Multiple Demographic Attributes Prediction in Mobile and Sensor devices

Yiwen Jiang   Wei Tang   Neng Gao

State Key Laboratory of Information Security, Chinese Academy of Sciences
Attribute Prediction

- Given public information of some users
- Infer private attributes of some target users
Attribute Prediction

Intelligent marketing and automatic advertising

### User Age
- 25-30 (33.1%)
- 15-24 (30.3%)
- 31-35 (18.4%)
- >40 (10.2%)
- >15 (1%)
Attribute Prediction

Friendly human-computer interaction

Hello, handsome boy.

Hello, beautiful girl.

is male?

is female?
Existing Attribute Prediction

Text, Image, Video, Network behaviors, Relationships......

[Social media logos: Facebook, Google+, Twitter, YouTube]
Mobile and Sensor devices

**mobile devices**: pedometer, gyroscope, accelerometer, vibrometer, magnetometer

**wearable devices** (Fitbit, Apple Watch, and Android Wear): pedometer, accelerometer, heart rate monitor

Sensing devices generate trillions of sensor data points per year, including rich signals such as **step count variability**

Extracting knowledge and emerging patterns from mobile and sensor devices for user attribute prediction is a nontrivial task.
Our Attribute Prediction

Step count come from pedometer in mobile phone
Roadmap

- Problem Formalization
- CANEA Model
- Evaluation
- Conclusion
### Problem Formalization

#### Multi-task Multi-class problem

| Attributes | Value  |
|------------|--------|
| Gender     | male, female |
| Age        | young, adult, Middle-age, old |

#### User representation

$$ S : [s_1, s_2, \ldots, s_n] $$

#### Problem

$$ \mathcal{T} : S \rightarrow \{y_1, y_2, \ldots y_t\} $$
CANE Model

Different model architecture
CANEA Model

Correlation Aware Neural Embedding with Attention

Data

Separated Embedding Layer

Stacked Bidirectional RNN with Attention

Correlation Learning Layer

Stacked Bidirectional RNN with Attention

BiLinear Model

Full Connection for Age to Gender Correlation Learning

Gender

Full Connection for Gender to Age Correlation Learning

Age

Multi-Task Prediction Layer

BiLinear Model

Softmax Classifier

Softmax Classifier

UCAS
Stacked Bidirectional RNN: temporality, longer term dependencies, time connection

Attention Mechanism: different days in each week play differential roles in tasks

Separated embedding branches representation:

\[ w_i = \text{softmax}(\tanh(v_ix + b_i)) \]

\[ E_i = \mathcal{L}_{\theta_i}(w_ix) \]
Correlation Learning Layer

The key components of this correlation learning pipelines are a **full connection network** and a **bilinear mixer**

**Correlation representation for i-th task:**

\[ C_i = \mathcal{F}_{\phi_i}(w_i \bar{E}_i) \]

**Combination function:**

\[ M_i = E_i W_i C_i \]
Multi-Task Prediction Layer

Prediction probability of a given user:

\[ p(y_i|S) = \text{softmax}(O_iM_i) \]

Loss function for i-th task:

\[ \text{Loss}_i = - \sum_{j=1}^{m} \log p(y_{i,j}|S_j) \]

The total multi-task loss function:

\[ \text{Loss} = \sum_{i=1}^{t} \lambda_i \text{Loss}_i \]
The pedometer dataset collected from WeChat, a famous mobile application in China. The dataset contains 39,246 users' 300-days walking step counts during the period from 2018.6.11 to 2019.4.6.
Consider all classes to be equal important:

\[ wP = \sum_{y \in Y} \left( \frac{\sum_{i=1}^{u} I(y_i^* = \hat{y}_i \& y = \hat{y}_i)}{\sum_{i=1}^{u} I(y = \hat{y}_i)} \right) \times \text{weight} \]

\[ wR = \sum_{y \in Y} \left( \frac{\sum_{i=1}^{u} I(y_i^* = \hat{y}_i \& y = \hat{y}_i)}{\sum_{i=1}^{u} I(y = y_i^*)} \right) \times \text{weight} \]

\[ wF1 = 2 \times \frac{wP \times wR}{wP + wR} \]
Baseline Models

**POP** POP is a naive method that always predicts the given sample as the majority classes.

**JNE** Joint Neural Embedding maps users' all walking histories into latent vectors.

**SNE** Structured Neural Embedding has similar structure with JNE. The only difference between SNE and JNE is that the loss of SNE is computed via a log-bilinear model with structured predictions.

**ETN** Embedding Transformation Network uses a shared embedding just as SNE. The shared embedding is fed into an embedding transformation layer to obtain the transformed representation for prediction.

**ETNA** Embedding Transformation Network with Attention is an improved version of ETN. The transformed representation produced by embedding transformed layer is fed to a task-specific attention layer to take into account the importance of each element in users' profile.
**Performance comparison of different models**

| model Name | Results       | Results       | wF1  |
|------------|---------------|---------------|------|
|            | wP            | wR            |      |
| POP        | 0.137         | 0.354         | 0.198|
| JNE        | 0.609         | 0.643         | 0.626|
| SNE        | 0.615         | 0.648         | 0.631|
| ETN        | 0.634         | 0.679         | 0.656|
| ETNA       | 0.641         | 0.688         | 0.660|
| CANEA\(^a\) | 0.496   | 0.513         | 0.504|
| CANEA\(^b\) | 0.681   | 0.726         | 0.703|
| CANEA      | **0.695**     | **0.741**     | **0.717**|

\(^a\) abandon correlation learning layer
\(^b\) abandon attention mechanism
Visualization of Attention

Comparison of attention weights calculated by separated embedding layer

Mean value (y-axis) of attention scores from Sunday to Saturday (x-axis)

Gender

Age

UCAS
# Experiments on Transaction Data

The transaction dataset is the first public dataset containing both transaction records and demographic information in previous paper.

| model Name | Results |
|------------|---------|
|            | wP      | wR      | wF1    |
| POP        | 0.086   | 0.294   | 0.134  |
| JNE        | 0.321   | 0.348   | 0.334  |
| SNE        | 0.295   | 0.351   | 0.321  |
| ETN        | 0.310   | 0.368   | 0.336  |
| ETNA       | 0.339   | 0.382   | 0.360  |
| CANEA      | **0.356** | **0.401** | **0.377** |
Conclusion

- Extend our sight to the ubiquitous mobile and sensor devices to bridge the gap between pedometer data and users’ demographic attributes

- A new model named Correlation Aware Neural Embedding with Attention (CANEA) is proposed for multiple demographic prediction, which has strong ability to learn the complementary correlations between tasks. In addition, we explicitly analyze different days’ effect of pedometer records on demographic prediction

- Extensive experiments are conducted on a real world pedometer dataset as well as a public transaction dataset, which all demonstrate the effectiveness of the proposed method
THANKS