Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)

TweetFit: Fusing Multiple Social Media and Sensor Data for Wellness Profile Learning

Aleksandr Farseev
farseev@u.nus.edu
School of Computing
National University of Singapore
13 Computing Dr, Singapore 117417

Tat-Seng Chua
chuats@comp.nus.edu.sg
School of Computing
National University of Singapore
13 Computing Dr, Singapore 117417

Abstract

Wellness is a widely popular concept that is commonly applied to fitness and self-help products or services. Inference of personal wellness-related attributes, such as body mass index or diseases tendency, as well as understanding of global dependencies between wellness attributes and users' behavior is of crucial importance to various applications in personal and public wellness domains. Meanwhile, the emergence of social media platforms and wearable sensors makes it feasible to perform wellness profiling for users from multiple perspectives. However, research efforts on wellness profiling and integration of social media and sensor data are relatively sparse, and this study represents one of the first attempts in this direction. Specifically, to infer personal wellness attributes, we proposed multi-source individual user profile learning framework named "TweetFit". "TweetFit" can handle data incompleteness and perform wellness attributes inference from sensor and social media data simultaneously. Our experimental results show that the integration of the data from sensors and multiple social media sources can substantially boost the wellness profiling performance.

Introduction

During the past decade, social multimedia services have drastically increased its impact on people's daily life. For example, more than half of American smartphone users were reported to spend an average of 144 minutes per day browsing their mobile devices, aiming to stay socially connected with their friends. Meanwhile, these users often follow the so-called Quantified Self tendency, which includes measuring and publishing various signals from wearable sensors (such as heart rate, body acceleration or physical location). These data is of crucial importance for research in wellness domain since it describes users' actual physical condition, which is related to users' well-being. At the same time, recent works demonstrated the great potential of social media data for wellness-related research (Mejova et al. 2015; Akbari et al. 2016). However, most of these works are descriptive in nature and do not study the integration of data from social media and wearable sensors. Considering that most Internet-active adults actively use more than four social media services in their everyday life (GlobalWebIndex 2016) with wide availability of data from wearable sensors, it seems reasonable to combine multimodal content from different social networks with sensor data for joint processing (Jain and Jalali 2014). Such integration will narrow the gap between users' online representation and actual physical status, which is the right step towards realizing the ideal of 360° user profiling (Farseev et al. 2016).

This article focuses on the problem of individual wellness user profiling based on data from multiple social networks and wearable sensors. Here, an individual wellness profile involves personal user attributes (Farseev et al. 2016) such as demographics (age, gender, occupation, etc.) (Farseev et al. 2015), Body Mass Index (BMI) category[1], personality (Buraya et al. 2017), or chronic disease tendency (Akbari et al. 2016). In our study, we focus on two important personal wellness attributes - BMI category and "BMI Trend' (the direction of BMI fluctuation over time - Increase/Decrease). Both attributes are closely related and correlated to one's overall health. For example, Field et al. (2001) discovered that people whose BMI is higher than 35.0 are approximately 20 times more likely to develop diabetes. Other benefits of such attributes include: a) BMI category can be further used in public health domain to monitor wellness tendencies of social media users at the global level; b) "BMI Trend" information can be utilized by users to rectify their lifestyle (i.e. via interactive mobile application or a "Smart Watch"), and by doctors to gain a complete picture of patient's health.

There are three challenges in addressing individual wellness profiling: 1) **Data gathering.** The data from modern social media services and sensor devices is often stored in independent web resources and hidden behind the privacy settings. Furthermore, the data from wearable sensors as well as personal attributes such as BMI or demography are often not publicly accessible. It is thus necessary to implement data collection and cross-source user account mapping techniques to support large-scale social media research. 2) **Data representation.** Besides the textual data, social media services involve data of various modalities. For example, in Instagram, users share recently taken pictures and

[1]The BMI measure is defined as the body mass divided by the square of the body height. Based on BMI, an individual can be categorized into one out of 8 BMI categories, namely "Severe Thinness", "Moderate Thinness", "Mild Thinness", "Normal", "Pre Obese", "Obese", "Obese II", and "Obese III" (WHO 2011).

videos, while in Endomondo[2] users post information about their workouts, which is strongly dependent on the temporal and spatial aspects. Integration of such heterogeneous multimodal data sources requires development of efficient and mutually consistent data representation approaches. 3) **Data modeling:** Efficient data integration for individual wellness profile learning is a tough challenge since the data from independent media sources is different in nature. Furthermore, multi-source data is often incomplete, which means that some users may not be active on all social networks. Finally, high dimensionality of multi-source feature space often leads to the so-called "curse of dimensionality" problem. Development of a learning framework that can handle all these issues is a hard task.

Inspired by previous studies and the challenges above, in this work we seek to address two research questions. First, to support the assumptions behind this study, it is important to understand: **(RQ1) Is it possible to improve the performance of BMI category and "BMI Trend" inference by fusing multiple social media and sensor data?** Second, for further wellness profiling improvement, it is essential to gain insight into: **(RQ2) What is the contribution of sensor data towards BMI category and "BMI Trend" inference?**

To answer the above research questions, we present a new computational wellness profiling framework named "TweetFit". We introduce the techniques to gather and represent data from a novel sensor data source (the Endomondo workouts) and other social media sources: Twitter, Foursquare, and Instagram, from which we predict users' BMI category and "BMI Trend". To do so, we treat individual wellness profiling as a regularized multi-task learning (MTL) problem, where different data source combinations for each inference category are represented as MTL "tasks". To facilitate further research, we release our multi-source multimodal sensor-social dataset (Farseev 2017) for public use.

The main contributions of this study are twofold: first, we present a **multi-source multi-task learning framework for wellness attribute inference**, which performs personal wellness profiling via regularized multi-task learning; second, we release a **large-scale social-sensor dataset** — a new benchmark towards wellness profiling with multi-source multimodal data and data from wearable sensors.

Related Work

Recently, medical and healthcare communities suggested the use of social media and sensor data as a meaningful resource for different wellness applications. For example, Eggleston et al. (2014) used social media to monitor food-related habits for obese and diabetes patients, while Fried et al. (2014) predicted diabetes and overweight rates for 15 US cities. At the same time, Mejova et al. (2015) testified the predictive power of Foursquare-based features for group obesity inference task and cultural differences analysis, while Abbar et al. (2015) leveraged Twitter data in an attempt to predict obesity and diabetes statistics based on food names in tweets. Finally, Akbari et al. (2016) proposed a multi-task learning framework for personal wellness events categorization. These research efforts were made towards wellness lifestyle analysis and the results show the great potential of social media data to assist in wellness related research. However, most of the works mentioned above are either descriptive in nature, use only a single data source, or built on naive data analytics approaches. They may not be useful to gain deeper insights from multi-source social media data and wearable sensors.

Meanwhile, there were several research efforts done on multi-source user profile learning. In an earlier work, Liu et al. (2009) embedded the so-called $\ell_{2,1}$ regularization in the multi-task learning to obtain sparse data representations for feature selection purpose, which is useful in high-dimensional data processing. However, the data source integration was carried out in an "early-fusion" manner, where all the features were fused into one vector before model training. Such a data integration strategy may result in high dimensionality and suboptimal final results. Farseev et al. (2015) introduced efficient ensemble learning solution, aiming to combine multi-source multimodal data for demographic user profile learning. The model was trained independently on each data source and consolidated in a "late-fusion" manner, which does not fully take advantage of multi-source data. Finally, Song et al. (2015) employed the structure-constrained multi-task learning framework for user interests inference from multi-source data. However, the framework relies on an external knowledge and data completion techniques, which makes it biased towards particular datasets and tasks. Due to the reasons above, the development of a fully-automated multi-source individual wellness profiling approach that would not rely on external knowledge and data completion techniques is of crucial importance for wellness profile learning.

NUS-SENSE: Sensor-Social Dataset

To build a comprehensive user profile, it is essential to integrate multimodal data from various sources that represent users from multiple perspectives (Song et al. 2015). At the same time, a complete wellness profile must incorporate information about users' physical health (Corbin et al. 2001). In the following, we describe the commonly-used data modalities and their potential for individual wellness profiling. First, it was noted that textual information is one of the most valuable contributors towards user profile learning, mainly because of its high availability and its ability to describe users' daily routines comprehensively (Farseev et al. 2016). Second, it was also observed (Farseev et al. 2015) that visual data plays an important role in age and gender prediction. It is reasonable to hypothesize that this data is also useful for individual user profile learning in wellness domain. Third, it was reported (Mejova et al. 2015) that data from location-based social networks is helpful in obesity estimation at the group level, which demonstrates its potential for use in personal BMI category and "BMI Trend" prediction. Finally, data from wearable sensors was found to be valuable for user activity recognition and health monitoring (Banaee, Ahmed, and Loutfi 2013); this shows its potential towards individual wellness profile learning. Considering the above, we harvested data from multiple social media sources. In

[2]endomondo.com

Table 1: Number of data records in NUS-SENSE dataset

Source	Twitter	End-do	Four-re	Inst-m
# Posts	1676310	140926	19743	48137
# Users	5375	4205	609	2062
#Age	3974	3111	427	1525
# Gender	5375	4025	609	2062
# BMI Cat.	1052	870	116	372
# BMI Tr.	147	136	18	51

particular, we utilized **Twitter** tweets as a textual data source; **Instagram** pictures and its descriptions (comments) as image and textual data sources; **Foursquare** check-ins and its corresponding comments (shouts) as venue semantics, mobility, and textual data sources; and **Endomondo** workouts as a sensor data source and for ground truth construction.

It is noted that other than providing the so-called, exercise semantics, Endomondo also serves as a rich source of sequential data from wearable sensors and reliable wellness-related ground truth. The exercise data sequences are often publicly available and usually include a series of multi-dimensional data points, each of which may contain such attributes as Altitude, Longitude, Latitude, Time, Heart Rate, etc. (see Figure 1 (b)). The ground truth labels can therefore be derived from these publicly accessible Endomondo user profiles' web pages, which often include such personal attributes as Country of Residence, Postal Code, Age (Birthday), Gender (Sex), Height, and Weight (see Figure 1 (a)). These attributes are either manually input by Endomondo App. users or automatically measured by connected "smart" sensors (i.e. FitBit Aria Smart Scale[3]). In fact, Endomondo data goes beyond representing users from just one more modality, but bridges the gap between online social media-based users' representation and their actual offline physical activities and condition.

The data was harvested in the period of 1 May 2015 to 28 Aug 2015. It was conducted in the following three steps: **1) Search of seed users.** We collected a "seed" set of Twitter users, who were recently active in Endomondo, by performing a search via Twitter Search API. **2) User-generated content collection.** We then started a Twitter "stream" that involve all "seed" users and download the multi-source user-generated content of these users via URLs to the original Twitter posts (see Figure 1 (c)). **3) Ground truth collection.** During the Twitter crawling process, we monitored the Endomondo users' accounts daily and recorded all the BMI updates during the whole data collection period. The average between a user's Weight and Height updates was used to compute his/her BMI. The difference between a user's first and last Weight and Height updates was used to estimate his/her "BMI Trend". Table 1 lists the dataset statistics.

To preserve users' privacy, the dataset is released in the form of data representations (features) and anonymized multi-source user timelines, instead of the original user posts (Farseev 2017). In the dataset, users are well distributed in all BMI categories. From Figure 2, it is apparent that the highest percentage of users (38%) belong to "Normal" BMI category, and the lowerest percentage of users (3%) belong to "Moderate Thinness" BMI group. It suggests that there is sufficient data samples to train a supervised model for BMI category classification task, but the evaluation must be conducted on each BMI category separately to avoid the imbalanced dataset evaluation problem (Farseev et al. 2015). It is also noticed that the distribution of users among "BMI Trend" is slightly shifted to the "Decrease" (56%) category, which can be explained by Endomondo users' general intention to "lose weight".

Data Representation

We extracted the following features:

1) Text Features: In our study, we aggregated textual data from the following data sources: *Twitter tweets*, *Instagram image captions*, *Instagram image comments*, and *Foursquare check-in comments (shouts)*. More specifically, we extracted the following set of features: **a) Latent Topic Features.** We merged all the textual data of each user into a document. All documents from multiple users were projected into a latent topic space using Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2003), with empirically determined parameters of $T = 50, \alpha = 0.5, \beta = 0.1$. **b) Writing style features.** As in Farseev et al. (2015), we extracted such writing style features as: *number of mistakes per post*, *number of slang words per post*, *average post sentiment*, which, based on our preliminary experiments, were found to be significantly ($\alpha = 0.05$) correlated to users' BMI. **c) Lexicon-based features.** We used two crowd-sourced lexicons of terms associated with controversial subjects from the US press (Mejova et al. 2014) and the lexicon of terms' healthiness category (Mejova et al. 2015). Additionally, we extracted food type and average calorie content from each post by using the Twitter Food Lexicon (Abbar, Mejova, and Weber 2015).

2) Venue Semantics Features: Similar to Farseev et al. (2015), we represented location data as a distribution of users' check-ins among 764 Foursquare venue categories. To overcome the data sparsity problem, we further reduced the data dimensionality by extracting Top 86 principal components (Jolliffe 2002), which preserve 85% of variance.

3) Mobility and Temporal Features: The following mobility features were extracted from Foursquare based on users' *areas of interest (AOIs)* (Qu and Zhang 2013), which is, essentially, the geographical regions of high user's check-ins density (regardless of check-in venue semantics): **a) average number of posts during each of the 8 daytime durations**, where each time duration is 3 hours long (i.e. $15-18$); **b) number of areas of interest (AOI)**; **c) median size of AOIs**; **d) number of AOI outliers**; and **e) median distance between AOIs**.

4) Visual Features: Inspired by Farseev et al. (2015), we computed distribution of user's photos among the 1000 ImageNet visual concepts (Deng et al. 2009) for each Instagram user. Similar to venue semantics features, we extracted Top

[3] www.fitbit.com/aria

Figure 1: Endomondo user profile (a), Endomondo workout (b), and the repost of Endomondo workout in Twitter (c).

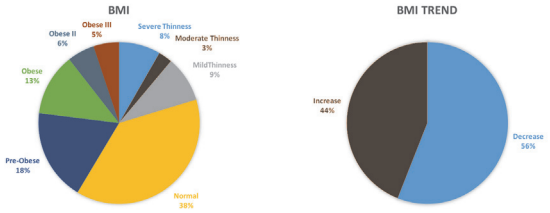

Figure 2: Distribution of users among different BMI categories and "BMI Trends" in NUS-SENSE dataset

150 principal components (preserves 85% of variance) from image concept distribution.

5) Sensor Features: To represent sensor data consistently with other data modalities, we incorporated the following feature types: **a) exercise statistics:** we computed the following averaged features using all sensor data samples for each user: *Distance (Ascend/Descend)*, *Speed*, *Duration*, *Hydration*; **b) external sensors statistics:** apart from wearable sensor features, we leveraged data from external weather sensors (where it is available) such as *Wind Speed* and *Weather Type*; **c) workout type distribution:** we also represented sensor data as a distribution of users' workouts among the 96 Endomondo workout categories; **d) frequency domain features:** we extracted frequency features for each workout by applying Fast Fourier Transform (Bracewell 1965) followed by low band-pass-filter (0 – 0.5 Hz) to construct the energy distribution among the 99 *frequency bins* for each of the five sensor signal types, namely, *Altitude, Cadence, Speed, Heart Rate (HR), and Oxygen Consumption (Oxygen)*. We then merged these 5 vectors together to obtain a frequency domain feature vector of size 495 for each user. Similar to venue semantics features, we extracted Top 54 principal components that preserves 85% of variance.

Individual Wellness Profiling

Problem Statement

In this work, we treat the problem of multi-source individual wellness profiling as a multi-task learning (Caruana 1997) problem. One significant issue in multi-task learning is how to define and employ the commonality among different tasks. Intuitively, different data source combinations may share common knowledge for predicting wellness attributes. By following this philosophy, we define a multi-task learning task as a unique combination of different sources for a given category. For notational convenience, in the following sections, we describe the case of single-category multi-source multi-task learning. In the case of multi-category inference (i.e. BMI category prediction), the single-category models can be naturally combined in one-vs-all manner (Rifkin and Klautau 2004).

Notation: In the rest of this paper, we use uppercase boldface letters (i.e. $\mathbf{M}$) to denote matrices, lowercase boldface letters (i.e. $\mathbf{v}$) to denote vectors, lowercase letters (i.e. s) to denote scalars, and uppercase letters (i.e. N) to denote constants. For matrix $\mathbf{M} = (m^i_j)$, $\|\mathbf{M}\|$ is the ℓ_2 (Frobenius) norm, while the $\|\mathbf{M}\|_{2,1} = \sum_{i=1}^{n} \left\|\mathbf{m}^i\right\|$ is the $\ell_{2,1}$ norm (Liu, Ji, and Ye 2009) ($\mathbf{m}^i$ is the ith row of the matrix $\mathbf{M}$).

Modeling Multi-Source Fusion

First, we propose a sparse model that mitigates the problem of joint learning from sensor and social media data aiming to infer BMI category and "BMI Trend" attributes.

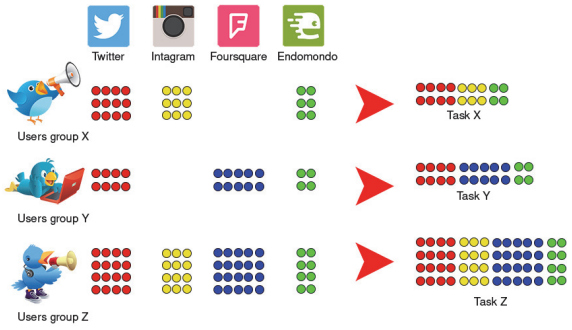

Figure 3: Incorporating block-wise incomplete data into multi-task learning model.

Suppose that there is a set of N exclusively labeled data samples and $S \geq 2$ data sources. We divide the dataset into T tasks, where *each task t is represented by the unique combination of available data sources* (see Figure 3). The number of features of task t is denoted as D_t; the maximum possible number of features of a task (i.e. when all data sources

are available) is denoted as D_{max}; the number of data samples of task t is denoted as N_t, and the number of different existing combinations of sources is denoted as T.

Figure 3 presents a toy example where four data sources (Twitter, Instagram, Foursquare, Endomondo) and three groups of social media users (X, Y, Z) are involved in multi-task learning process. The user group X consists of 3 users; the user group Y includes 2 users; the user group Z includes 4 users. These three user groups form three distinct multi-task learning task types, where the first task type (Task X) represents Twitter + Instagram + Endomondo data source combination; second task type (Task Y) represents Twitter + Foursquare + Endomondo data source combination; and third task type (Task Z) represents Twitter + Instagram + Foursquare + Endomondo (all data sources) data source combination. The aim is to train a model, which can predict the target category.

Formally, each of the T tasks can be defined as a set of pairs (jth data sample $\mathbf{x}_j^t$ and its corresponding label y_j^t):

$$t = \{(\mathbf{x}_j^t, y_j^t) \mid j = 1...N_t,\ \mathbf{x}_j^t \in\ R^{D_t},\ y_j^t \in\ \{-1; 1\}\}.$$

The prediction for jth data sample that corresponds to task t is then given by:

$$f_t(\mathbf{x}_j^t; \mathbf{w}^t) = \mathbf{x}_j^{t\intercal}\mathbf{w}^t$$

where $\mathbf{w}^t$ is the model parameter vector of task t. All model parameters are denoted as the block matrix $\mathbf{W}$:

$$\mathbf{W} = (\mathbf{w}^{1\intercal}, \mathbf{w}^{2\intercal}, ..., \mathbf{w}^{T\intercal}) \in \mathbb{R}^{D_{max}\ \times\ T}.$$

The optimal $\mathbf{W}$ can be found by solving:

$$\Gamma(\mathbf{W}) = \arg\min_{\mathbf{W}} \Psi(\mathbf{X}, \mathbf{W}, \mathbf{Y}) + \lambda\Upsilon(\mathbf{W}), \qquad (1)$$

where $\Psi(\mathbf{X}, \mathbf{W}, \mathbf{Y})$ is the loss function, $\Upsilon(\mathbf{W})$ is the sparsity regularizer that selects the discriminant features to prevent high data dimensionality (Liu, Ji, and Ye 2009), and $\lambda \geq 0$ controls the group sparsity.

The loss function $\Psi(\mathbf{X}, \mathbf{W}, \mathbf{Y})$ term can be replaced by a convex smooth loss function. In this work, we adopt the logistic loss:

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

To incorporate feature selection into the objective (Liu, Ji, and Ye 2009), we define $\Upsilon(\mathbf{W})$ as:

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

where F_s is the feature vector dimension of the data source s, and $\rho(s, f)$ is the index function that denotes all the model parameters of the fth feature from the data source s. The Υ term is the $\ell_{2,1}$ norm (Akbari et al. 2016), which leads to a sparse solution (controlled by $\lambda \geq 0$) via constraining all tasks that involve source s to share a common set of features.

Optimization

The objective function in Eq. 1 is convex but not smooth, since it consists of smooth (Ψ) and non-smooth (Υ) terms. This means that the conventional optimization approaches, such as Gradient Decent, are not directly applicable in our case. Inspired by the fast convergence rate of the Nesterov's approach (Liu, Ji, and Ye 2009), we reformulate the non-smooth problem from Equation (1) as:

$$f(\mathbf{W}) = \arg\min_{\mathbf{W}\in\mathbf{Z}} \Psi(\mathbf{X}, \mathbf{W}, \mathbf{Y})$$
$$s.t.\ \mathbf{Z} = \left\{\mathbf{W} \mid \|\mathbf{W}\|_{2,1} \leq z\right\},$$

where $z \geq 0$ is the radius of the $\ell_{2,1}$-ball, and there is a one-to-one correspondence between λ and z (proof is given in Liu et al. (2009)).

In Nesterov's method, the solution on each step ($\mathbf{W}_{i+1}$) is computed as a "gradient" of a search point $\mathbf{S}_i$:

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

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

where $\mathbf{S}_i$ is computed from the past solutions:

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

where α_i is the combination coefficient, and γ_i is the appropriate step size for $\mathbf{S}_i$ (can be determined by line search according to Armijo-Goldstein rule).

Evaluation

To answer our research questions, we compare the performance of "TweetFit" (trained based on all data sources) with "TweetFit" trained based on different data source combinations and various state-of-the-art baselines. For evaluation purposes, NUS-SENSE dataset was uniformly split into a train (80% of users) and testing (20% of users) sets.

Evaluation Metrics

We explicitly evaluate the performance of our proposed "TweetFit" framework ($\alpha = 0.1$) by solving the problem of individual wellness profiling. Specifically, we present the inference results of two personal wellness attributes: BMI category (eight attribute classes) and "BMI Trend" (binary classification). To perform BMI category inference, we first solved the problem in Equation 1 for each inference category, and then combined the obtained results in one-vs-all manner (Rifkin and Klautau 2004). To avoid the prevalence of popular BMI categories in evaluation, we use "Macro-Recall" (R_M), "Macro-Precision" (P_M), and "Macro-F_1" ($F_{1,M}$) metrics, which are the averaged "Precision", "Recall", and "F_1" measures across all categories (Farseev et al. 2015). To tackle the data imbalance problem at the training stage, we uniformly selected equal number of negative and positive samples for each binary classification task.

Evaluation Against Data Source Combinations

As mentioned above, in this work we utilized principal component analysis (PCA) (Jolliffe 2002) dimensionality reduction technique. Due to this reason and the space limitation, we do not compare the predictive performance of different individual feature types. Instead, we study the corresponding performance of individual data sources and its combinations. To do so, we evaluated "TweetFit" trained on different data source permutations. The Mobility and Venue Semantics data representations were treated as one data source, namely, "Venue Semantics & Mobility", since both of them were extracted from Foursquare check-ins data. We also did not evaluate the "BMI Trend" prediction performance on independent sources since there are only a few users with all data sources available in the "BMI Trend" test set.

Table 2: Evaluation of the "TweetFit" framework trained on independent data sources and data source combinations.

Data Source Combination	BMI category	
	R_M/P_M	$F_{1,M}$
Visual (V)	0.049/0.188	0.077
Ven. Sem. & Mob. (VSM)	0.194/0.107	0.137
Sensors (S)	0.153/0.158	0.155
Textual (T)	0.229/0.146	**0.178**
V + S	0.174/0.201	0.186
V + T	0.126/0.245	0.166
V + VSM	0.161/0.154	0.157
T + VSM	0.160/0.204	0.179
S + VSM	0.163/0.233	**0.191**
S + T	0.148/0.270	**0.191**
V + T + VSM	0.126/0.233	0.163
S + T + V	0.137/0.207	0.164
S + T + VSM	0.182/0.236	0.205
S + VSM + V	0.180/0.283	**0.221**
All Data Sources	0.214/0.292	**0.246**

First, we examine the contribution of different data sources towards wellness profile learning and source integration ability. An interesting observation comes from the data source combination results (see Table 2), where the combinations of "Sensors + Text" and "Sensors + Venue Semantics & Mobility" return the best performance and seem to be the most influential among other bi-source combinations. More impressive results can be gained from triplet combinations, where the combination of "Visual + Sensors + Venue Semantics & Mobility" performs the best. Based on these results, we can conclude that the Sensor data is of crucial importance for individual wellness profile learning since it is the only data source included in all best-performing data source combinations. This observation can also be interpreted by the ability of sensor data to represent users' actual physical condition, which is directly related to users' BMI category and "BMI Trend". Another explanation is the richness of the sensor data since in addition to exercise semantics it also carries the high-grained sequential data, which may not be available for other conventional social media sources. Summarizing the above, **we respond to RQ2 by highlighting the vital role of sensor data for the task of individual wellness profile learning** and suggesting its usage in further wellness-related research.

Let's now describe the single-source evaluation results (see Table 2). It is interesting to note that in the case of learning from independent data sources, our framework trained on Text modality performs the best, while those trained on Sensors and Venue Semantics & Mobility data ranks 2nd and 3rd place, respectively. First, the superiority of Text data over other modalities can be explained by its quantitative dominance (see Table 1). At the same time, the Sensor data holds the 2nd position, which again highlights its importance. Finally, being trained on visual data, "TweetFit" performs the worst among all other data sources. One possible explanation is the high level of noise in users' Instagram photos. Furthermore, the differences with the previous study can be interpreted by the generality of ImageNet image concepts (Deng et al. 2009) that could be useful for the general task of demographic attribute inference (Farseev et al. 2015), but less effective for the more narrow problem of individual wellness profile learning. In conclusion, we would like to highlight the Text and Sensor data sources as the strongest contributors towards wellness profile learning.

Evaluation Against Baselines

To answer **RQ1**, we compare the following user profiling approaches: **1)Random Forest** — strong baseline for the user profile learning (Farseev et al. 2015), where the number of trees equals to 105 and 25 for the "BMI category" and "BMI Trend" inference, respectively. **2)MTFL** (Liu, Ji, and Ye 2009) — the $\ell_{2,1}$ norm regularized multi-task learning with $\alpha = 0.5$. **3)iMSF** (Yuan et al. 2012) — the sparse $\ell_{2,1}$ norm regularized multi-source multi-task learning, with $\alpha = 0.4$; **4)MSE** — multi-source user profiling ensemble, proposed by Farseev et al. (2015). **5)TweetFit** — our framework trained based on all data sources, with $\alpha = 0.1$.

Table 3: Evaluation of the "TweetFit" framework against user profiling baselines.

Method	BMI category		"BMI Trend"	
	R_M/P_M	$F_{1,M}$	R_M/P_M	$F_{1,M}$
MSE	0.141/0.145	0.142	0.634/0.655	0.644
R.Forest	0.135/0.226	0.169	0.333/0.863	0.480
iMSF	0.171/0.174	0.172	0.649/0.649	0.649
MTFL	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**

The evaluation results are presented in Table 3. The results show that "TweetFit" achieve the best performance in both inference tasks as compared to all the baselines. This points conclusively towards the **positive answer to RQ1**. Specifically, we conclude that it is possible to improve the individual wellness profiling performance by integrating data from multiple social media sources and sensors. Moreover, "TweetFit" outperforms other state-of-the-art approaches from the multi-task learning family as well

as non-linear baselines. This shows the effectiveness of the framework in integrating data from wearable sensors and social media for wellness profiling.

Although "TweetFit" outperforms the baselines, the achieved BMI category prediction performance could not yet be used in real-world applications. This highlights BMI category inference as a challenging problem. An improvement could possibly be achieved by introducing inter-source correlation into the multi-source learning objective (Akbari et al. 2016). In future works, we also plan to apply a different BMI categorization scheme or treat BMI inference as a regression task, aiming to embed individual wellness profiling into the **bBridge**[4] social multimedia analytics platform (Farseev, Samborskii, and Chua 2016);

Conclusions

In this work, we presented one of the first studies on individual wellness profiling from sensor and social media data, which was handled by training the "TweetFit" framework to infer BMI category and "BMI Trend" personal wellness attributes. To facilitate further research, we released the multi-source multimodal dataset (Farseev 2017), which can be used for research on: user profiling (Farseev et al. 2016); multi-view timeline analysis (Jain and Jalali 2014; Akbari et al. 2016); and user identification across multiple social networks.

Acknowledgements

NExT research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its IRC@SG Funding Initiative.

References

Abbar, S.; Mejova, Y.; and Weber, I. 2015. You tweet what you eat: Studying food consumption through twitter. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM.

Akbari, M.; Hu, X.; Liqiang, N.; and Chua, T.-S. 2016. From tweets to wellness: Wellness event detection from twitter streams. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. AAAI.

Banaee, H.; Ahmed, M. U.; and Loutfi, A. 2013. Data mining for wearable sensors in health monitoring systems: a review of recent rrends and challenges. *Sensors*.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research*.

Bracewell, R. 1965. The fourier transform and it's applications. *New York*.

Buraya, K.; Farseev, A.; Filchenkov, A.; and Chua, T.-S. 2017. Towards user personality profiling from multiple social networks. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. AAAI.

Caruana, R. 1997. Multitask learning. *Machine Learning*.

Corbin, C. B.; Welk, G.; Corbin, W. R.; and Welk, K. 2001. *Concepts of Fitness and Wellness*. McGraw-Hill.

[4]bbridge.net

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.

Eggleston, E. M., and Weitzman, E. R. 2014. Innovative uses of electronic health records and social media for public health surveillance. *Current Diabetes Reports*.

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

Farseev, A.; Akbari, M.; Samborskii, I.; and 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).

Farseev, A.; Samborskii, I.; and Chua, T.-S. 2016. bbridge: A big data platform for social multimedia analytics. In *Proceedings of the 2016 ACM on Multimedia Conference*. ACM.

Farseev, A. 2017. NUS-SENSE Multi-Source Social-Sensor Dataset. http://nussense.farseev.com.

Field, A. E.; Coakley, E. H.; Must, A.; Spadano, J. L.; Laird, N.; Dietz, W. H.; Rimm, E.; and Colditz, G. A. 2001. Impact of overweight on the risk of developing common chronic diseases during a 10-year period. *Archives of Internal Medicine*.

Fried, D.; Surdeanu, M.; Kobourov, S.; Hingle, M.; and Bell, D. 2014. Analyzing the language of food on social media. In *Proceedings of the International Conference on Big Data*. IEEE.

GlobalWebIndex. 2016. Gwi social report q3 2016. http://insight.globalwebindex.net/social.

Jain, R., and Jalali, L. 2014. Objective self. *MultiMedia, IEEE*.

Jolliffe, I. 2002. *Principal Component Analysis*. Wiley Library.

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

Mejova, Y.; Zhang, A. X.; Diakopoulos, N.; and Castillo, C. 2014. Controversy and sentiment in online news. *arXiv preprint arXiv:1409.8152*.

Mejova, Y.; Haddadi, H.; Noulas, A.; and Weber, I. 2015. # foodporn: Obesity patterns in culinary interactions. In *Proceedings of the 5th International Conference on Digital Health*. ACM.

Qu, Y., and Zhang, J. 2013. Trade area analysis using user generated mobile location data. In *Proceedings of the 22nd International conference on World Wide Web*. International World Wide Web Conferences Steering Committee.

Rifkin, R., and Klautau, A. 2004. In defense of one-vs-all classification. *Journal of machine learning research* 5(Jan).

Song, X.; Nie, L.; Zhang, L.; Liu, M.; and Chua, T.-S. 2015. Interest inference via structure-constrained multi-source multi-task learning. In *Proceedings of the International Joint Conference on Artificial Intelligence*. ACM.

WHO, W. H. O. 2011. Global database on body mass index. 2011. *Global Database on Body Mass Index*.

Yuan, L.; Wang, Y.; Thompson, P. M.; Narayan, V. A.; and Ye, J. 2012. 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)*.