User Profiling through Deep Multimodal Fusion

Golnoosh Farnadi, Jie Tang, Martine De Cock, Marie-Francine Moens

By: Shadi Sadeghpour and Ali Jalalifar

Presentation Summary

- Introduction
- Related Work
- System Model
  - UDMF (User Profiling through Deep Multimodal Fusion)
- Evaluation
- Experimental Results
- Conclusion & Future Work

Introduction

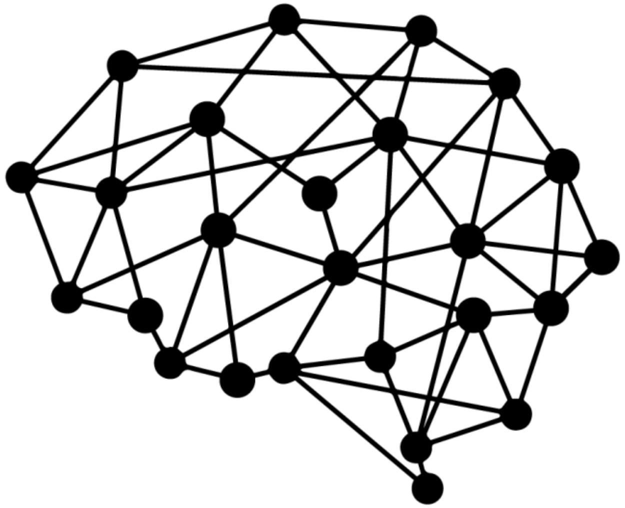
- **Personalization:** The adaption of the services to fit the user's interests, characteristics and needs.
- **User Profiling:** The key of effective personalization.
- **Multimodal Fusion:** The process of integrating information from various input modalities and combining them into a complete command.

- **User Contents:**
  - Textual Content
  - Visual Content
  - Relational Content

Introduction *(Cont'd)*

Deep neural network(DNNs)

- Neural networks lend themselves well for integrating multiple data sources, as they allow a non-linear combination of data sources to be trained to solve the problem.
- DNNs are arguably the best known method for most pattern recognition problems involving perception.



Paper Structure

- A hybrid DNN based framework to integrate multiple source of user data is introduced, UDMF.
- A mechanism of stacking to take advantage of the dependency among target variables to more accurately infer user attributes is proposed.
- Incorporation of shared and non-shared of representations among the data sources and integrates is presented, Power-set combination.
- A Node2Vec embedding for extracting features from social relational content is applied to infer users' age, gender and personality traits of social media users

Related Work

User Profiling

- Computational personality recognition in social media. *[G. Farnadi et al.]*
  Machine learning models to infer the age, gender, and personality traits of users.

- PAN *[http://pan.webis.de/]*
  Author profiling focus on various features and techniques to predict age and gender of authors, or shared tasks such as WCPR

- The YouTube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs *[J.-I. Biel and D. Gatica-Pere]*
  Vloggers (YouTube bloggers) to predict their personality based on their visual and audio content.

- Analyzing personality through social media profile picture choice *[Liu et al.]*
  Oxford project feature

Related Work *(Cont'd)*

Multimodal Modeling in Deep Neural Networks

- Multimodal deep learning *[J. Ngiam et al.]*
  A deep Autoencoder network to learn a multimodal feature representation in audio-visual speech recognition

- Modeling spatial-temporal clues in a hybrid deep learning framework for video classification *[Wu et al.]*
  CNN features with a Long Short Term Memory (LSTM) network

Stacking in Deep Neural Networks

- Deep neural networks employing multi-task learning and stacked bottleneck features for speech synthesis *[Z. Wu et al.]*
  Stack-Bottleneck features, task of speech recognition

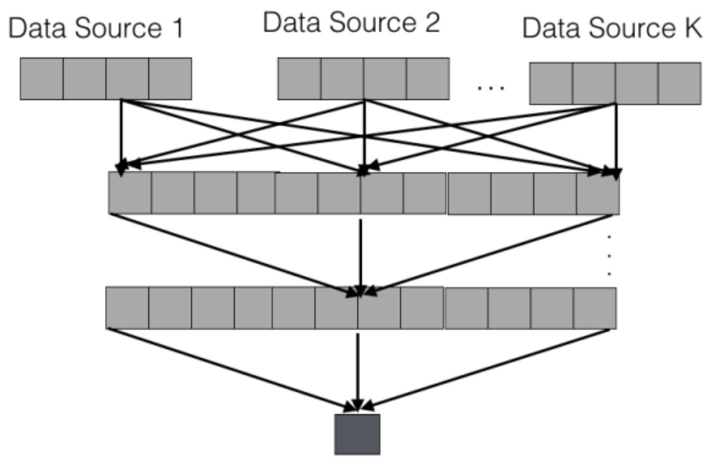
- Stacking-based deep neural network: Deep analytic network on convolutional spectral histogram features *[A. B.-J. Low et al.]*
  Stacking two deep neural networks: Unsupervised network & Supervised network

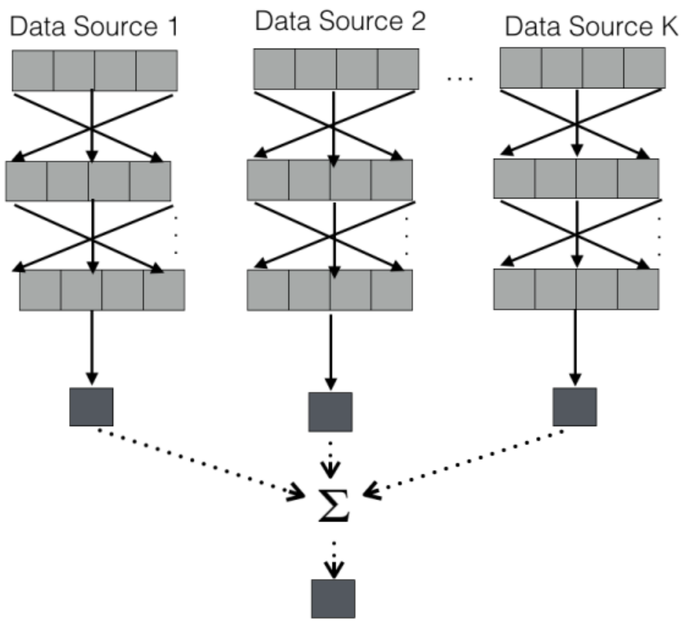
System Model

UDMF: User Profiling through Deep Multimodal Fusion

- Goal: Integrate two or more sources of data/knowledge and create a single representation that provides a more accurate description of the data sources than any of the individual ones.

- Early and Late approaches of integrating multiple data source in a deep neural network architecture





System Model *(Cont'd)*

A Hybrid Model

1. Leverages all sources of users' data and incorporates the correlation between modalities by mapping all combinations of data sources into shared representations.
2. Integration of data sources at the decision level, combine the decision of all combinations of data sources.
3. Utilize the decision of the data integration framework for dependent tasks in the learning process.

Deep Neural Networks

- Easy to combine various data sources
- Combine data sources with non-linear functions
- Extract features using unsupervised approaches (the Node2Vec embedding)

System Model *(Cont'd)*

Mechanism designed To integrate data sources in UDMF:

1. Stacking
2. Power-set Combination

General setting for a multilayer feedforward network described by an acyclic graph

1. Single data source D as input. The degree of activation of unit i on layer h

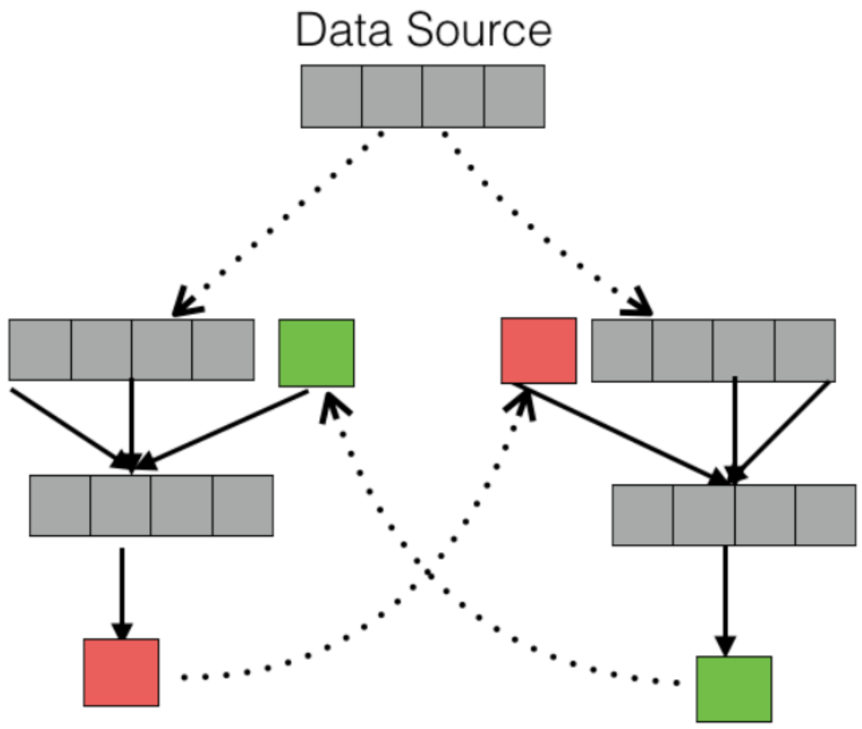
$$U_i^h(D) = f(\sum_j w_{ij}^{hl} \cdot U_j^l(D)) \qquad (1)$$

2. The degree of activation of unit i on layer 0

$$U_i^0(D) = f(\sum_j w_{ij} \cdot D_j) \qquad (2)$$

System Model - Stacking

Stacking is suitable for multi-task learning where target variables are correlated with each other



Stacking of 2 target variables given one data source

System Model - Stacking *(Cont'd)*

- The degree of activation of unit i on layer 0 at epoch q:

$$U_i{}^{0q}(D) = f(\sum_j w_{ij}.D_j + \sum_z w_{iz}.\alpha_z.t_z{}^{q-1}) \qquad (3)$$

  - $t_z{}^{q-1}$: target variable "z" at epoch q-1
  - α_z: Gating Variable

- the value of the target neurons Where q=0:

$$U_i^{00}(D) = f(\sum_j w_{ij} \cdot D_j) \qquad (4)$$

System Model - Power-set Combination

- Incorporate correlations among features and data sources by an early integration approach of all subsets of DS.
- Combine the predicted outcome as a late integration approach with an ensemble method.

$$DS = \{D_1, D_2, \ldots, D_k\}$$

- Counterpart of Equation (2) in the UDMF framework

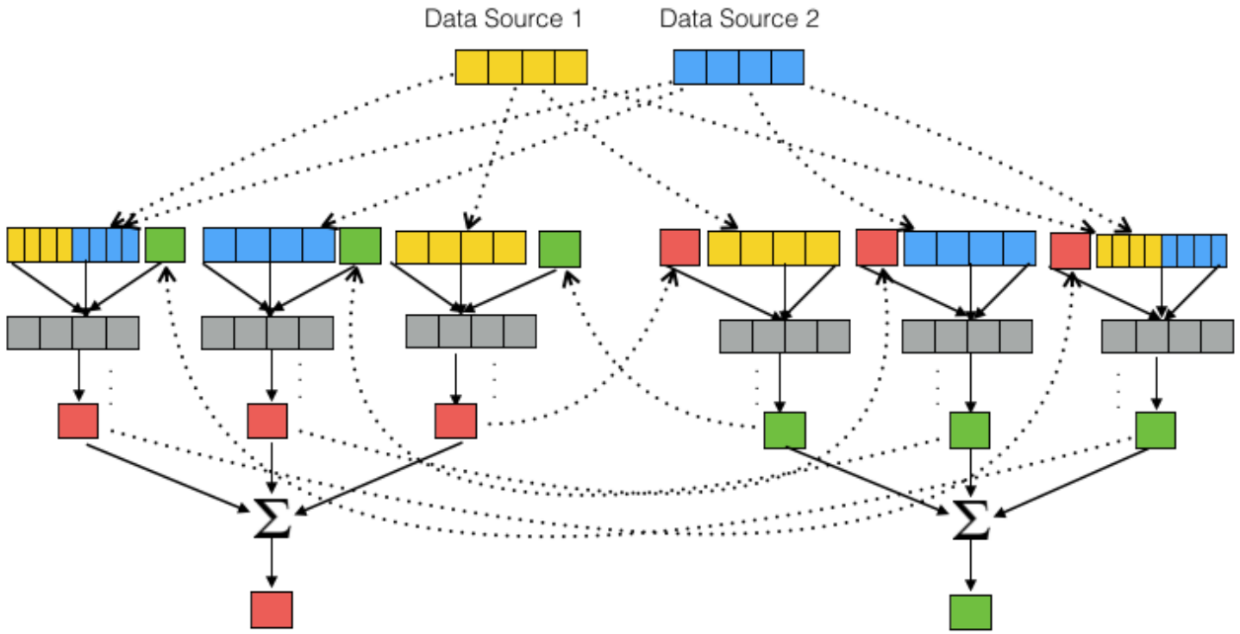
$$U_i^{0q}(\mathcal{D}) = f\Big(\sum_{D \in \mathcal{D}} \sum_{j} w_{ij} . D_j + \sum_{z} w_{iz} . \alpha_z . t_z^{q-1}\Big) \qquad (5)$$

System Model - Power-set Combination *(Cont'd)*

Example:

$DS = \{A, B\}$

Power-set $\mathcal{P}(DS) = \{\{A\}, \{B\}, \{A, B\}, \{\quad\}\}$



The architecture of UDMF with stacking of 2 target variables and power-set combination of two data sources.

- The three trained mini-DNNs for each target variable.
- Total: six trained mini-DNNs.
- The output of each mini-DNN is stacked as input to the sister mini-DNNs at the end of each epoch for the training at the next epoch.

Evaluation

MyPersonality project dataset [http://mypersonality.org/]

- A popular Facebook application introduced in 2007 in which users took a standard Big Five Factor Model psychometric questionnaire, [https://openpsychometrics.org/printable/big-five-personality-test.pdf]

A = 14 - (2) ____ + (7) ____ - (12) ____ + (17) ____ - (22) ____ + (27) ____ - (32) ____ + (37) ____ + (42) ____ + (47) ____ = ______

C = 14 + (3) ____ - (8) ____ + (13) ____ - (18) ____ + (23) ____ - (28) ____ + (33) ____ - (38) ____ + (43) ____ + (48) ____ = ______

N = 38 - (4) ____ + (9) ____ - (14) ____ + (19) ____ - (24) ____ - (29) ____ - (34) ____ - (39) ____ - (44) ____ - (49) ____ = ______

O = 8 + (5) ____ - (10) ____ + (15) ____ - (20) ____ + (25) ____ - (30) ____ + (35) ____ + (40) ____ + (45) ____ + (50) ____ = ______

The scores you calculate should be between zero and forty. Below is a description of each trait.

- **Extroversion (E)** is the personality trait of seeking fulfillment from sources outside the self or in community. High scorers tend to be very social while low scorers prefer to work on their projects alone.
- **Agreeableness (A)** reflects much individuals adjust their behavior to suit others. High scorers are typically polite and like people. Low scorers tend to 'tell it like it is'.
- **Conscientiousness (C)** is the personality trait of being honest and hardworking. High scorers tend to follow rules and prefer clean homes. Low scorers may be messy and cheat others.
- **Neuroticism (N)** is the personality trait of being emotional.
- **Openness to Experience (O)** is the personality trait of seeking new experience and intellectual pursuits. High scores may day dream a lot. Low scorers may be very down to earth.

Evaluation *(Cont'd)*

- Use the median value to create binary classes for each characteristic, the median value:
  age = 23, Opn = 4, Con = 3.5, Ext= 3.5, Agr = 3.65, and Neu = 2.75

- The tasks : predicting age, gender and personality traits of Facebook users using their textual (status updates),visual (profile picture) and relational data (page likes).

Configuration Parameters	
DNNs	3 Layers - Input layer Hidden layer(100 neurons per each data source) Sigmoid layer(Results)
ReLU	activation function - Model a non-linear combination of inputs
Adam	optimization algorithm
epochs	100
batch size	128

Evaluation *(Cont'd)*

How represent each data source for the task of user profiling in social media?

Data source embedding:

- Textual data source embedding
- Visual data source embedding
- Relational data source embedding

Textual data :

- Combine the status updates of each user in the dataset into one document per user.
- Each user with 88 Linguistic Inquiry and Word Count (LIWC) features extracted from her/his status updates, consisting of features related to

Evaluation *(Cont'd)*

Linguistic Inquiry and Word Count (LIWC)

- ❖ The **LIWC** program includes the main text analysis module along with a group of built-in dictionaries.

  a) Standard counts (word count)
  b) Psychological processes (the number of anger words such as hate, annoyed, … in the text)
  c) Relativity (the number of verbs in the future tense)
  d) Personal concern (the number of words that refer to occupation such as job, majors, ...)
  e) Linguistic dimensions (the number of swear words)

- ❖ LIWC outperformed **GloVe** and ***n*-gram** models**.**

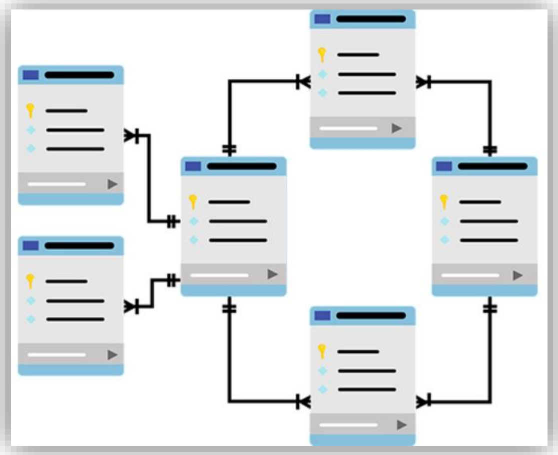
Evaluation *(Cont'd)*

Visual data

- For each user, his/her profile picture is used and extract 64 facial features using the Oxford Face API.
- The performance of the Oxford features is compared and assigned as the input layer, with a 128-dimensional activation vector extracted from the last layer (before the softmax) of the pre-trained VGG-16 and VGG-19 models in ImageNet.
- The DNNs with the Oxford features as the input layer significantly outperform the VGG-based models specifically for the task of age and gender prediction.

Evaluation *(Cont'd)*

Relational data



- An unsupervised deep neural network is trained, *Node2Vec* on the relational graph.
- Node2Vec is extending the Skip-gram architecture to networks.

relational graph $G = (V, E)$

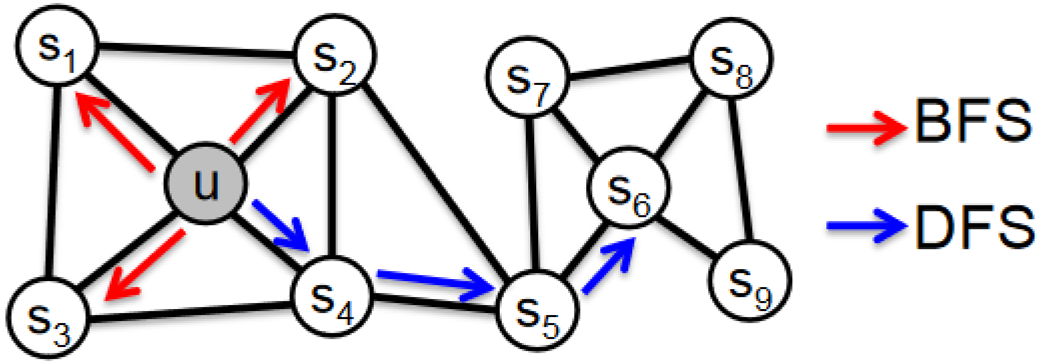
mapping function to represent nodes with features f $f : v \rightarrow \mathbb{R}^d$

- The Node2Vec model maximizes the log-probability of observing a network neighborhood $NS(u)$ for a node u conditioned on its feature representation, given by f

$$\max_{f} \sum_{u \in V} \log Pr(NS(u)|f)$$

Evaluation *(Cont'd)*

- A mapping of users to a low-dimensional space of features is learnt that maximizes the likelihood of preserving network neighborhoods of users and pages.
- A Node2Vec model is trained using the page like relations.
- Using features extracted from the Node2Vec model, not only users with pages that they like (i.e., their neighbors) can be represented, but also similar users by a flexible biased random walk procedure (Node2Vec walks) will be found to produce $N_S(u)$ that can explore neighborhoods in both a Breadth-First Sampling (BFS) as well as a Depth-First Sampling (DFS) fashion.
- It is iteratively performed the random Node2Vec walk on the graph to sample nearest neighbors for each node and then train a Skip-gram architecture to find embeddings for each node.



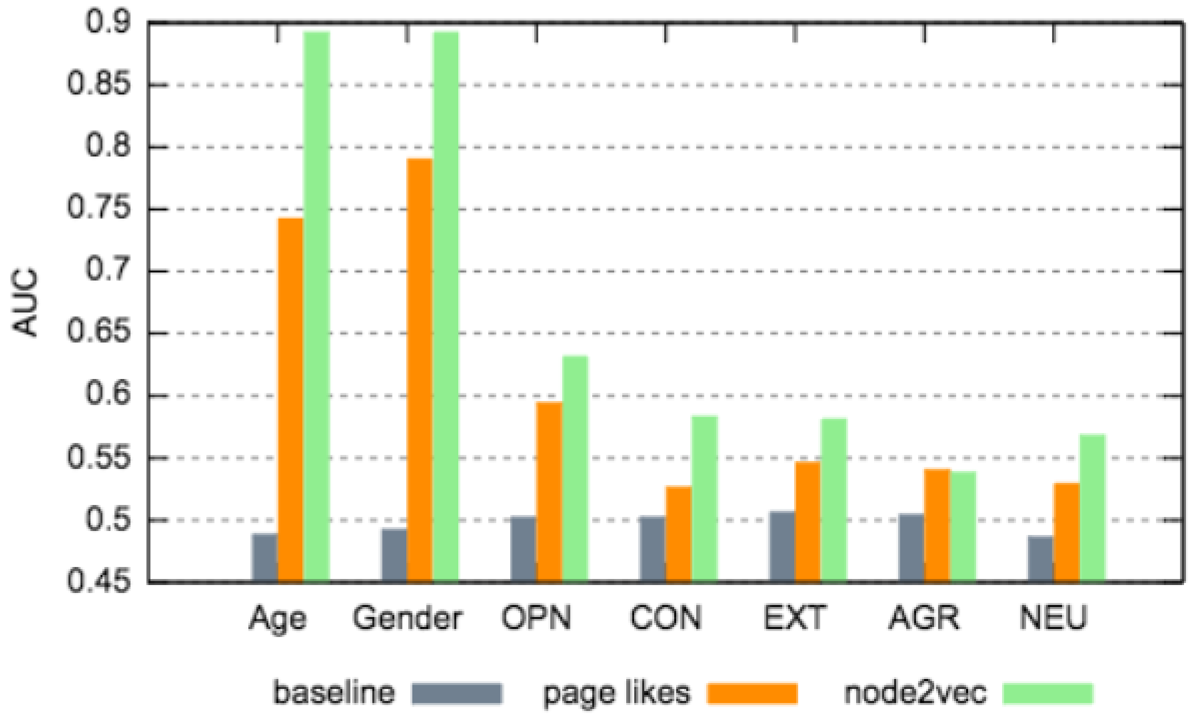
BFS and DFS search strategies from node u(k= 3)

Evaluation *(Cont'd)*

- Page likes model

- Node2Vec embeddings for the user profiling task.

$$\begin{array}{cc} & \text{Pages} \\ \text{Users} & \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ & \cdots & \end{bmatrix} \end{array}$$



- Node2Vec features extracted from users' page likes outperform using only pages that users like as features for the tasks of inferring gender, age, and Big Five personality traits.

Evaluation *(Cont'd)*

Data Source Integration:

1. Design a Mini-DNNs architecture, with one layer for the input neurons, one hidden layer which is fully connected to the input neurons and one output layer which has the output of the learning task using the sigmoid activation function.

2. Similar mini-DNNs per each target variable that are integrated in the stacking process.

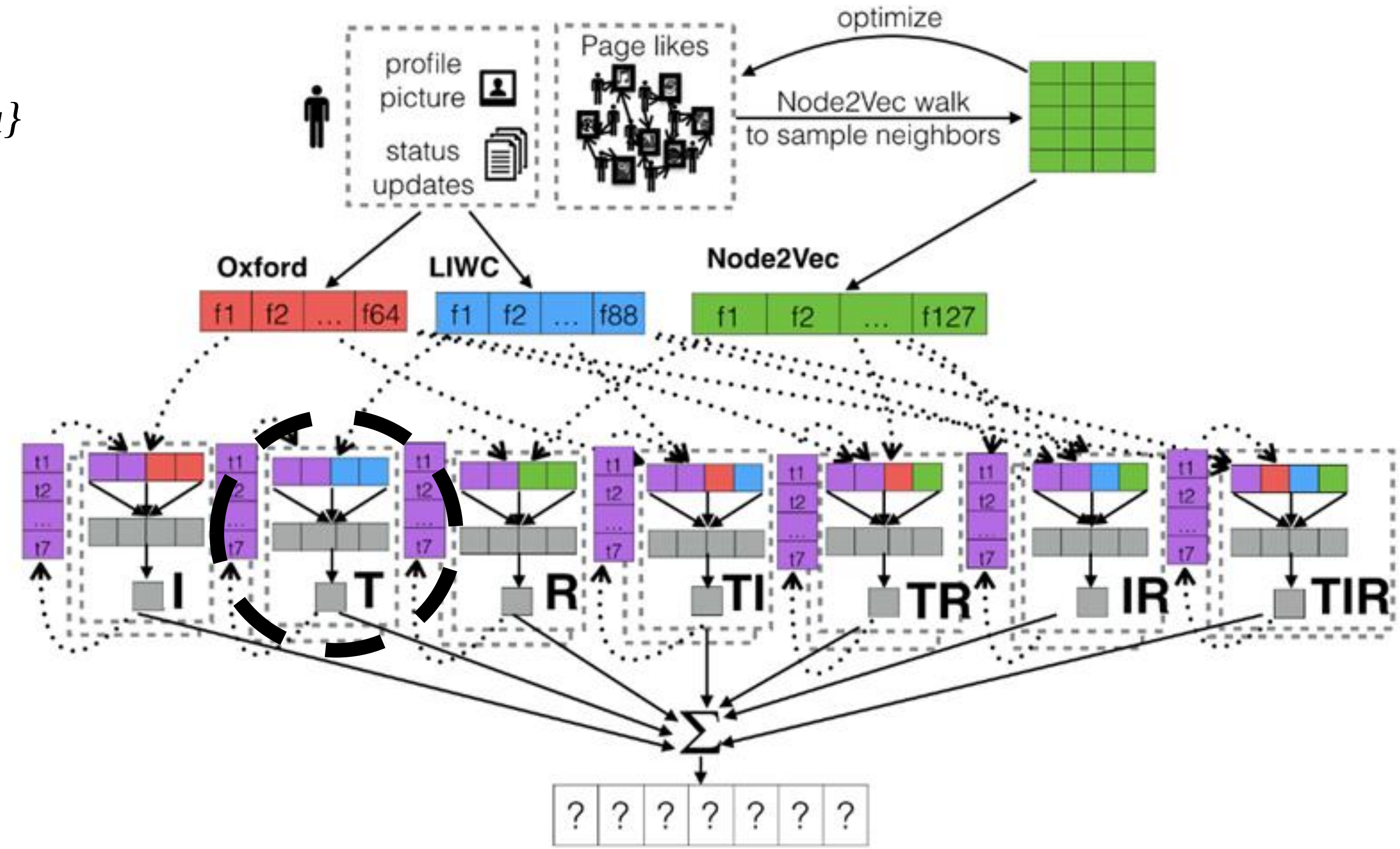
3. For each subset of data sources, p mini-DNNs getting updated at each epoch from the output of $p-1$ other mini-DNNs.

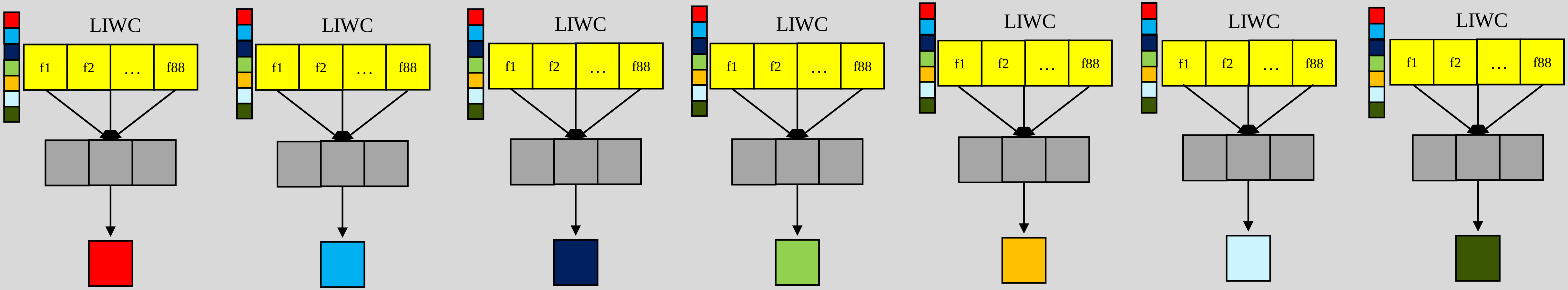
4. Apply majority voting to determine the label.

Evaluation *(Cont’d)*

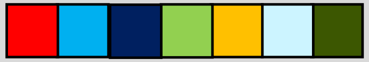
L = { *female, young, Opn, Con, Ext, Agr, Neu*}
P *target variable*

- Each network is trained for one target variable from a set of labels l.





Target Variables:



Experimental Results

- UDMF evaluation for the task of inferring users 'age, gender and personality traits in Facebook.
- Examine the output of the UDMF framework using Stacking only stacking with single source, combination of two sources, all three sources. ***[Unimodal baselines & Multimodal baselines]***

Model	Stack	Age	Gender	Opn	Con	Ext	Agr	Neu
Baseline		0.488	0.492	0.502	0.502	0.506	0.506	0.486
One source								
Text	✗	0.741±0.022	0.668±0.020	0.550±0.016	0.575±0.017	0.536±0.016	0.547±0.016	0.523±0.016
	✓	0.748±0.022	0.668±0.020	0.553±0.017	0.574±0.017	0.545±0.016	0.550±0.016	0.524±0.016
Image	✗	0.552±0.016	0.915±0.027	0.502±0.015	0.500±0.015	0.504±0.015	0.512±0.015	0.520±0.016
	✓	0.550±0.016	0.897±0.027	0.516±0.015	0.511±0.015	0.518±0.015	0.519±0.015	0.541±0.016
Relation	✗	0.875±0.026	0.886±0.027	0.601±0.018	0.571±0.017	0.567±0.017	0.525±0.016	0.558±0.017
	✓	0.893±0.027	0.898±0.027	0.622±0.018	0.589±0.018	0.573±0.017	0.533±0.016	0.563±0.016
Two sources								
Early approach	✗	0.734±0.022	0.873±0.026	0.569±0.017	0.588±0.018	0.536±0.016	0.545±0.016	0.547±0.016
TI	✓	0.746±0.022	0.864±0.026	0.546±0.016	0.568±0.017	0.542±0.016	0.546±0.016	0.536±0.016
Early approach	✗	0.878±0.026	0.896±0.027	0.610±0.018	0.586±0.018	0.567±0.017	0.535±0.016	0.554±0.017
TR	✓	0.891±0.027	0.899±0.027	0.627±0.019	0.601±0.019	0.572±0.017	0.551±0.016	**0.574±0.017**
Early approach	✗	0.878±0.026	0.951±0.028	0.606±0.018	0.574±0.017	0.569±0.017	0.524±0.016	0.562±0.017
IR	✓	0.895±0.027	0.951±0.028	0.633±0.019	0.592±0.018	0.577±0.017	0.537±0.016	0.564±0.017
Three sources								
Ensemble	✗	0.876±0.026	**0.952±0.028**	0.603±0.018	0.587±0.018	0.569±0.017	0.537±0.016	0.562±0.017
(Late approach)	✓	0.893±0.027	0.949±0.028	0.626±0.019	0.606±0.018	**0.582±0.017**	0.549±0.016	0.570±0.017
Early approach	✗	0.887±0.027	0.947±0.028	0.617±0.018	0.577±0.017	0.567±0.017	0.541±0.016	0.566±0.017
TIR	✓	**0.899±0.027**	0.934±0.028	**0.635±0.019**	**0.607±0.018**	0.560±0.018	**0.551±0.016**	0.572±0.017

Experimental Results *(Cont'd)*

- Examine the capabilities of the hybrid UDMF framework with both stacking and power-set combination of two and three data sources in modeling multiple data sources.

Model	Age	Gender	Opn	Con	Ext	Agr	Neu
			One/Two sources				
Page likes	0.743±0.020	0.699±0.022	0.605±0.017	0.516±0.016	0.555±0.016	0.540±0.0161	0.527±0.016
LR (T)	0.711±0.021	0.654±0.020	0.564±0.017	0.568±0.017	0.551±0.016	0.548±0.016	0.530±0.016
LR (I)	0.584±0.017	0.858 ±0.026	0.514±0.015	0.520±0.015	0.528±0.016	0.528±0.016	0.525±0.016
LR(T,I)	0.711±0.017	0.852 ±0.025	0.555±0.017	0.564±0.017	0.551±0.016	0.550±0.016	0.542±0.016
UDMF(T,I)	0.756±0.023	0.886±0.027	0.569±0.017	0.575±0.017	0.552±0.017	0.552±0.016	0.539±0.016
UDMF(T,R)	0.879±0.026	0.943±0.028	0.628±0.019	0.607±0.018	0.580±0.017	**0.564±0.017**	0.575±0.017
UDMF(I,R)	0.892±0.027	0.955±0.029	0.630±0.019	0.607±0.018	0.587±0.018	0.551±0.016	0.571±0.017
			Three sources				
Weighted Soft Voting	0.656±0.019	0.861±0.026	0.523±0.016	0.0523±0.016	0.508±0.015	0.507±0.015	0.518±0.015
Random Forest (100)(T,I,R)	0.786 ±0.023	0.900±0.027	0.588±0.018	0.564 ±0.017	0.544±0.016	0.549±0.016	0.538±0.016
LR(T,I,R)	0.808 ±0.024	0.888±0.027	0.603±0.018	0.585 ±0.018	0.550±0.017	0.550±0.016	0.572±0.017
UDMF(T,I,R)	**0.903±0.027**	**0.956±0.029**	**0.647±0.019**	**0.615±0.018**	**0.592±0.018**	0.556±0.017	**0.580±0.017**

Mean and standard deviation of AUC scores in inferring age, gender and personality traits by fusing two and three data sources in UDMF.

Conclusion

- A hybrid user profiling architecture in deep neural networks is proposed, UDMF.
- UDMF has two simple and yet effective properties, stacking and power-set combination strategies.
- UDMF combines different modalities both at the feature level and decision level to predict accurate multiple attributes of social media users given their user generated content and social relational content.
- Stacking and power-set combination in UDMF enhance the learning power in combining the data sources.
- To make a user profile, UDMF predicted age, gender and personality traits of users with highly accurate results.

Limitation and Future Direction

- Small dataset of textual and visual, no NN for learning features.
- Add other data sources, temporal and geological, …
- Very simple DNN model with only three layers to collect these results, more layers and using regularizer to improve the results
- …

YORK U
UNIVERSITÉ
UNIVERSITY

THANK YOU