerful for such applications as Activity Classification.**

2. **For BMI classification, sensor features alone may not be sufficient.**

Class / Physical Exercise	F1 Score					
	Adaboost	SVM	NN	CNN	BDT	Fusion
Walking	1.00	1.00	0.67	1.00	0.56	**1.00**
Aerobics	0.83	0.83	0.33	0	0.83	**0.83**
Running	0.93	0.93	0.84	0.78	0.91	**0.95**
Indoor Cycling	0.60	0.76	0.25	0.59	0.55	**0.89**
Weight Training	0.86	0.75	0.46	0.75	0.29	**0.75**

* A. Kumar Chowdhury, A. Farseev, P.-R. Chakraborty, D. Tjondronegoro and V. Chandran (2017). Automatic Classification of Physical Exercises from Wearable Sensors using Small Dataset from Non-Laboratory Settings. In Proceedings of IEEE Life Science Conference (LSC).

> *Every successful individual knows that his or her achievement depends on a community of persons working together.*
>
> *- Paul Ryan*

Group Multi-View Learning

And Its Application In Urban Mobility Domain

(NUS-MSS)

User Community Definition

Communities are distinct groups of people, divided by a certain property.

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Why do we need to detect user communities (2)?

For personalized advertisement and marketing

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Idea III

Going further towards realizing the ideal of 360° User Profile Learning

Input One

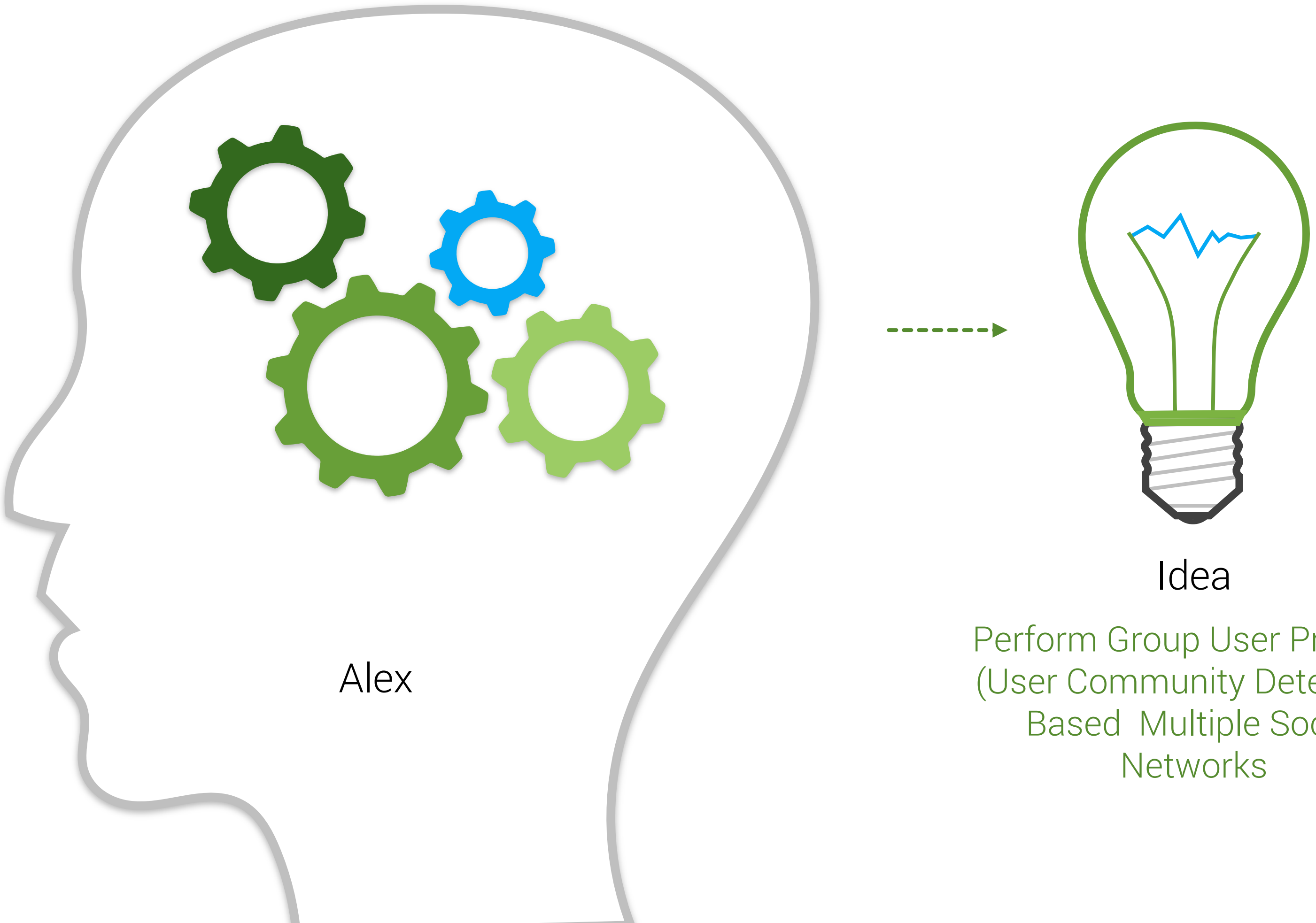
1

Integration of Data from multiple social networks helps in improving Individual Profiling Performance.

Input Two

2

It was not well investigated if multi-source data allows for group profiling performance improvement.



Idea

Perform Group User Profiling (User Community Detection) Based Multiple Social Networks

E-Mail: farseev@u.nus.edu l Website: http://farseev.com

Research Questions

Question One

1

What the contribution is of each data source (modality) towards group user profiling?.

Question Two

2

Inter-source relationship information enable us to find better user communities.

Question Three

3

Does incorporation of group user profiling boost recommendation performance?

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Related Research Directions: Group User Profiling

Single-Source Clustering

Mostly mono-mo... mono-source ... processing...

- Gr...g
- ...Clustering
- ...Clustering
- ...ctral clustering
- Quality Optimization

MONO-SOURCE OR MONO-MODAL

Multi-Source Clustering

Li... umber of ... to the field. ...n ... con... ...sion manne... ...ed jointly. Int... relationship is... considered.

LATE OR EARLY FUSION

Urban Mobil... ...ns

Mostly m... mono... pr... ...tion, ...qualitative ...s.

MONO-SOURCE OR MONO-MODAL

...ss-Domain Recommend...tion

INTER-SOURCE REL. NOT CONSIDERED

Usually, this problem of group user profiling is treated as a unsupervised learning task and performed based on mono-source mono-modal datasets. Research on multi-source group profiling is relatively sparse and limited by lack of inter-source relationship modeling. Implicit evaluation is adopted in most of the related studies.

E-Mail: farseev@u.nus.edu l Website: http://farseev.com

Contributions

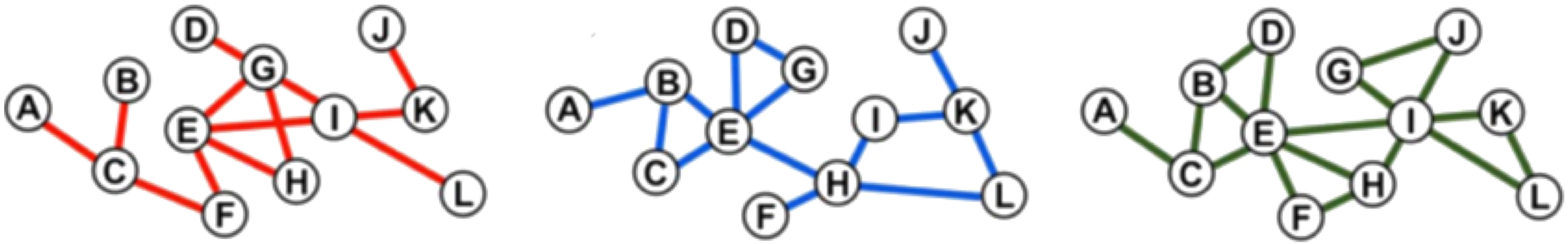
01 New Generic Approach For Clustering On Multi-Layer Graphs that considers inter-layer relationship

02 New Schema For Automatic Inter-Layer Relationship Inference From Data

*Farseev, A., Samborskii, I., Filchenkov, A., & Chua, T. S. (2017 August). Cross-Domain Recommendation via Clustering on Multi-Layer Graphs. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval.

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Multi-source data representation via multi-layer graphs



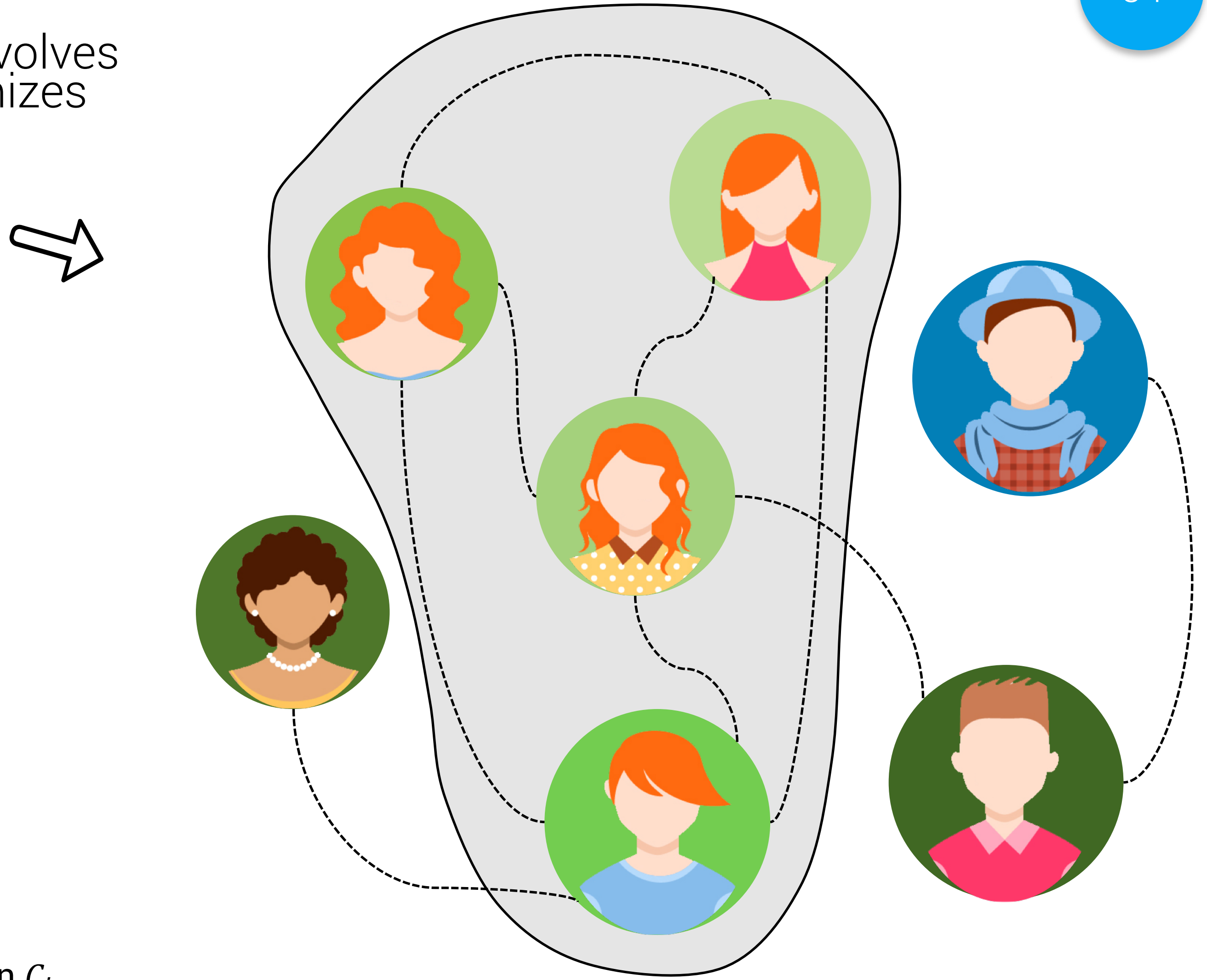
Multi-layer graph – graph G, where $G = \{G_i\}$, $G_i = (V, E_i)$

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Problem Formulation: MinCut Problem

For a given number k of subsets, the $MinCut$ involves choosing a partition $C_1, \ldots, C_k$ such that it minimizes the expression:

$$cut(C_1, \ldots, C_k) = \sum_{i=1}^{k} W(C_i, \bar{C}_i)$$



The state-of-the-art approximation of $MinCut$ is formulated in a form of
standard trace minimization problem

$$\min_{U \in R^{n \times k}} \mathrm{tr}(U^T L U), \mathrm{s.t.}\, U^T U = I$$

Can be solved by Spectral Clustering

Or for multi-layer graph:

$$\min_{U \in R^{n \times k}} \mathrm{tr}(U^T L_{sum} U), \text{where } L_{sum} = \sum_{i=1}^{M} L_i$$

Unable to handle cross-source relationship

*W is the sum of weights of edges attached to vertices in C_i

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Extending definition of trace minimization

$$\min_{U \in R^{n \times k}} \sum_{i=1}^{M} \mathrm{tr}(U^T L_i U), \mathrm{s.t.}\, U^T U = I$$

$$\min_{U \in R^{n \times k}} \mathrm{tr}(U^T L_{sum} U), \text{where } L_{sum} = \sum_{i=1}^{M} L_i$$

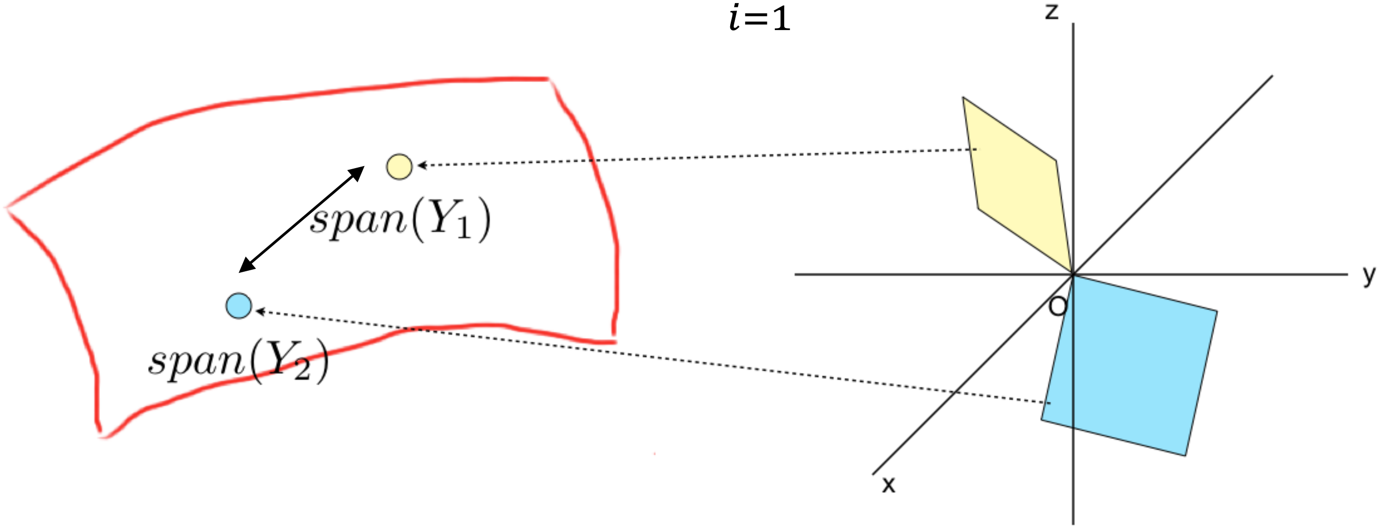
Suffers from poor generalization ability.

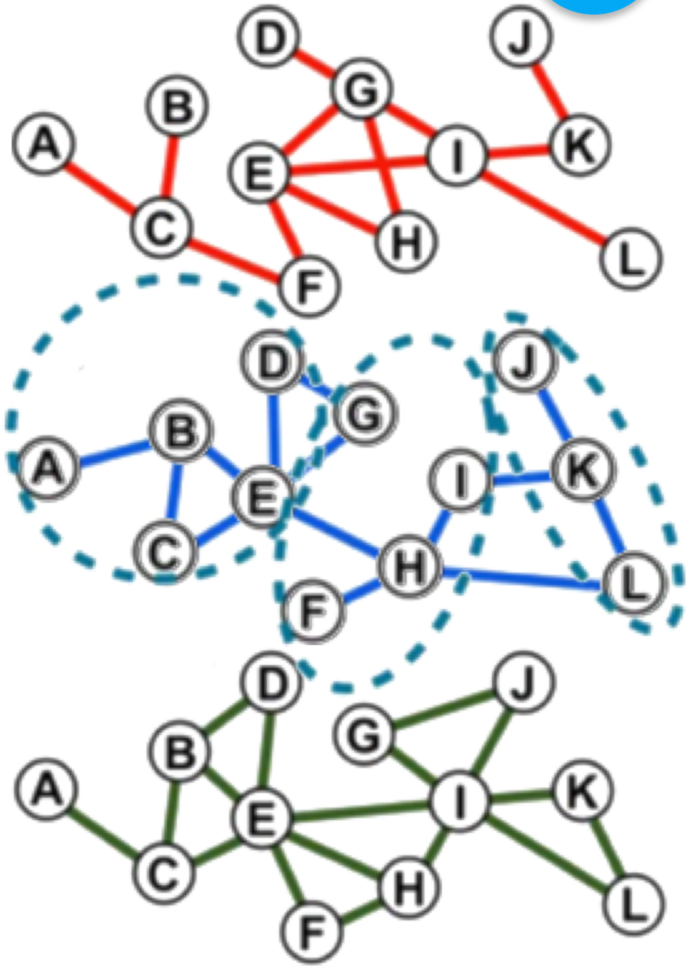
Regularized Clustering on Multi-layer Graph -1

Distance between two spaces Y_1 and Y_2 (graph layers) can be measured on Grassman Manifolds:

$$d^2_{Proj}(Y_1, Y_2) = \frac{1}{2}\left\|Y_1 Y_1^T - Y_2 Y_2^T\right\|_F^2$$

$$d^2_{Proj}\left(S, \{S_i\}_{i=1}^M\right) = kM - \sum_{i=1}^{M} tr(SS^T - S_i S_i^T)$$





* X. Dong, P. Frossard, P. Vandergheynst, and N. Nefedov. Clustering on multi-layer graphs via subspace analysis on grassmann manifolds. IEEE Tr. on Signal Processing, 2014.

E-Mail: farseev@u.nus.edu I Website: http://farseev.com

Regularized Clustering on Multi-layer Graph: Inter-Layer Smoothness

Extends the objective function to introduce the subspace analysis regularization

$$\min_{U\in\mathbb{R}^{n\times k}} \sum_{i=1}^{M} \operatorname{tr}\left(U^T L_i U\right) + \alpha\left(kM - \sum_{i=1}^{M} \operatorname{tr}\left(UU^T U_i U_i^T\right)\right), \text{s.t.}\ U^T U = I$$

$$\min_{U\in\mathbb{R}^{n\times k}} \operatorname{tr}(U^T \boxed{L_{mod}} U)$$

$$L_{mod} = \sum_{i=1}^{M} (L_i - \alpha U_i U_i^T)$$

Regularized Clustering on Multi-layer Graph: Inter-Layer Relationship

By using distance on Grassman Manifolds, we present the new objective function for the i^{th} layer:

$$\min_{\hat{U}_i \in \mathbb{R}^{n \times k}} \operatorname{tr}\left(\hat{U}_i^T L_i \hat{U}_i\right) + \beta_i \left(kM - \sum_{j=1, j \neq i}^{M} w_{i,j} \operatorname{tr}\left(\hat{U}_i \hat{U}_i^T U_j U_j^T\right) \right)$$

$$\min_{\hat{U}_i \in \mathbb{R}^{n \times k}} \operatorname{tr}\left(\hat{U}_i^T \boxed{\hat{L}_i} \hat{U}_i\right)$$

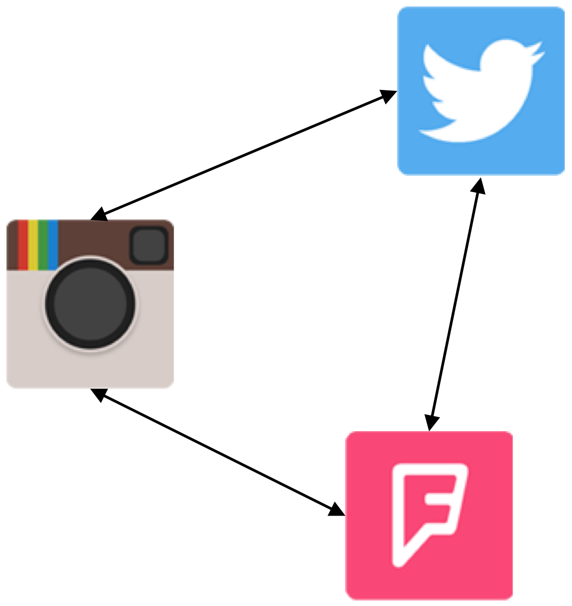
$$\boxed{\hat{L}_i = L_i - \beta_i \sum_{j=1, j \neq i}^{M} w_{i,j} \operatorname{tr}\left(U_j U_j^T\right)}$$

New Inter-Layer Relationship Regularization ☆

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Heuristic to estimate $\mathbf{w}_{i,j}$ when computing i-th layer ?

$$\min_{\hat{U}_i \in \mathbb{R}^{n\times k}} \operatorname{tr}\left(\hat{U}_i{}^T \hat{L}_i \hat{U}_i\right), \hat{L}_i = L_i - \beta_i \sum_{j=1,j\neq i}^{M} \boxed{w_{i,j}} \operatorname{tr}\left(U_j U_j^T\right)$$



Inter-layer relationship graph $\boldsymbol{R(V,E)}$ – weighted graph which represents the similarity between layers.

$$\forall (i,j) \in E, w_{i,j} = \frac{\sum_{k=2}^{K}\left(1 - \frac{\left\|M_{i,k} - M_{j,k}\right\|}{\sqrt{N(N-1)}}\right)}{K-1}$$

where $M_{i,k}$ is clustering co-occurrence matrix of layer i, $m_{a,b} = 1$, if users a and b assigned to the same cluster, and 0 otherwise.

Generic Heuristic for Layer Similarity Computation for Multi-Layer Graph Clustering

Regularized Clustering on Multi-layer Graph: Final objective function

Let's combine equations from previous slides to define the final objective function:

$$\min_{U \in \mathbb{R}^{n\times k}} \sum_{i=1}^{M} \operatorname{tr}\left(U^T \hat{L}_i U\right) + \alpha\left(kM - \sum_{i=1}^{M} \operatorname{tr}\left(UU^T \widehat{U}_i \widehat{U}_i^T\right)\right) =$$

$$= \min_{U \in \mathbb{R}^{n\times k}} \operatorname{tr}\left(U^T \boxed{\sum_{i=1}^{M} \left(\hat{L}_i - \alpha \widehat{U}_i \widehat{U}_i^T\right)} U\right)$$

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Explicit Evaluation: Examples of Detected User Communities in Singapore

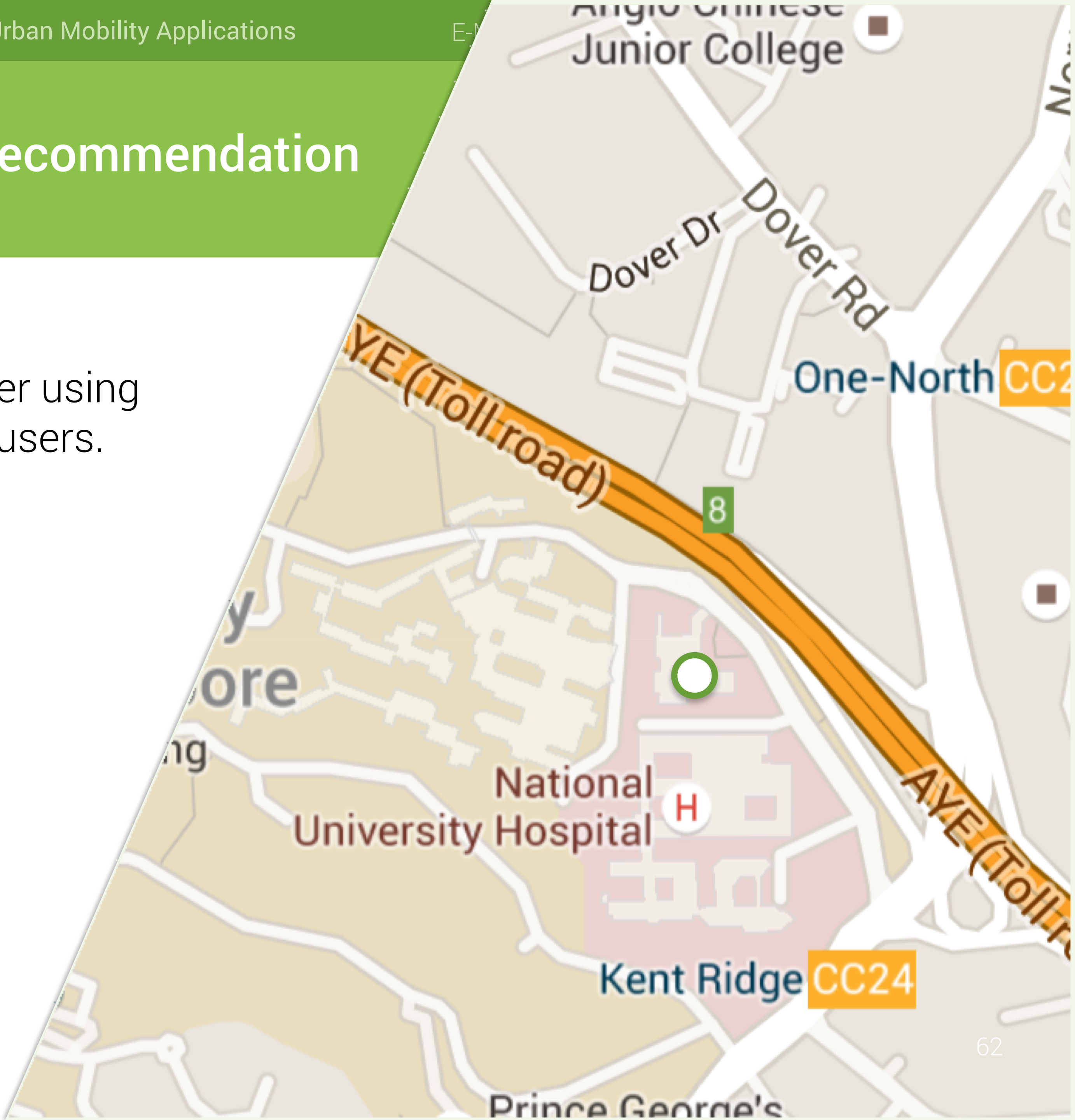
Name	Bag of Words for different modalities		
	Text	Visual	Location
Gadgets 832 users	device, launcher, android	mouse, digital clock, hard disc	electronics store, tech startup, technology building
Arts 538 users	painting, landscape, reflection	obelisk, paint-brush, pencil box	arts & crafts store, arts & entertainment, museum
Food 446 users	dining, coffee, cooking	pineapple, microwave, frying pan	italian restaurant, pizzeria, macanese restaurant

Implicit Evaluation: Venue Category Recommendation

Venue category recommendation – recommendation of venue categories to user using information about his past visits and other users.

Result: ranked venue category list:
Clothing Store Rank (Rank 1)
Hotel Rank (Rank 2)
...
Ice Cream Shop (Rank K = 764)

Total 764 different categories



Venue Category Recommendation

Collaborative Venue Category Recommendation – recommendation of venue categories (i.e. restaurant, cinema) to user using information about his/her profile (i.e. past visits) and/or information about users from the same domain.

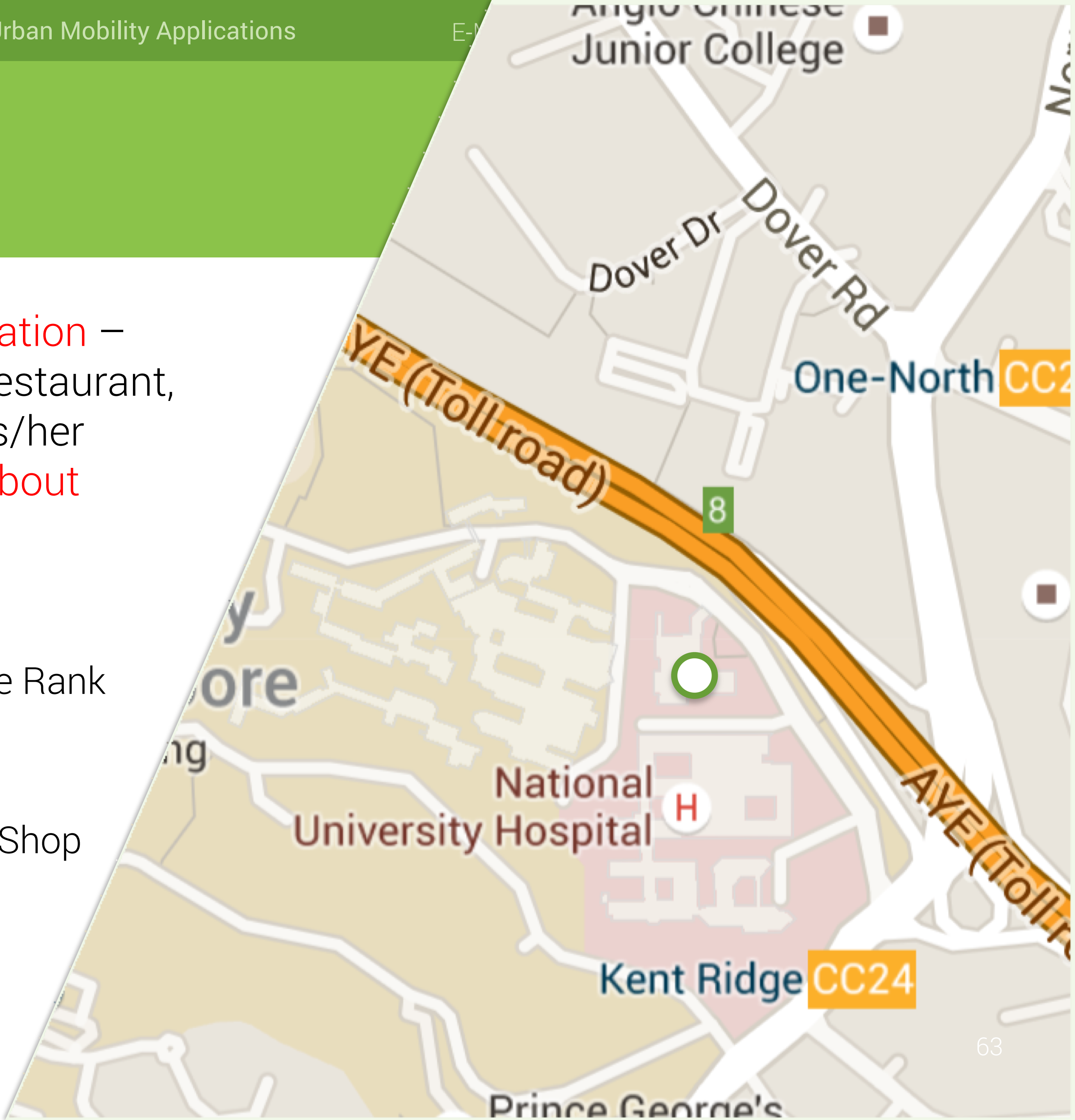
Ranked list

1. Clothing Store Rank
2. Hotel Rank

...

764. Ice Cream Shop

Venue visits in past:



Implicit Evaluation: Cross-Domain Venue Category Recommendation

Cross Domain - Venue category recommendation – recommendation of venue categories (i.e. restaurant, cinema) using information about his/her profile (i.e. past visits) and/or information about users from other sources (i.e. images, texts, location types).

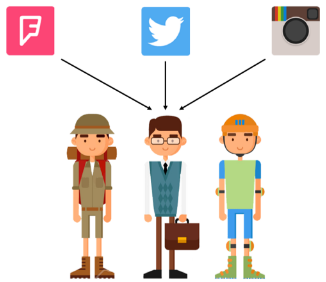
Ranked list:

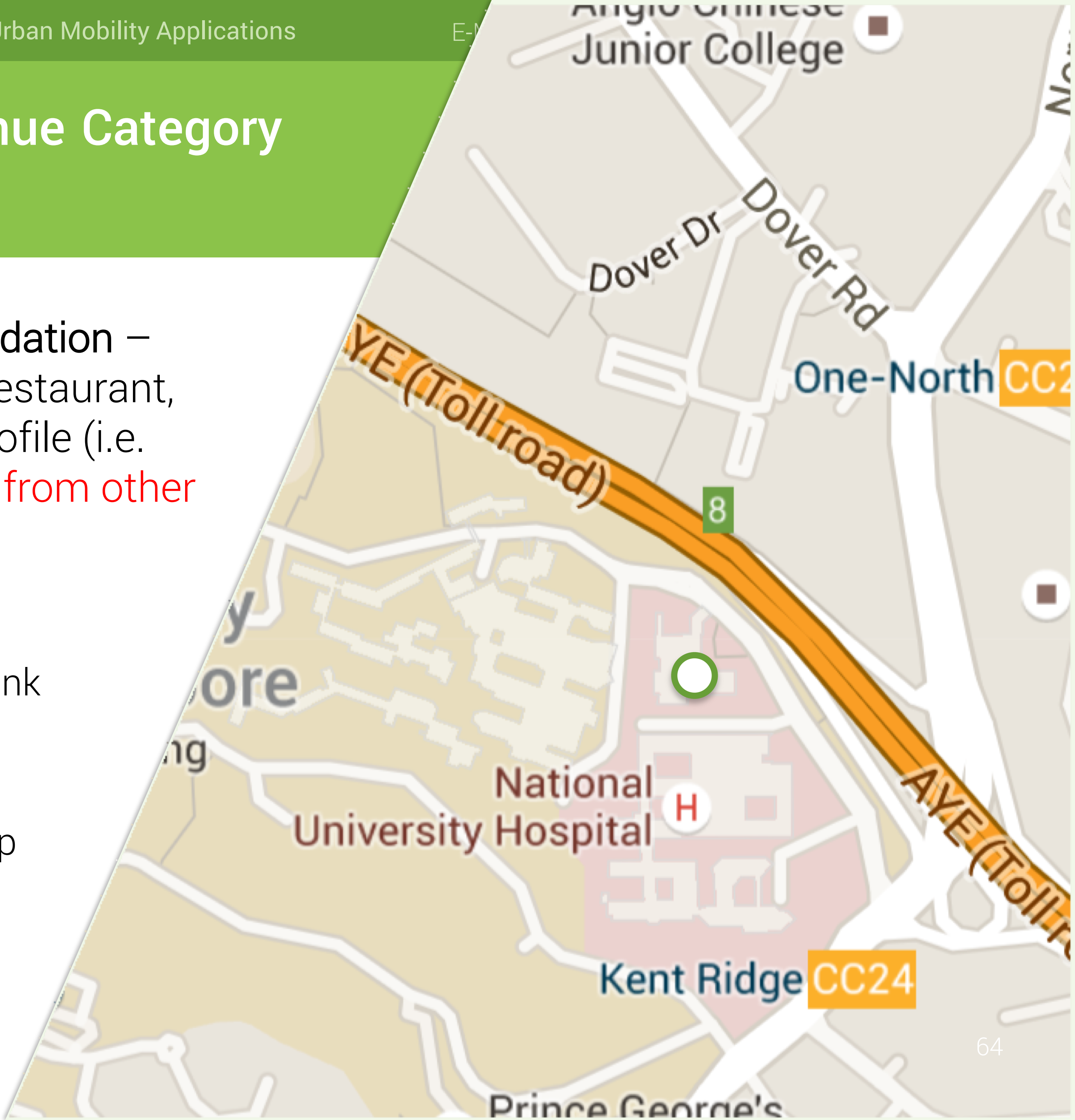
1. Clothing Store Rank
2. Hotel Rank

...

764. Ice Cream Shop

Multi-Source Data:



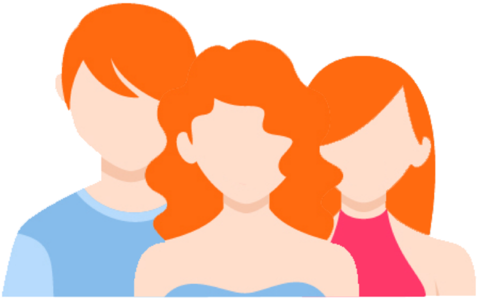


E-Mail: farseev@u.nus.edu l Website: http://farseev.com

Cross-Source Community-Based Collaborative Recommendation

We perform venue category recommendation based on both Individual and Group Knowledge:

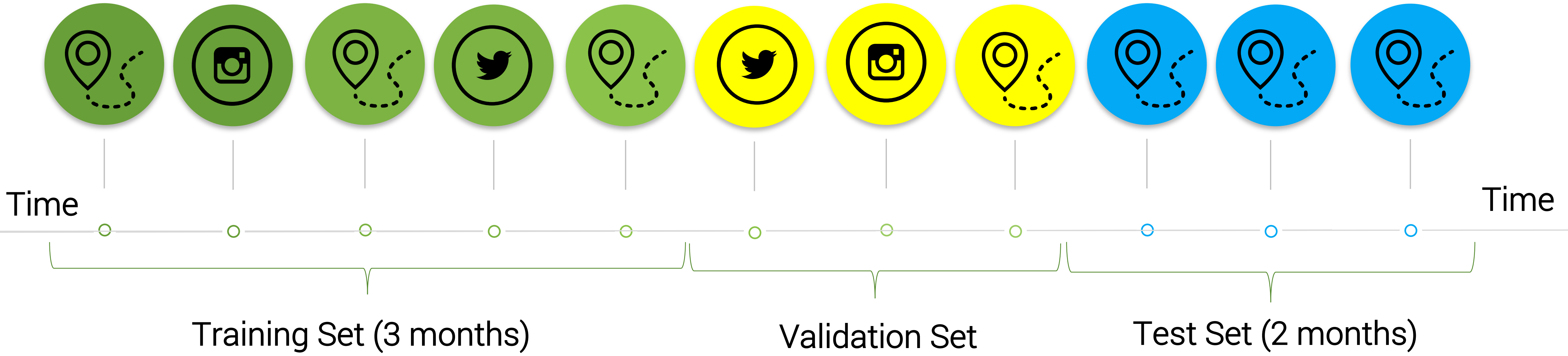
$$rec(u) = sort\left(\gamma \cdot vec_u + \theta \frac{\sum_{v \in C_u} vec_v}{|C_u|}\right)$$


+


E-Mail: farseev@u.nus.edu l Website: http://farseev.com

Data

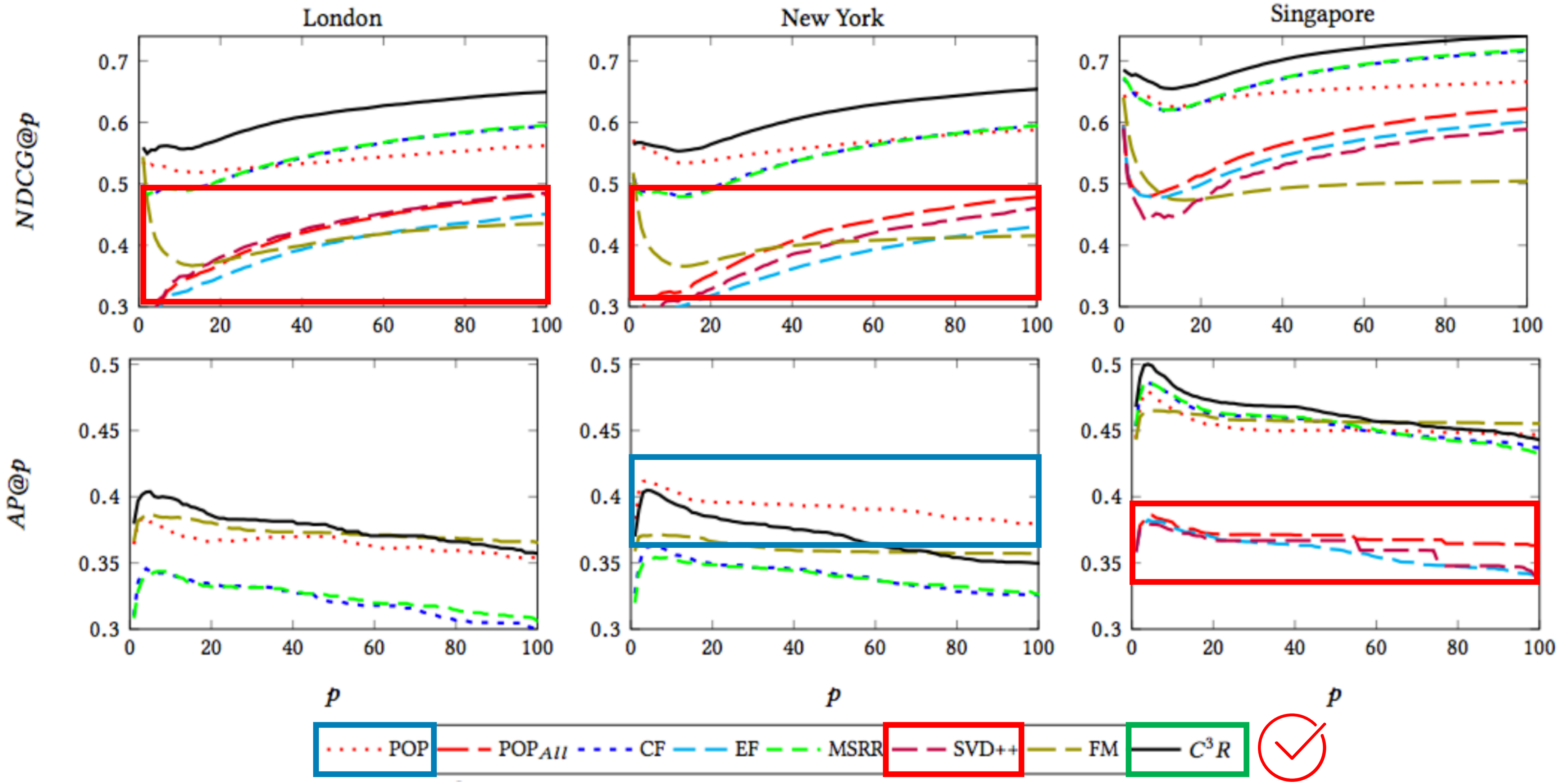
NUS-MSS Dataset was split into training, validation and testing sets on a time basis:



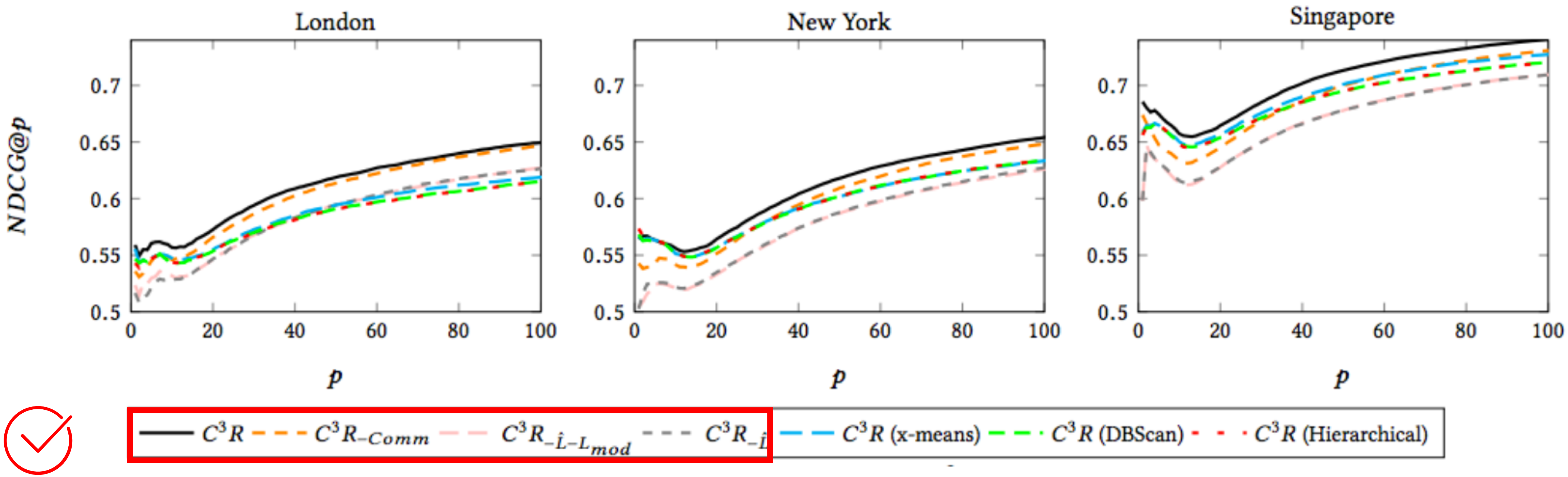
1801 users from Singapore, 813 users from London, and 1602 users from New York with data from all three social networks in training, validation and testing sets.

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Evaluation Against Recommender Systems



Evaluation Against Community Detection Approaches

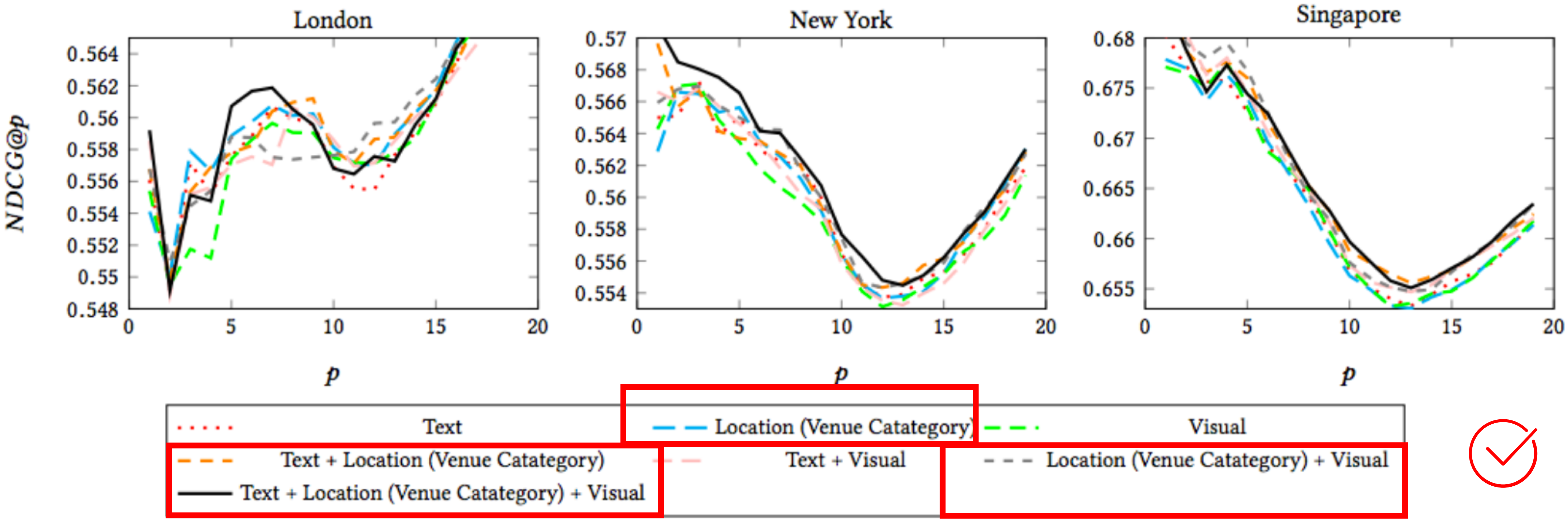


+ Incorporation of group knowledge is is important

+ Multi-modal clustering performs better than single-source clustering

+ Incorporation of Inter-Source relationship is crucial.

Comparing Data Source Combinations



+ In different geo regions, different data sources are of different importance

+ Location data is more powerful than other data modalities

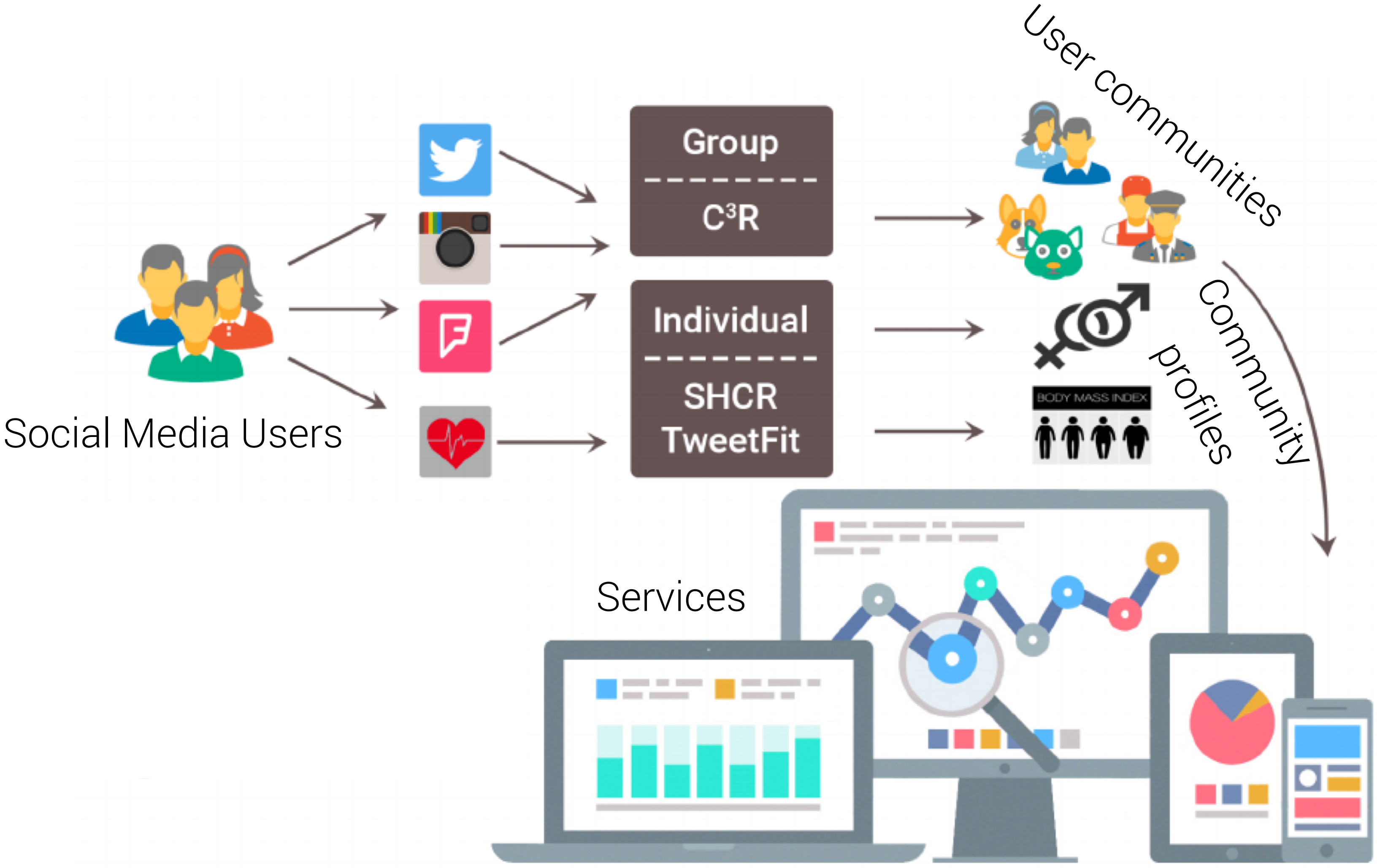
$$W_R = \left(\begin{array}{c|ccccc} & \textbf{tw} & \textbf{4sq} & \textbf{inst} & \textbf{tmp} & \textbf{mob} \\ \hline \textbf{tw} & 1 & 0.632 & 0.621 & 0.643 & 0.561 \\ \textbf{4sq} & 0.632 & 1 & 0.614 & 0.631 & 0.570 \\ \textbf{inst} & 0.621 & 0.614 & 1 & 0.621 & 0.551 \\ \textbf{tmp} & 0.643 & 0.631 & 0.621 & 1 & 0.560 \\ \textbf{mob} & 0.561 & 0.570 & 0.551 & 0.560 & 1 \end{array}\right)$$

Overall Framework

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

360° User Profile Learning Framework

Multi-Source User Profiling And Its Applications in One Framework….



E-Mail: farseev@u.nus.edu | Website: http://farseev.com

The Framework in Action

Services

DEMO

“

If you can’t explain it simply, you don’t understand it well enough.

- Albert Einstein, Physicist

User Profiling Analytics

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Idea IV

Understanding inter-source relationship from the data perspective

Input One

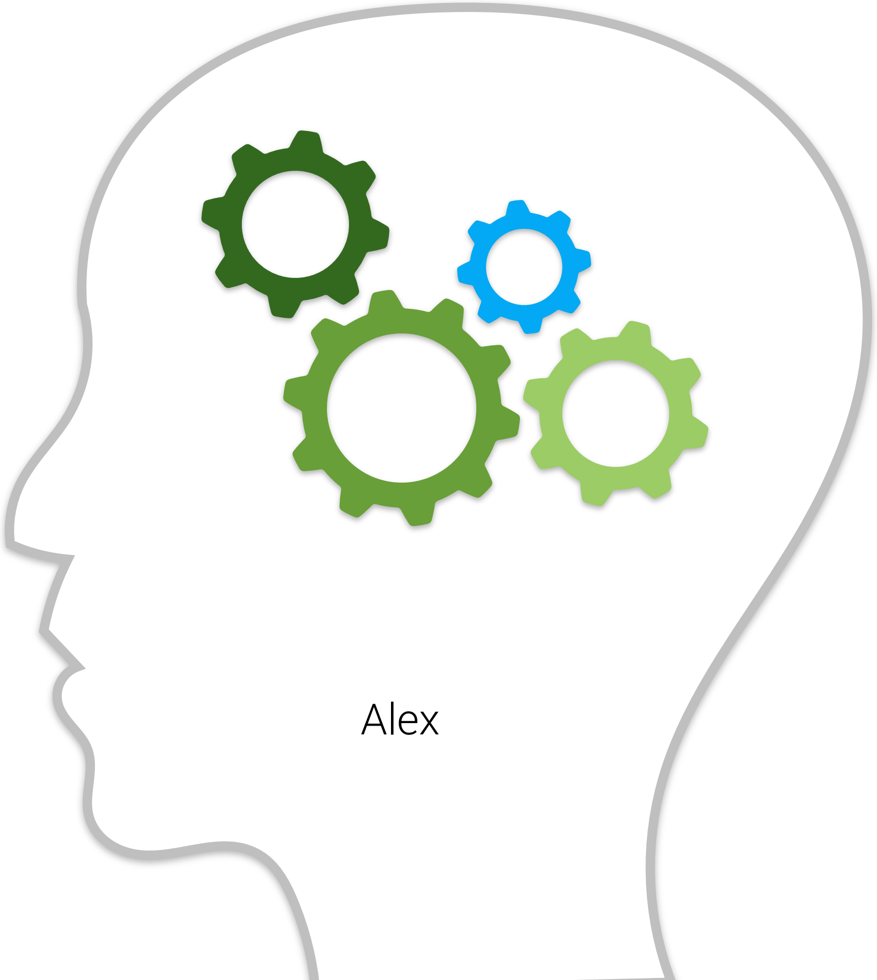
1

Multi-source supervised an unsupervised machine learning clearly outperforms single-source learning baselines.

Input Two

2

The statistical relationship between data sources was not yet studied.



Idea

Perform inter-source and intra-source Pearson correlation analysis between data sources

E-Mail: farseev@u.nus.edu l Website: http://farseev.com

Research Questions

Question One

1

What is the relation between different data modalities, data sources, and individual user attributes?

Question Two

2

What is the relation between different data modalities, data sources, and detected user communities?

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Related Research Directions: Social Media Data Analytics

Descriptive Analytics

Description and discussion regarding collected social media (mainly Twitter). Main ap mmarization and visualiz Word Clouds, Distribution Topics, etc.

NOT INFORMATIVE

Census vs Social Media

In most cases, comparison of publicly available census e. US population Obesity) edia data statistic pregn 14], by Mejova et ories from I pictures by Sharma et al. [20], publishing traits by Park et al. [2016], diet success on MyFitnessPal by Weber et al. [2015]

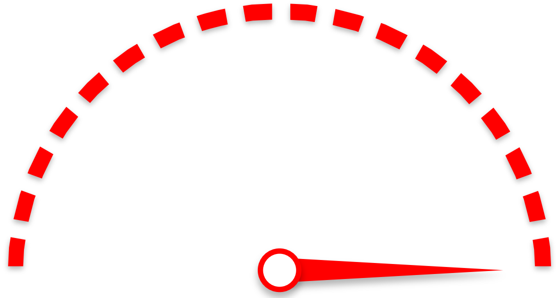
SINGLE-SOURCE OR SINGLE-MODAL

Most of the works apply statistical analysis on data from single source/modality, but did not study the relationship within sources / modalities.

Contributions

First study on Cross-Modal Statistical Analysis of Users from Multiple Social Networks

01



*Farseev, A., & Chua, T. S. (2017). Tweet can be Fit: Integrating Data from Wearable Sensors and Multiple Social Networks for Wellness Profile Learning. ACM Transactions on Information Systems (TOIS).

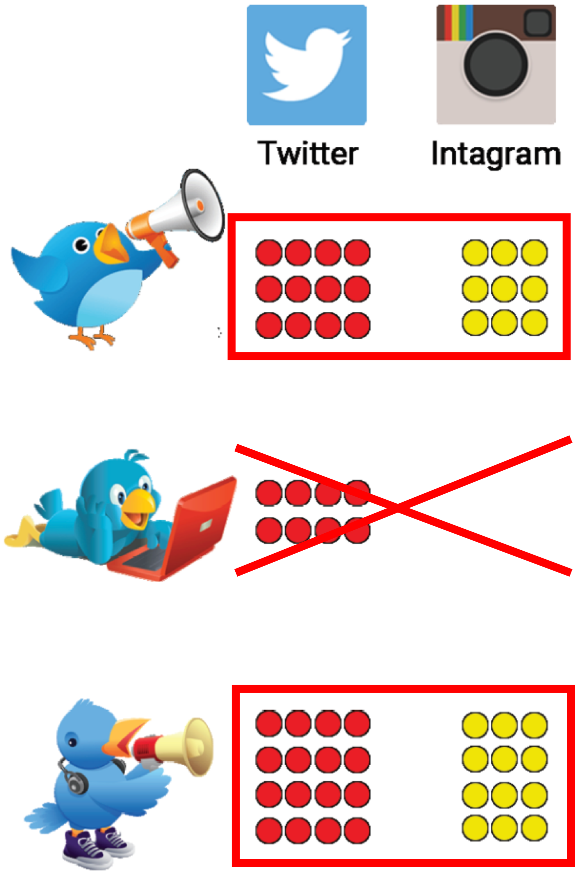
E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Pearson Correlation to Visualize Significant Data Relationships

Sample correlation coefficient r – an estimate of the unknown correlation coefficient for a representative sample of size n:

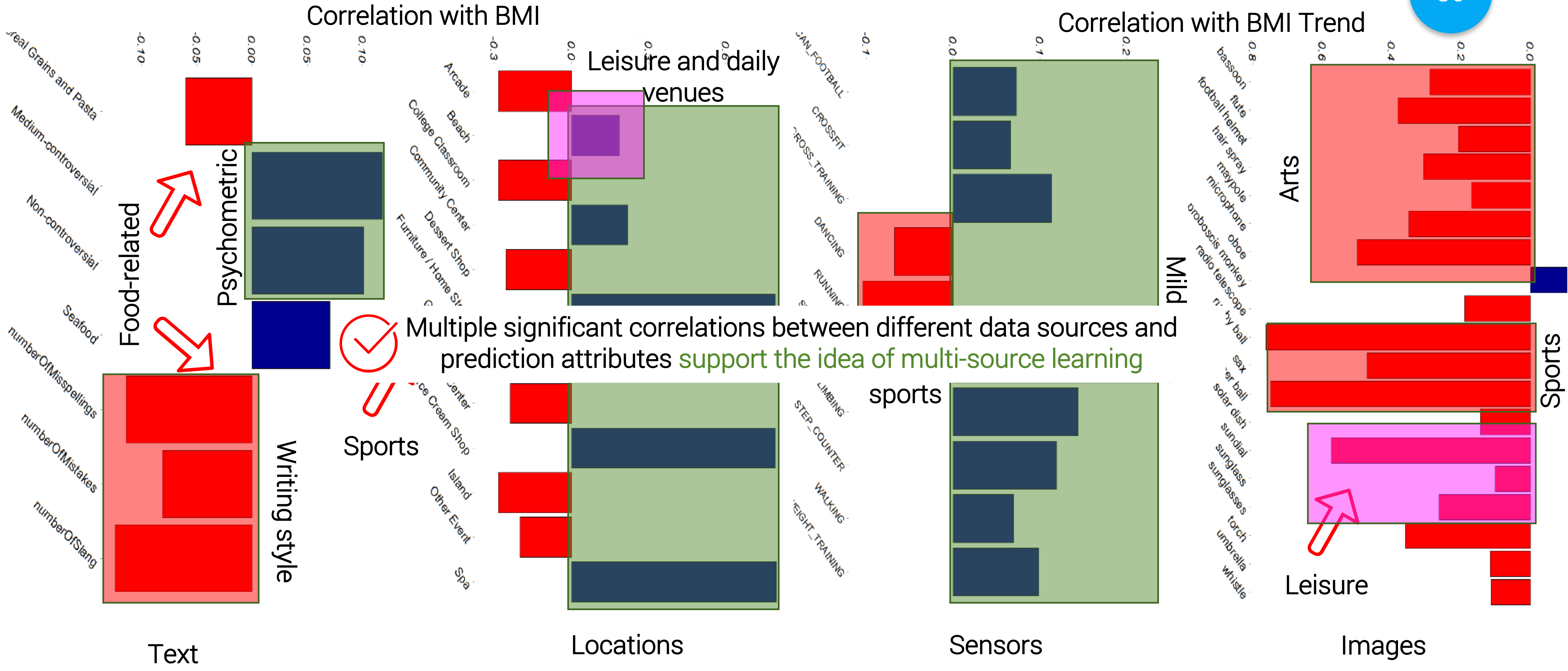
$$r = \frac{Cov(x,y)}{\sqrt{Var(x)Var(y)}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}},$$

where x_i, y_i are i-th population samples and $\overline{x}$, $\overline{y}$ are the population means.



E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Individual Profile Analytics: Correlation with individual attributes



Multiple significant correlations between different data sources and prediction attributes support the idea of multi-source learning

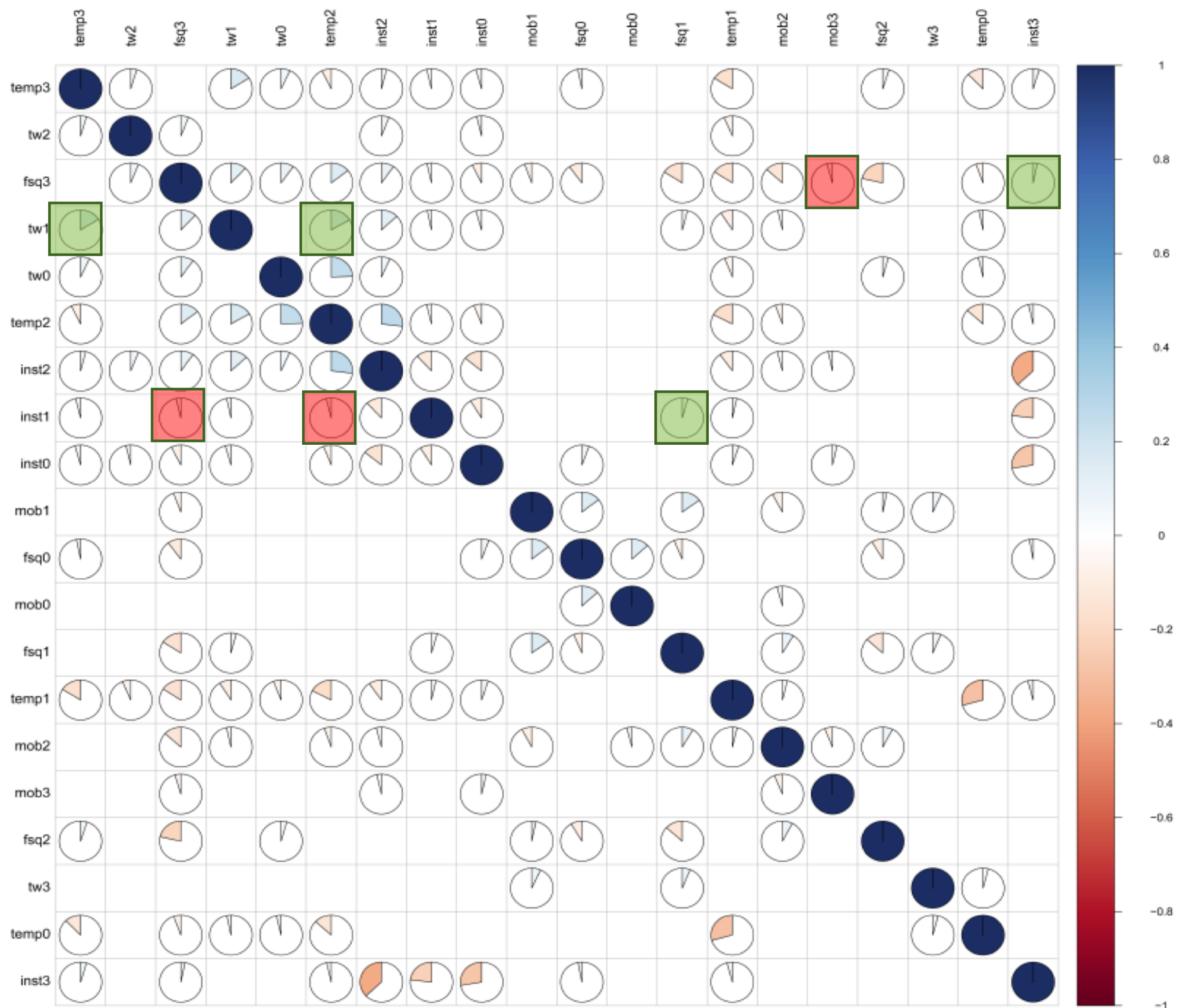
E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Group Profile Analytics: Correlation between user communities

ID	Name	Profile (Terms)
	Twitter	
1	Work/Active	#misspellings, #words/tweet, work
	Foursquare	
1	Shop/Dine	Mall, Chinese Restaurant, Fast Food, Japanese Restaurant, Food Court
3	Business/Travel	Airport, Café, Hotel, Bakery, Bus Station
	Instagram	
1	Beach/Leisure	Lotion, Sunscreen, Sea slug, Swan
3	Home/Family	Gong, Backpack, Dog, Pot, Library (Book)
	Temporal	
2	Misc.	Weekends 12 – 15
3	Misc.	Weekends 6 – 12, Weekdays 18 – 24
	Mobility	
3	Explorers	#AOI outliers

- Shoppers/Diners & Leisure Lifestyle explains positive correlation
- Business users tend to spent more time on family activities
- Lack of business activities by Shopper & Diners explains negative relation with Business community

Significant overlap in community detection results show the ability of data source to compliment each other.



“

It is a mistake to try to look too far ahead…

- Winston S. Churchill

Future Work

Future Work:

- Utilization of Deep Neural Networks for Individual Multi-View Supervised Learning
- Utilization of Temporal Data Component for Modeling Multi-View Temporal Dependencies.
- Exploring Semi-Supervised Learning for Better Multi-View Clustering

<Preliminary Results>

Multi-View Neural Temporal Learning

for Personality Profiling

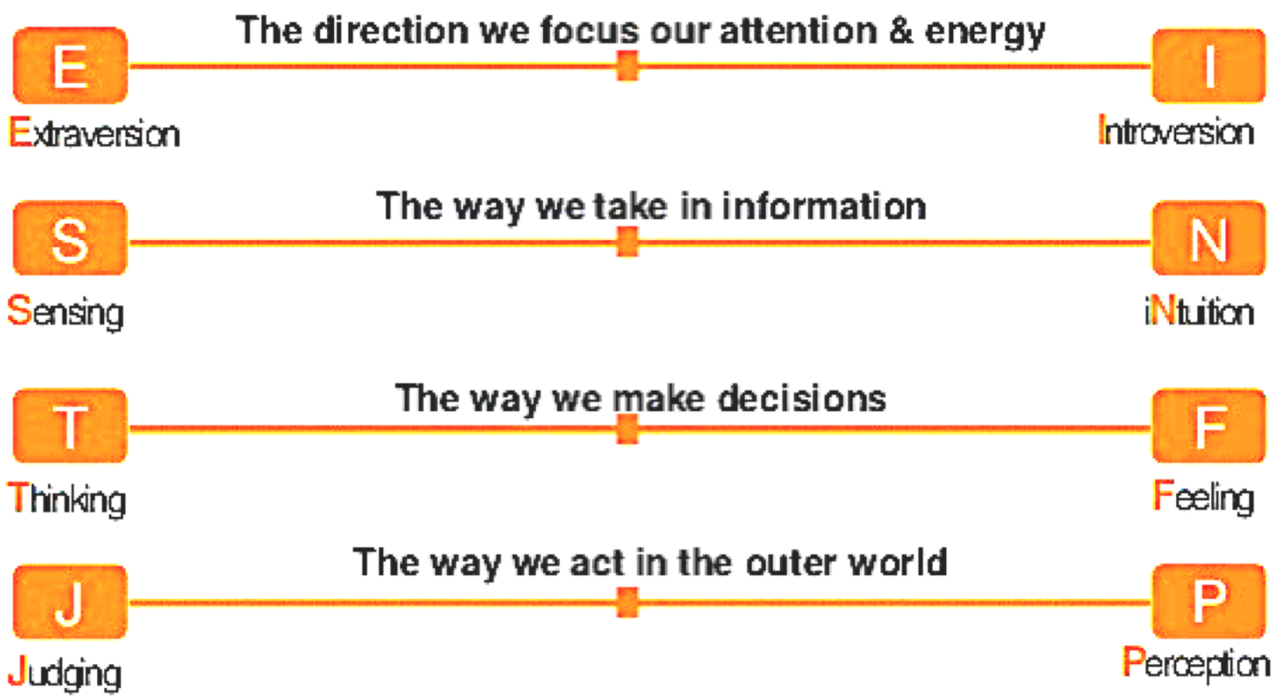
Preliminary Study on Multi-View Temporal Learning for Personality Profiling

Relationships *Life challenges*

Personality
The powerful description of an individual that affects:

Emotions *Behavior*

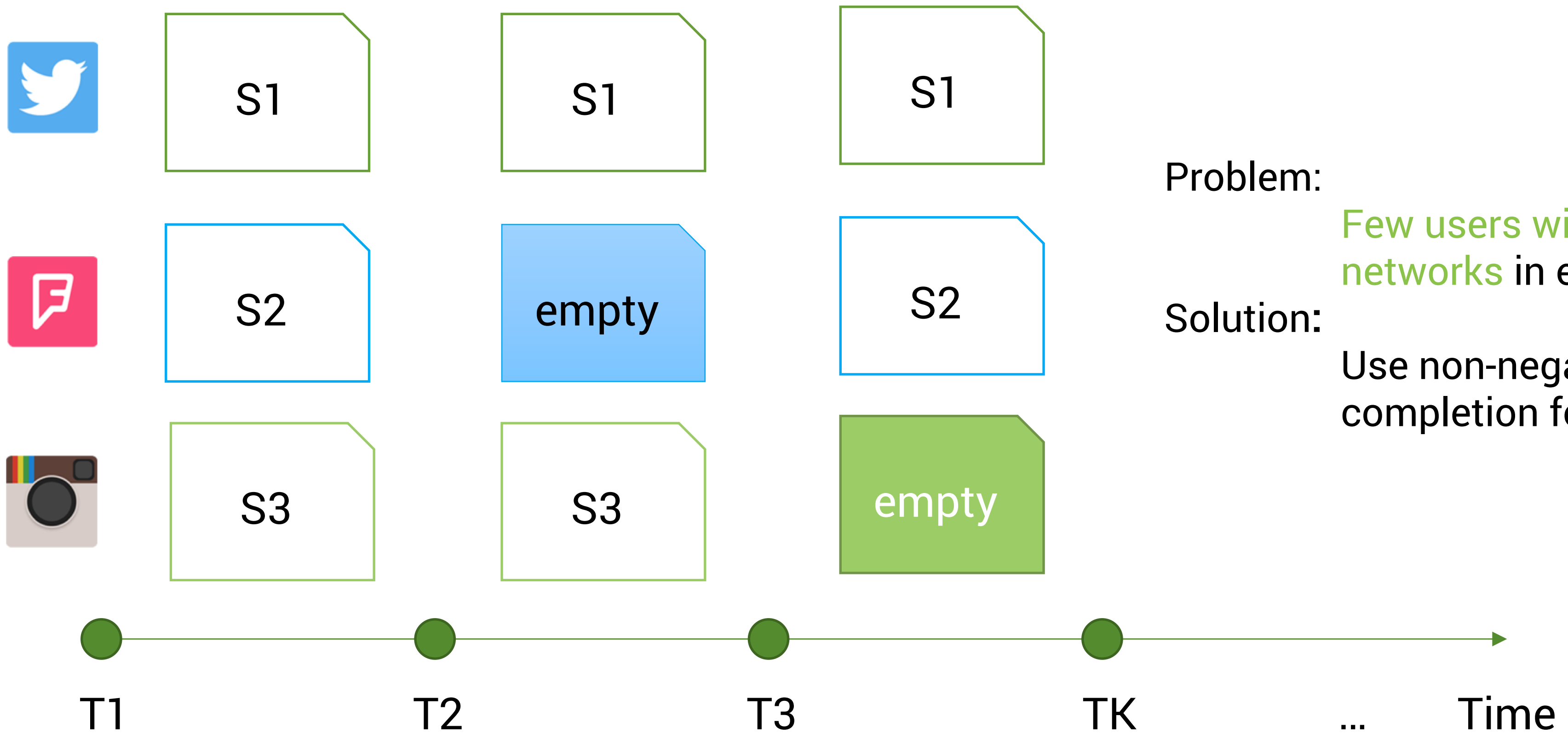
MBTI Personality Scale



E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Preliminary Study: Dealing with missing data

Divide user activity into K time periods



Problem:

Few users with data from all three social networks in each window

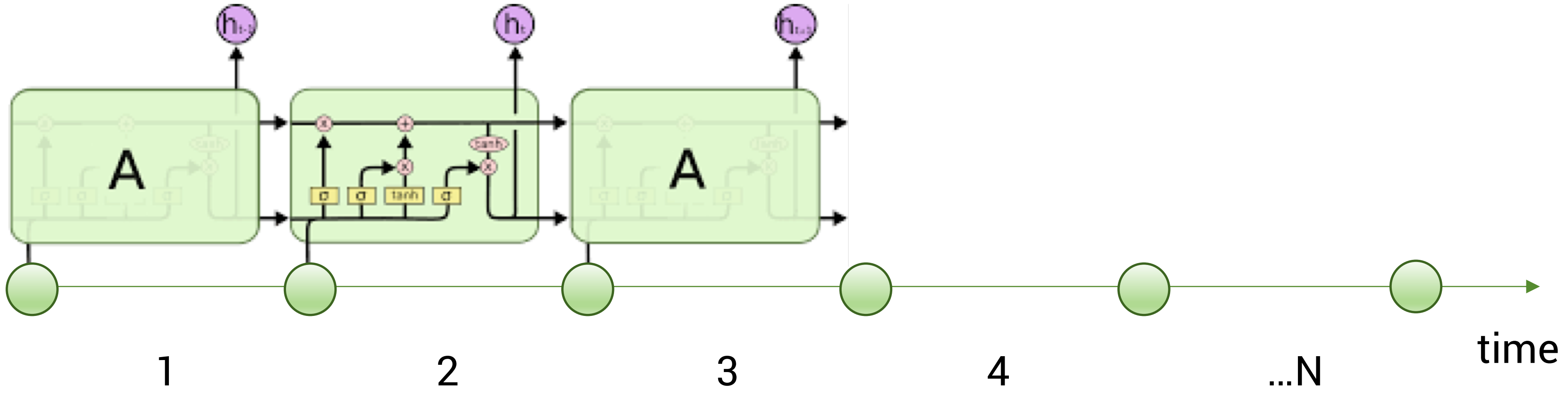
Solution:

Use non-negative matrix factorization for data completion for every k-period window

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Preliminary Study: Incorporation of temporal aspect

- Model: LSTM for time periods. Neural network that will take into account multiple modalities and temporal aspect of the data.



Experiment. Test the temporal model on different data modalities

and their combinations (use only complete data)

Preliminary Results: Different Data Source Combinations

Table 3: The results for combination of data of different modalities, the best window size is shown in brackets. Evaluation metrics: Macro F_1.

	E/I	S/I	T/F	J/P
Text (T)	0,362 (2)	0,456 (8)	0,380 (6)	0,369 (9)
Media (M)	0,349 (2)	0,409 (5)	0,493 (9)	0,372 (7)
Location (L)	0,531 (4)	0,43 (4)	0,493 (9)	0,394 (5)
T, M	0,349 (3)	0,494 (3)	0,465 (9)	0,443 (2)
M, L	0,511 (4)	0,505 (5)	0,496 (4)	0,495 (3)
T, L	0,521 (5)	0,518 (3)	0,495 (3)	0,496 (7)
T, M, L	**0,542 (6)**	**0,528 (9)**	**0,534 (7)**	**0,530 (5)**

Table 2: Results obtained by non-temporal baselines and PLSTM framework. Evaluation metrics: macro F_1

	Gradient Boosting	Logistic Regression	Naive Bayes	LSTM
E/I	**0,615**	0,600	0,440	0,541
S/N	0,495	0,425	0,33	**0,524**
T/F	**0,570**	0,435	0,440	0,534
J/P	0,510	0,395	0,470	**0,530**

Table 4: Most important features for MBTI categories

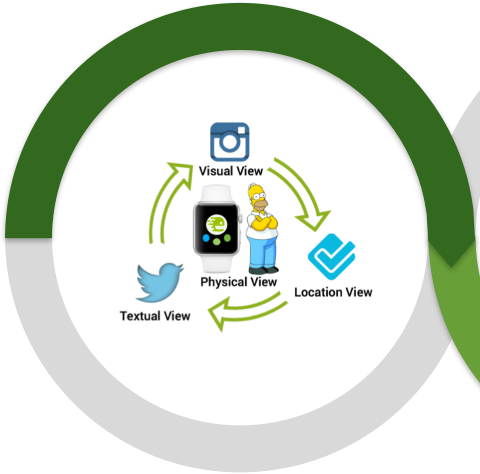
Category	E/I	S/N
Heuristic features	Num. of hashtags Dairy food	Num. of misspellings Time 12 am - 3 am
LIWC features	"past" "anxiety"	"anxiety" "achievement"
Location features	Car rental locations Coworking spaces	Financial services Health food stores

Category	T/F	J/P
Heuristic features	Num. of repeated chars Num. of misspellings	Num. of misspellings Time 12 am - 3 am
LIWC features	"affection" "feelings"	"motion" "affection"
Location features	Office supplies places Food places	Shops Volleyball courts

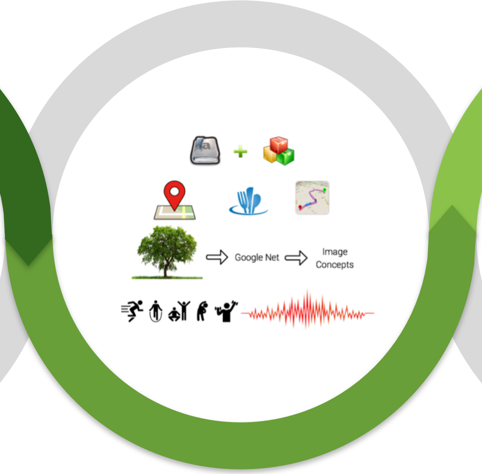
Having here or take away?

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

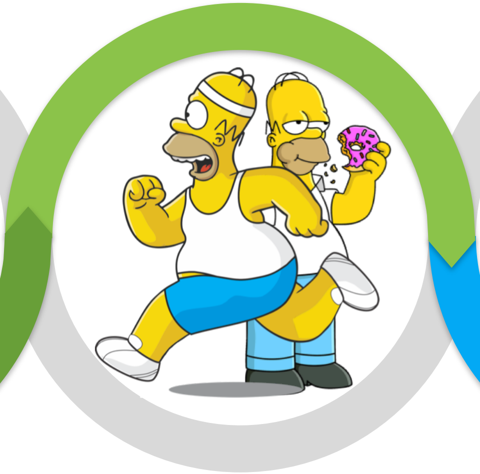
Main takeaways



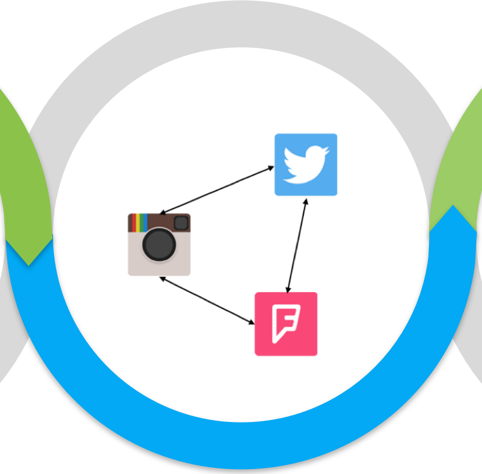
Multi-Source Data Describes Users From Multiple Perspectives



Learning From Multi-Source Data Allows For Achieving Significant Improvement of Classification



Join Learning From Social Media Data And Sensor Data Helps To Improve Wellness Profiling Performance



Considering Inter-Source Relationship is Useful for Multi-Source Clustering and Recommendation

Future Work: Multi-View Temporal Neural Learning

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Top 10 involved publications

Farseev, A., Nie, L., Akbari, M., & Chua, T. S. (2015 June). **Harvesting multiple sources for user profile learning: a big data study.** In Proceedings of the 5th ACM Conference on Multimedia Retrieval (pp. 235-242). ACM.

Farseev, A., Kotkov, D., Semenov, A., Veijalainen, J., & Chua, T. S. (2015 June). **Cross-social network collaborative recommendation.** In Proceedings of the 7th ACM Conference on Web Science (p. 38-39). ACM.

Farseev, A., Akbari, M., Samborskii, I., & Chua, T. S. (2016). **360° user profiling: past, future, and applications.** by Aleksandr Farseev, Mohammad Akbari, Ivan Samborskii and Tat-Seng Chua with Martin Vesely as coordinator. ACM SIGWEB Newsletter, (Summer), 4.

Farseev, A., Samborskii, I., & Chua, T. S. (2016 October). **bBridge: A Big Data Platform for Social Multimedia Analytics.** In Proceedings of the 25th ACM Conference on Multimedia (pp. 759-761). ACM.

Farseev, A., & Chua, T. S. (2017 February). **TweetFit: Fusing Multiple Social Media and Sensor Data for Wellness Profile Learning.** In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (pp. 87-93). AAAI.

Farseev, A., & Chua, T. S. (2017). **Tweet can be Fit: Integrating Data from Wearable Sensors and Multiple Social Networks for Wellness Profile Learning.** ACM Transactions on Information Systems (TOIS).

Farseev, A., Samborskii, I., Filchenkov, A., & Chua, T. S. (2017 August). **Cross-Domain Recommendation via Clustering on Multi-Layer Graphs.** In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval.

Chowdhury, A.-K., Farseev, A., Chakraborty, P.-R., & Tjondronegoro, D. (2017 December). **Automatic Classification of Physical Exercises from Wearable Sensors using Small Dataset from Non-Laboratory Settings** In Proceedings of the First IEEE Life Science Conference. IEEE.

Buraya, K., Farseev, A., Filchenkov, A., & Chua, T. S. (2017 February). **Towards User Personality Profiling from Multiple Social Networks.** In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. AAAI.

Nie, L., Zhang, L., Wang, M., Hong, R., Farseev, A., & Chua, T. S. (2017). **Learning user attributes via mobile social multimedia analytics.** ACM Transactions on Intelligent Systems and Technology (TIST), 8(3), 36.

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

This PhD would not be possible without…

Prof. Stephane Bressan

Internship @ NUS &
Thesis Advisory

Prof. Kian-Lee Tan

Internship @ NUS

Mr. Thomas Kister

Internship @ NUS

Dr. Liqiang Nie

Ex P.Doc. @ LMS,
For help with my first paper

Dr. Mohammad Akbari

Ex Ph.D. @ LMS,
For multiple fruitful discussions

Mr. Ivan Samborskii

For helping me to make
bBridge Project happen

Prof. Andrey Filchenkov

Main collaborator @ ITMO

Prof. Paolo Rosso

Main collaborator @ UPV

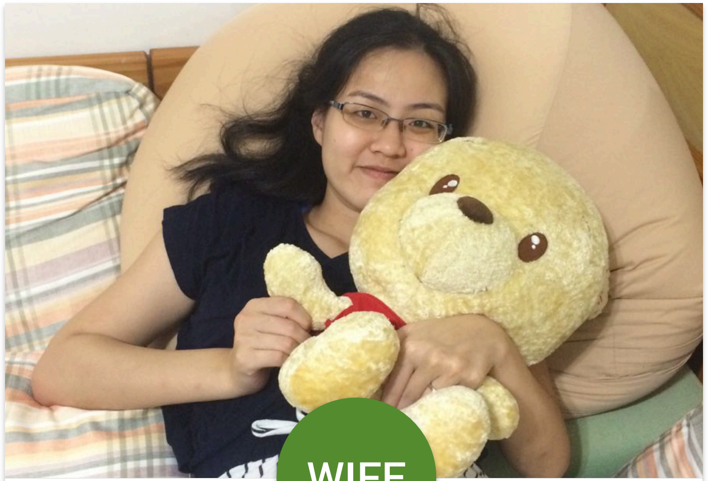
Mr. Choon Meng Seah

For shaping bBridge
Project

Prof. Mun Choon Chan

Thesis Advisory

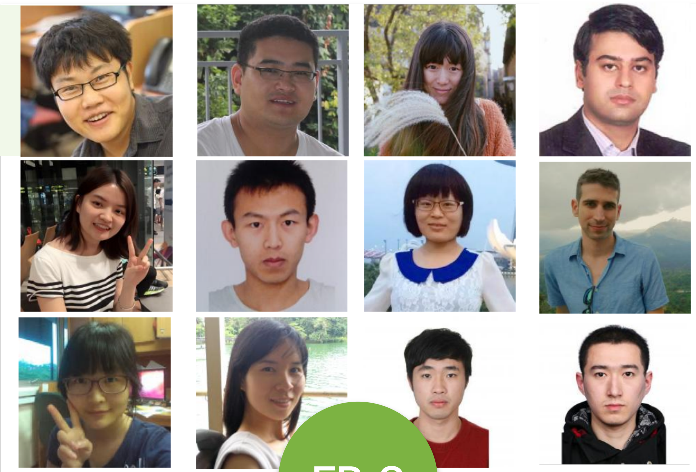
This PhD would not be possible without…



WIFE

Almost Dr. Joy Chu

Extremely Supportive Wife



FR-S

LMS Group

Friends

BOSS

Prof. Tat-Seng Chua

Who has guided me though the Jungle of Research
and Deep Russian Forests

Thank You

Questions?

Backup I

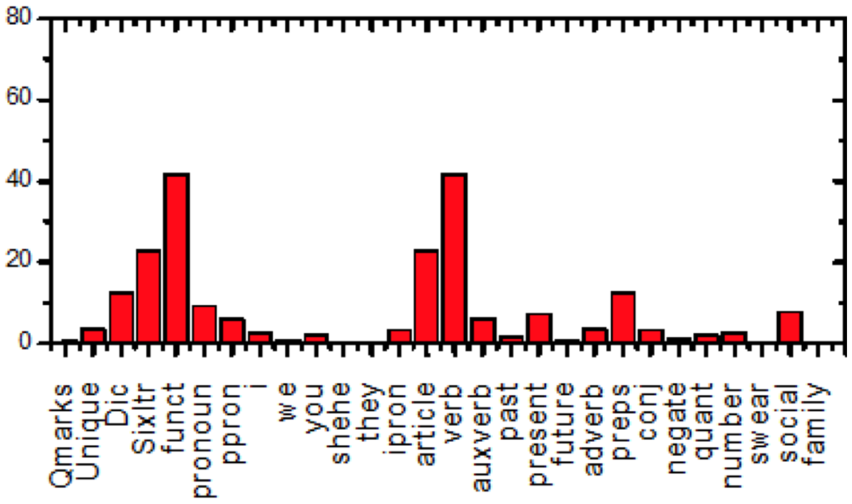
Multi-View Features

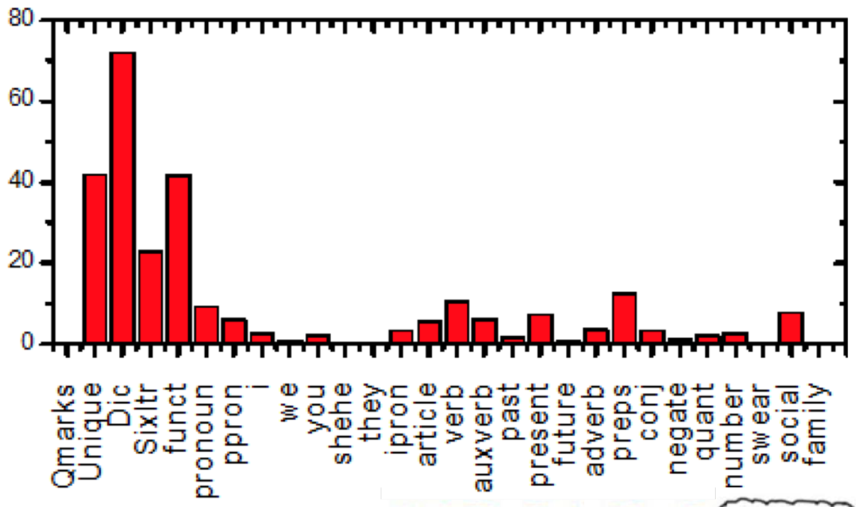
E-Mail: farseev@u.nus.edu | Website: http://farseev.com

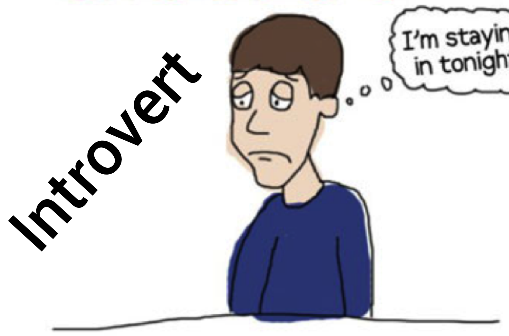
Textual data representation (1): LIWC

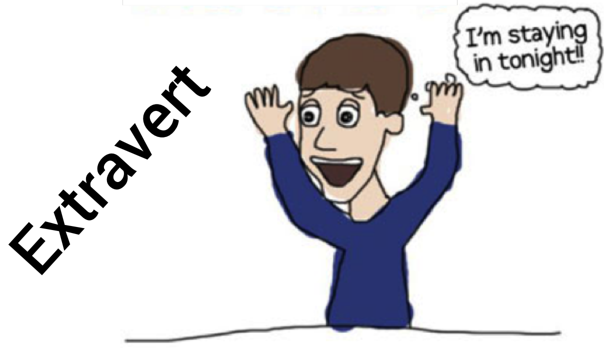
Pennebaker's LIWC Descriptors

69 LIWC Categories









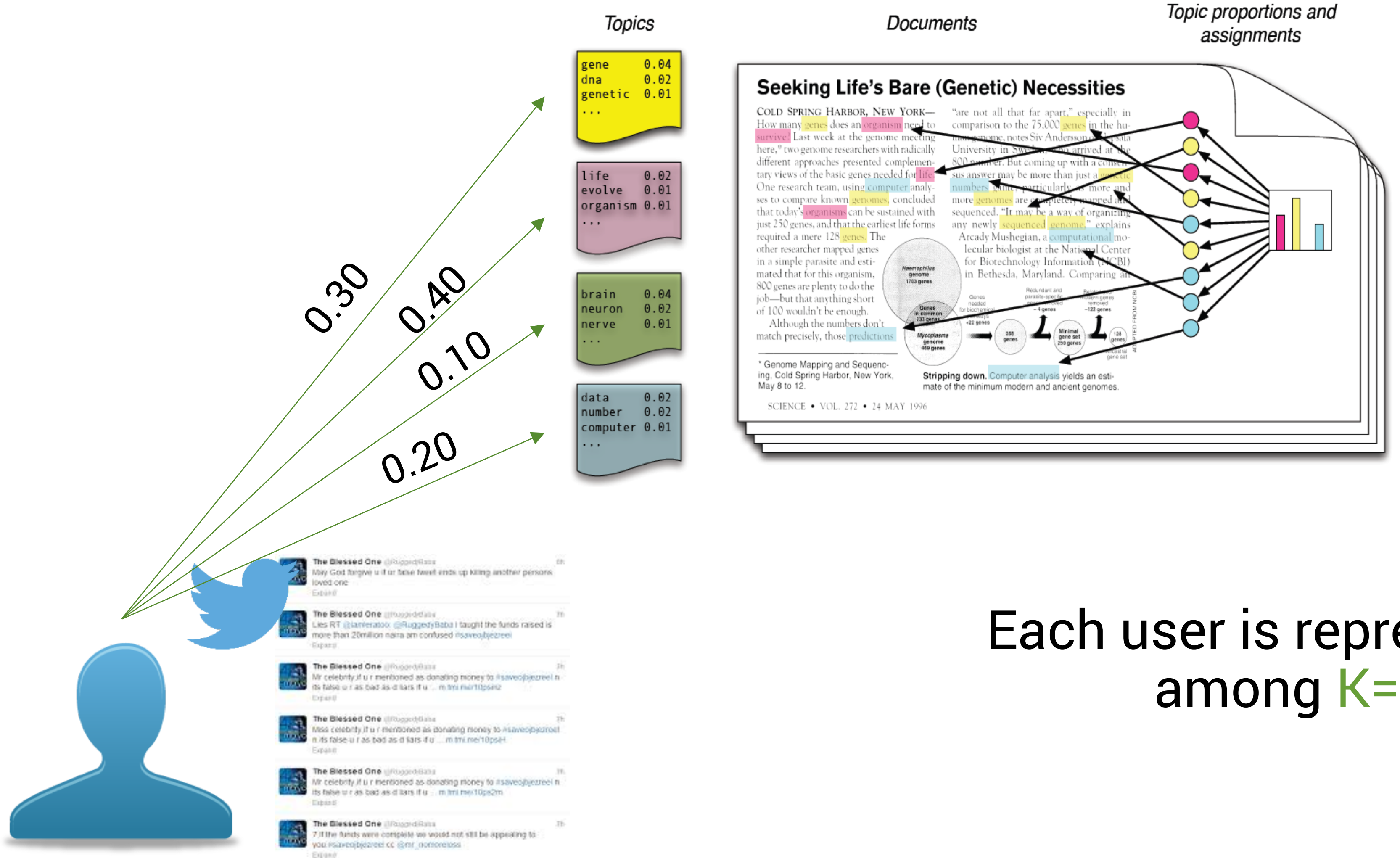
Each user is represented as distribution among K=69 LIWC categories

* Pennebaker, J. W. (2011). The secret life of pronouns.

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Textual data representation (2): Latent Topic Modeling

Latent Dirichlet Allocation



Each user is represented as distribution among K=50 latent topics

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Textual data representation (3): Writing style features

Heuristically inferred and found to be useful…

- # hashtags
- # repeated chars
- Avg. sentiment score
- # slang words
- # emotion words
- # Misspellings
- # URLs
- # emoticons
- # Unknown words
- # User mentions
- Avg. sentiment level
- Avg. # terms

Image Data Representation: Image Concept Detection

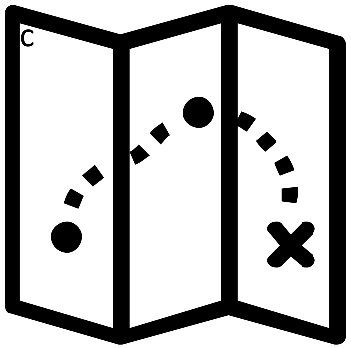
Going deeper with convolutions with Google Net.

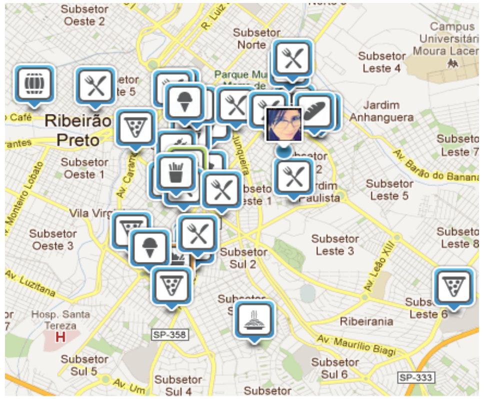
Each user is represented as distribution among K=1000 image concepts

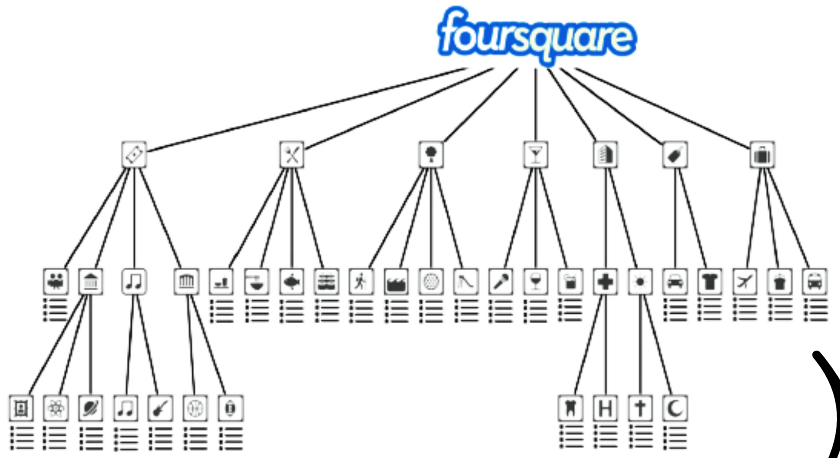
*Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-9).

Location Data Representation (1): Utilizing Venue Semantics

Venues are not just GEO points, but multidimensional objects with rich semantics







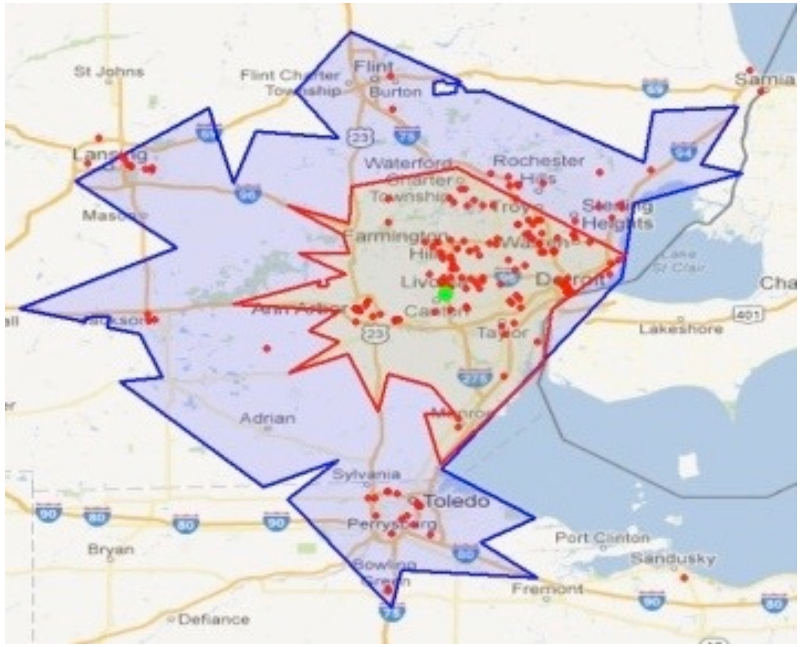
For case when user performed check-ins in two restaurants and airport
but did not perform check-ins in other venues:

	Category 1	…	Category Restaurant	…	Category Airport	…	Category K
User 1	0	0	2	0	1	0	0
…	*	*	*	*	*	*	*
User N	*	*	*	*	*	*	*

Each user is represented as distribution among 764 venue categories

Location Data Representation (2): Mobility Features

- # of posts in 8 daytime intervals
- # of AOIs
- median AOI size
- # of AOI outliers
- median distance between AOIs



User's Area of Interests (AoI) – area the most frequently visited by the user. *In fact, it is the convex hull over the dense user's check-in region (DB-Scan cluster).*

*Qu, Y., & Zhang, J. (2013, May). **Trade area analysis using user generated mobile location data**. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 1053-1064). International World Wide Web Conferences Steering Committee

E-Mail: farseev@u.nus.edu l Website: http://farseev.com

Sensor Data Representation

Heuristically inferred and found to be useful...

- # workouts
- Avg. heart rate (HR)
- Avg. ascent/descend
- Age / Gender
- HR
- Oxygen consumption

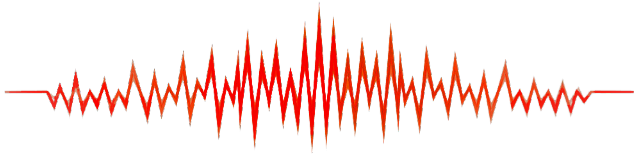
Each user is represented as distribution among 500 + 95 + 6 = 601 sensor features

Distribution over 95 sport categories

Furrier’s Spectrum

500 frequency bins

$$F_{Max} = 0.5\ dHz$$



Backup II

Demographic Profiling Training Details

Age and Gender Prediction: Details

Some details that you may find in the paper

- Bias of estimated ages does not exceed ±2.28 years. It is thus reasonable to use the estimated age for age group prediction task.
- We have adopted SMOTE* oversampling to obtain balanced age-group labeling
- By performing 10-fold cross validation, we determine the optimal number of constructed random trees for each classifier with iteration step equal to 5 as 45, 25, 35, 40, 105 random trees for Random Forest Classifiers learned based on location, LIWC, heuristic, LDA 50, and image concept features respectively.
- We jointly learn the l_i model "strength" coefficient by performing "Hill Climbing" optimization** with step 0.05 and 1000 random restarts.

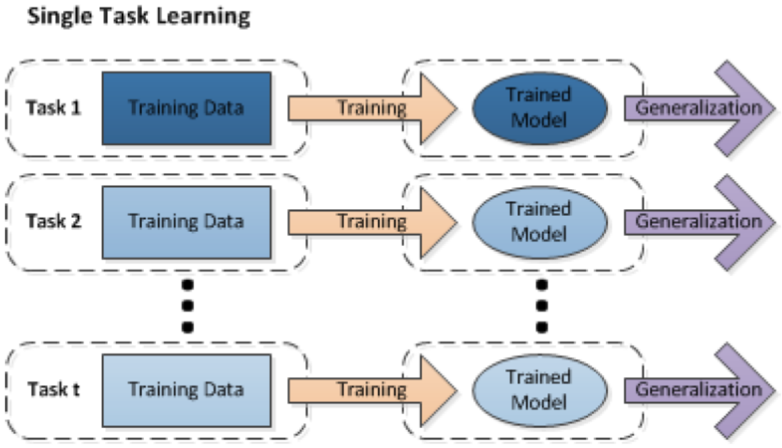
*N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research, 2002.

**An iterative algorithm that starts with an arbitrary solution to a problem, then attempts to find a better solution by incrementally changing a single element of the solution. If the change produces a better solution, an incremental change is made to the new solution, repeating until no further improvements can be found.

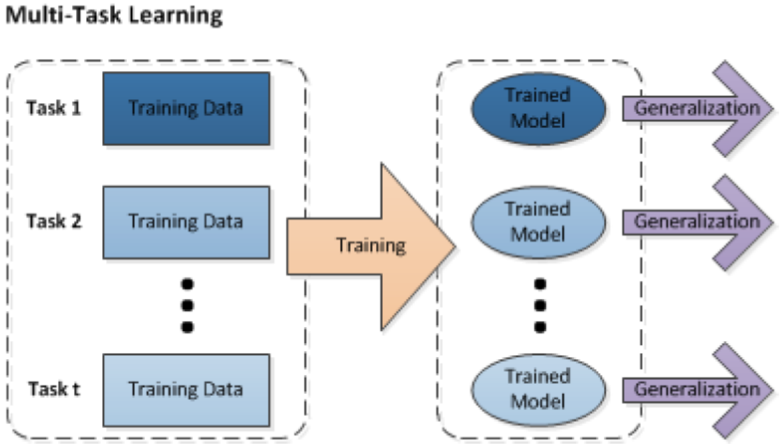
Backup III

Multi-Task Learning

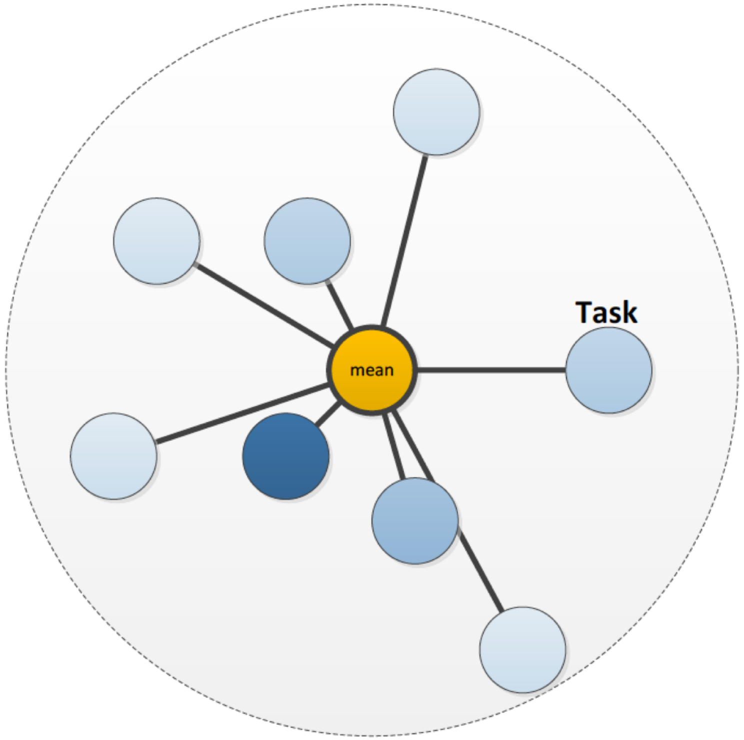
Inspiration: Multi-Task Learning (MTL)



Multi-task Learning is different from single task learning in the training (induction) process.



Inductions of multiple tasks are performed simultaneously to capture intrinsic relatedness.



Example of Regularized MTL:
Assumption: task parameter vectors of all tasks are close to each other.

Optimization of Multi-Source Multi-Task Objective

Smooth Reformulation of the objective:

$$f(\mathbf{W}) = \underset{\mathbf{W}\in\mathbf{Z}}{\arg\min}\ \Psi(\mathbf{X},\mathbf{W},\mathbf{Y}) + \mu\Omega(\mathbf{W}) \quad s.t.\ \mathbf{Z} = \left\{\mathbf{W} \mid \|\mathbf{W}\|_{2,1} \le z\right\},$$

Nesterov's Optimization Solution on
Each Step, S_i computed from past solutions:

$$\mathbf{W}_{i+1} = \underset{\mathbf{W}}{\arg\min}\ M_{\gamma_i,S_i}(\mathbf{W}),$$

$$M_{\gamma_i,S_i}(\mathbf{W}) = f(\mathbf{S}_i) + \langle \nabla f(\mathbf{S}_i), \mathbf{W} - \mathbf{S}_i \rangle + \frac{\gamma_i}{2}||\mathbf{W} - \mathbf{S}_i||^2,$$

$$\mathbf{S}_i = \mathbf{W}_i - \alpha_i(\mathbf{W}_i - \mathbf{W}_{i-1}).$$

Backup IV

Wellness Profiling Baselines

Wellness Profiling Baselines

Wellness Profiling Baselines

- **aMTFL$_2$** [85] — the $\ell_{2,1}$ norm regularized multi-task learning with the least squares lost and $\alpha = 0.5$.
- **iMSF** [160] — the sparse $\ell_{2,1}$ norm regularized multi-source multi-task learning, with $\alpha = 0.4$;
- **MSESHC** — weighted ensemble proposed in [47]. The modality weights $\mathbf{s}$ were learned by Stochastic Hill Climbing (SHC): $\mathbf{s} : \{0.75, 0.2, 0.25, 0.45, 0.2, 0.3, 0.45, 0.2\}$; for venue categories, image concepts, behavioral text, LDA 50 text, sport categories, sensors freq. bins, workout statistics, and mobility features, respectively.
- **TweetFit** [44] — multi-source multi-task learning framework "TweetFit" (equivalent to **M^{2}WP** framework, trained without inter-category relatedness regularization (Equation 4.1)).

[44] A. Farseev and T.-S. Chua. Tweetfit: Fusing multiple social media and sensor data for wellness profile learning. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. AAAI, 2017.

[47] A. Farseev, L. Nie, M. Akbari, and T.-S. Chua. Harvesting multiple sources for user profile learning: a big data study. In *Proceedings of the ACM International Conference on Multimedia Retrieval*. ACM, 2015.

[85] J. Liu, S. Ji, and J. Ye. Multi-task feature learning via efficient l 2, 1-norm minimization. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence*. AUAI Press, 2009.

[160] L. Yuan, Y. Wang, P. M. Thompson, V. A. Narayan, and J. Ye. Multi-source learning for joint analysis of incomplete multi-modality neuroimaging data. In *Proceedings of the 18th International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, 2012.

Backup V

Recommender Evaluation Metrics

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Recommendation Evaluation Metrics

Average Precision

$$AP@p = \frac{1}{\sum_{i=1}^{p} r_i} \sum_{i=1}^{p} r_i \left(\frac{\sum_{j=1}^{i} r_j}{i} \right), r_i = \begin{cases} 1, & \text{item } i \text{ is relevant} \\ 0, & \text{otherwise.} \end{cases}$$

Normalized Discounted Cumulative Gain

$$NDCG@p = \frac{DCG@p}{IDCG@p}, DCG@p = \sum_{i=1}^{p} \frac{2^{rel_i}}{\log_2(i+1)}, rel_i = \frac{Cat_i}{N_{Cat}}$$

Backup VI

Recommender System Baselines

E-Mail: farseev@u.nus.edu | Website: http://farseev.com

Venue Category Recommendation Evaluation Baselines

Popular (POP) —recommendation based on user's past experience

Popular All (POP All) —recommendation based on experience of all users

Multi-Source Re-Ranking (MSRR) — linearly combines recommendation results from all data modalities

Nearest Neighbor Collaborative Filtering (CF) — recommendation based on top k most similar Foursquare users

Early Fusion (EF) — fuses multi-source data into a single feature vector

SVD++ — makes use of the "implicit feedback" information

FM— brings together the advantages of different factorization-based models via regularization.

$\mathbf{C^3R}$ - $\mathbf{\hat{L}_i}$ — $\mathrm{C^3R}$ recommendation without inter-layer regularization

$\mathbf{C^3R}$ - $\mathbf{\hat{L}_i}$ - $\mathbf{\hat{L}_{Mod}}$ — $\mathrm{C^3R}$ recommendation without inter-layer regularization and sub-space regularization

$\mathbf{C^3R}$−$\boldsymbol{Comm}$ — $\mathrm{C^3R}$ recommendation without user community extraction

$\mathbf{C^3R}$ (DBScan) — $\mathrm{C^3R}$ recommendation, where user communities are detected by Density-Based clustering (DBScan)

$\mathbf{C^3R}$ (x-means) — $\mathrm{C^3R}$ recommendation, where user communities are detected by x-means clustering

$\mathbf{C^3R}$ (Hierarchical) — $\mathrm{C^3R}$ recommendation, where user communities are detected by Hierarchical Clustering

$\mathbf{C^3R}$ — Our Approach

Backup VII

Weighted Cohen’s Kappa

Weighted Cohen's Kappa

$$\kappa_w = \frac{p_{o(w)} - p_{e(w)}}{1 - p_{e(w)}},$$

$$p_o(w) = \sum_i^{\kappa} \sum_j^{\kappa} w_{ij} p_{ij} \qquad \mathbb{W} = (w_{i,j}) \in \mathbb{R}^{k \times k} \qquad p_e(w) = \sum_i^{k} \sum_j^{k} w_{ij} p_{i.} p_{.j}$$

Kappa Statistics	Strength of Agreement
<0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

Backup VIII

Personality Profiling

Mayers-Briggs Type Indicator (MBTI)

MBTI – the typology, which is designed to exhibit psychological preferences in how people perceive the world around them and distinguishes 16 personality types.