


**Aleksandr Farseev**

farseev@u.nus.edu


**Tat-Seng Chua**

chuats@comp.nus.edu.sg

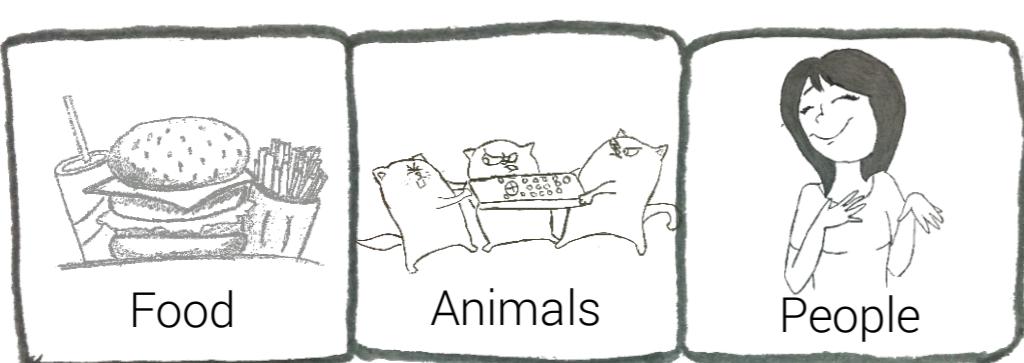
## I Harvesting Multiple Social Media...



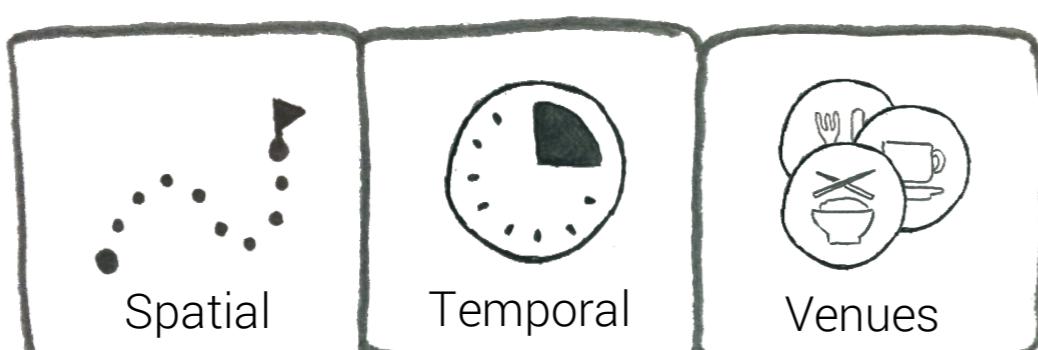
## II Representing Social Media Sources...



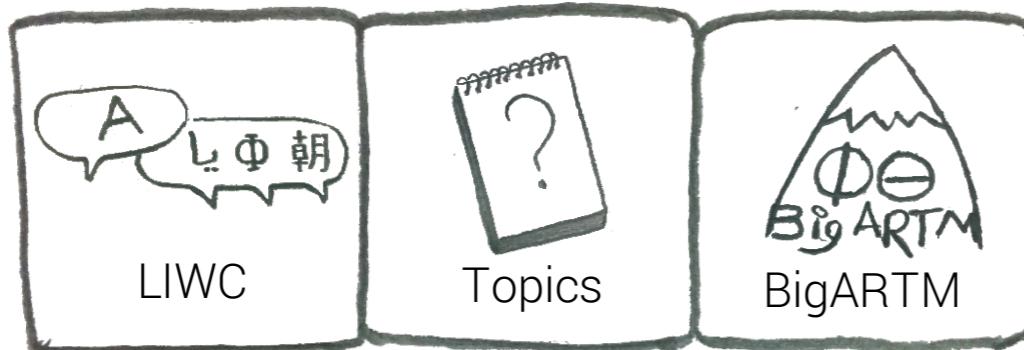
Image Concepts



Mobility &amp; Venue Semantics



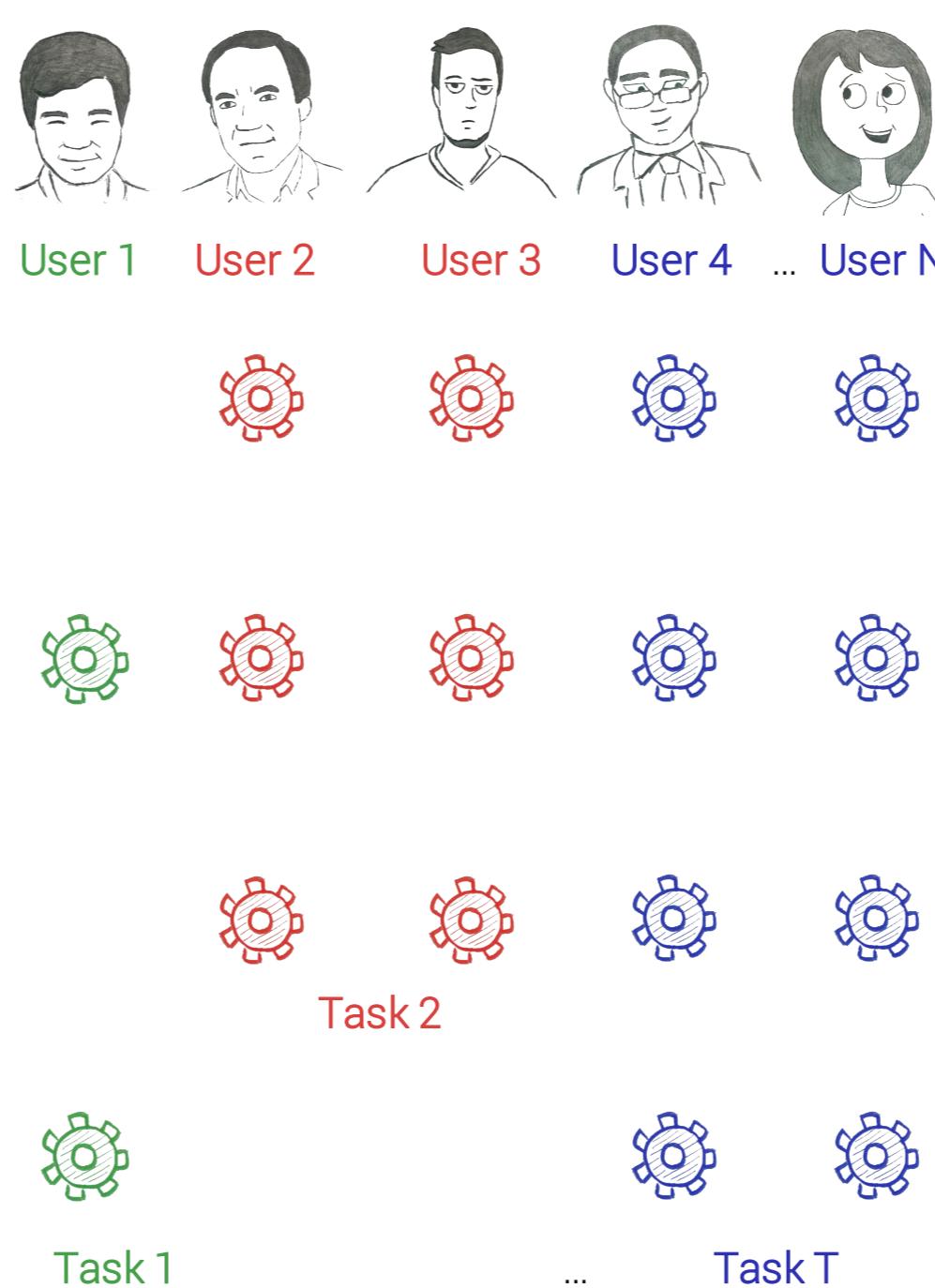
Linguistic Features &amp; Topics



Sensor Features



## III Multi-Source Multi-Task Learning for BMI Inference...



Data source combinations are Multi-Task learning “tasks”:

$$\arg \min_w \frac{1}{T} \sum_{t=1}^T \frac{1}{N_t} \sum_{i=1}^{N_t} \log \left( 1 + e^{-y_i f_t(x_i^t; w^t)} \right) + \gamma \sum_{s=1}^S \sum_{f=1}^{F_s} \|w_{p(s,f)}\|$$

- $t: 1 \dots T$  – multi-task learning tasks
- $N_t$  – number of data samples in the task  $t$
- $s: 1 \dots S$  – data sources (Twitter, Instagram, Foursquare, Sensors)
- $F_s$  – number of features in the data source  $s$
- $W = (w^1, w^2, \dots, w^T)$  – block matrix of the model parameters
- $x_i^t$  –  $i$ -th data sample of the task  $t$
- $y_i^t$  –  $i$ -th label of the task  $t$
- $f_t(x_i^t; w^t) = x_i^t w^t$
- $p(s, f)$  – index function that denotes all model parameters of the  $f$ -th feature from the source  $s$

Restricts same sources from different tasks to use same set of features

## IV Evaluating TweetFit Framework...

Data Source Combinations

TweetFit On Source Combinations	BMI (8 Cat.)
<b>F<sub>1Macro</sub></b>	
Visual (V)	0.077
Ven. Sem. & Mob. (VSM)	0.137
Sensors (S)	0.155
<b>Textual (T)</b>	<b>0.178</b>
V + S	0.186
V + T	0.166
V + VSM	0.157
T + VSM	0.179
<b>S + VSM</b>	<b>0.191</b>
<b>S + T</b>	<b>0.191</b>
V + T + VSM	0.163
S + T + V	0.164
S + T + VSM	0.205
<b>S + VSM + V</b>	<b>0.221</b>
<b>All Data Sources</b>	<b>0.246</b>

Multi-Source Learning Baselines

Method	BMI (8 Cat.)	BMI Tr.
	F <sub>1Macro</sub>	F <sub>1</sub>
MSE	0.142	0.644
R.Forest	0.169	0.480
iMSF	0.172	0.649
MTFL	0.184	0.710
<b>TweetFit</b>	<b>0.202</b>	<b>0.718</b>

Concluding...

In this work, we presented the first study on wellness profile learning from sensor and social media data, which was handled by training the “TweetFit” framework to infer BMI category and BMI Trend wellness attributes. “TweetFit” can handle data incompleteness and perform attributes inference from sensor and social media data simultaneously. To facilitate further research, we released the multi-source multimodal dataset NUS-SENSE\*. It can be used for research on multi-source user profiling, multi-view timeline analysis, and cross-network user identification.