A

Tweet can be Fit: Integrating Data from Wearable Sensors and Multiple Social Networks for Wellness Profile Learning

ALEKSANDR FARSEEV, National University of Singapore
TAT-SENG CHUA, National University of Singapore

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 (BMI) category 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. At the same time, 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, we infer personal wellness attributes by utilizing our proposed multi-source multi-task wellness profile learning framework — "WellMTL", which can handle data incompleteness and perform wellness attributes inference from sensor and social media data simultaneously. To gain insights into the data at a global level, we also examine correlations between first-order data representations and personal wellness attributes. Our experimental results show that the integration of sensor data and multiple social media sources can substantially boost the performance of individual wellness profiling.

Categories and Subject Descriptors: H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing

General Terms: Algorithms, Experimentation, Performance

Additional Key Words and Phrases: Multiple Sources Integration; Wellness Profile Learning; Multi-Task Learning; Wearable Sensors; Personal Lifestyle Assistance

ACM Reference Format:
Aleksandr Farseev and Tat-Seng Chua, 2017. Tweet can be Fit: Integrating Data from Wearable Sensors and Multiple Social Networks for Wellness Profile Learning *ACM Trans. Inf. Syst.* V, N, Article A (January YYYY), 35 pages.
DOI: 0000001.0000001

1. INTRODUCTION

In the past decade, the impact of social multimedia services on people's daily life have drastically increased [Filchenkov et al. 2014]. For example, more than half of American smartphone users were reported to spend an average of 144 minutes per day browsing their mobile devices [Morningside 2013], aiming to stay socially connected. 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). This data is of crucial importance for research in the wellness domain since it describes users' actual physical condition [Park et al. 2016],

NExT research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its IRC@SG Funding Initiative.
The Microsoft Windows Azure Cloud and Microsoft MSDN subscription were provided by "MB-Guide"[1] and "bBridge"[2] projects.
Author's addresses: A. Farseev and T.-S. Chua, School of Computing, 13 Computing Drive, Singapore 117417.

DOI: 0000001.0000001

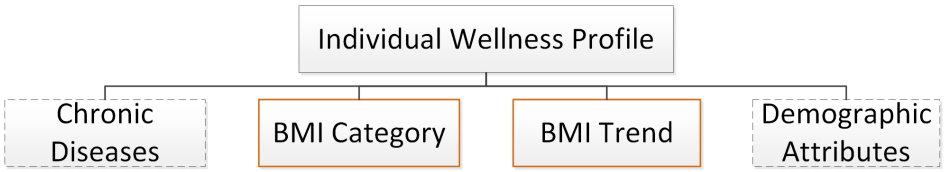


Fig. 1: Personal wellness attributes inside individual wellness profile

which is related to users' well-being. At the same time, recent works demonstrate the great potential of social media data for wellness-related research [Abbar et al. 2014; Chou et al. 2014; Mejova et al. 2015; Weber and Achananuparp 2015; Akbari et al. 2016; Park et al. 2016]. Considering that nowadays most Internet-active adults actively use more than three social network services in their everyday life [GlobalWebIndex 2016], and with the wide availability of data from wearable sensors, it is natural to combine multimodal content from different social networks with sensor data for joint modeling [Jain and Jalali 2014; Jalali and Jain 2015; Park et al. 2016]. Such integration will narrow the gap between users' online representation and actual physical status, which is the right step towards realizing the ideal of wellness profiling.

This article focuses on the problem of individual wellness user profiling based on data from multiple social networks and wearable sensors. An individual wellness user profile typically involves personal user attributes [Farseev et al. 2016] such as BMI[3] category, demographics [Farseev et al. 2015b], personality [Buraya et al. 2017], or chronic disease tendency [Akbari et al. 2016], as outlined in Figure 1. In our study, we focus only on two personal wellness attributes — BMI category and "BMI Trend". According to World Health Organization (WHO), BMI measure is categorized into one out of 8 categories, namely "Severe Thinness", "Moderate Thinness", "Mild Thinness", "Normal", "Pre Obese", "Obese", "Obese II", and "Obese III" [WHO 2011]. The "BMI Trend" is the direction of BMI fluctuation over time (Increase/Decrease) for those whom we have multiple measurements. The BMI category and "BMI Trend" attributes are closely related and correlated to one's overall health. For example, in [Field et al. 2001], it was 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: (i) BMI category can be further used in public health domain to monitor wellness tendencies of social media users at the global level; and (ii) "BMI Trend" information can be utilized by users to rectify their lifestyle (i.e. via an interactive mobile APP), and by doctors to gain a complete picture of patients' health.

There are **two major challenges** in addressing individual wellness profile learning. First, multi-source multi-modal data must be **efficiently represented**. Besides the textual data, social media services involve data of various modalities. For example, in Instagram[4], users share recently taken pictures and videos, while in Endomondo[5] users post information about their workouts, which is strongly dependent on the temporal and spatial aspects. Integration of such heterogeneous multimodal data sources requires the development of efficient and mutually consistent data representation approaches. Second, it is essential to perform effective joint **multi-source data modeling.** Joint data modeling for individual wellness profile learning is a tough challenge since the data from independent media sources is different in nature. Furthermore,

[3]The BMI measure is defined as the body mass divided by the square of the body height. It is widely used to quantify the amount of tissue mass (muscle, fat, and bone) in an individual [WHO 2011]
[4]http://instagram.com
[5]http://endomondo.com

multi-source data is often incomplete, which means that some users may not be active on all social networks. Lastly, personal wellness attribute categories (classes) are often inter-related (i.e. the adjacent BMI categories could unite users of similar lifestyle and social media content [Mejova et al. 2015]), which must be taken into account at the learning stage. Development of a learning framework that can handle all these issues is a crucial challenge.

Inspired by the aforementioned challenges, we seek to address the following **research questions**:

(1) **What is the relationship between social media data representations, wearable sensor data representations, and wellness attributes?**
(2) **Is it possible to improve the performance of BMI category and "BMI Trend" inference by joint modeling of social media and sensor data?**
(3) **Is it possible to improve the performance of BMI category and "BMI Trend" inference by incorporating inter-category relatedness into the learning process?**

To answer the above research questions, we present a new computational wellness profiling framework "WellMTL". First, we study relationships between different multimodal data representations (image concepts, venue categories, sports activities, sensor features, mobility patterns, latent topics) and wellness attributes. Second, we construct users' individual wellness profiles by predicting their BMI category and "BMI Trend". We treat the individual wellness profile learning as a regularized multi-task learning (MTL) problem [Evgeniou and Pontil 2004]. Specifically, we represent different data source combinations as MTL "tasks". Concurrently, we consider inter-category relationship by regularizing the MTL model by learning "similar" categories in a mutually-consistent fashion. Third, we perform error and feature importance analysis aiming to understand which BMI categories can be inferred more accurately and which particular feature types facilitate better inference of each BMI category. Lastly, we discuss the inter-source data correlation and the application of our framework.

The main **contributions** of our study are summarized as follows:

— **Formulation of a multi-task learning framework for wellness attribute inference.** We formulated personal wellness profiling of users from multiple social networks as a regularized multi-task learning problem. In this proposed "WellMTL" framework, we represented the social media and sensor data in a unified manner.
— **Extensive data analysis.** To offer insights on the need to combine data from multiple social networks and wearable sensors within one integrated framework, we performed first-order and high-order statistical analysis of the relationship between different data representations and wellness attributes. Additionally, we conducted inference error analysis and feature importance analysis on each wellness category.

The rest of the paper is organized as follows: Sections 2 and 3 present related work and data description, respectively. The data analysis at a global level and the individual wellness profiling are described in Sections 4 and 5, respectively. We evaluate individual wellness profiling performance, conduct error analysis, and perform feature importance in Section 6. This is followed by a discussion on inter-source data relationships in Section 7. The real-world application of our proposed framework is outlined in Section 8, while framework's limitations and future work directions are outlined in Section 9. We then conclude the paper in Section 10.

2. RELATED WORK

Recently, several research studies were devoted to cross-social network user activity analysis and its application in wellness domain. For example, medical and healthcare

communities suggested the utilization of social media data as a useful resource to monitor food-related habits for obese and diabetes patients [Eggleston and Weitzman 2014]. Culotta et al. [2014] performed an analysis on Twitter data aiming to predict US citizens' health-related statistics such as obesity or teen pregnancy, while Chou et al. [2014] conducted a qualitative study on weight-related interactions of Twitter and Facebook users. Another study [Fried et al. 2014] was targeted to predict diabetes and overweight rates for 15 US cities. Mejova et al. [2015] and Silva et al. [2014] attested to the predictive power of Foursquare-based features for the group obesity inference task and cultural differences analysis, while Sharma et al. [2015] leveraged Instagram data in an attempt to measure food calories from image-related hashtags. Later, Jurgens et al. [2015] drew first-order statistics and performed preliminary analysis at a group level of Fitocracy[6] users' exercising activity, while Weber et al. [2015] guided machine learning models based on data from MyFitnessPal[7] to predict a success of users' personal diet plans. At last, Park et al. [2016] studied publishing traits of social media users who openly share their individual health and fitness related information, while Akbari et al. [2016] introduced a learning framework for personal wellness events categorization. As noted, there were research efforts made towards wellness lifestyle analysis and the results showed enormous potential of social media data to assist 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 some research efforts done on multi-source learning for user profiling. For example, Liu et al. [2009] implemented the so-called $\ell_{2,1}$ regularization to obtain sparse data representations for feature selection purpose, which is useful in learning from high-dimensional data. However, the data source integration was carried out in an early fusion manner, where all the features were combined into one vector before model training. Such a data integration strategy may result in the so-called "curse of dimensionality" problem and, consequently, suboptimal classification performance. Later, Farseev et al. [2015b] introduced an ensemble learning solution, aiming to combine multi-source multimodal data for demographic user profile learning. Nonetheless, the model was trained independently on each source and consolidated in a "late-fusion" manner, which does not fully take advantage of multi-source data [Song et al. 2015b]. Finally, Song et al. [2015a] proposed a multi-source learning framework for Volunteerism tendency prediction for users of different social networks. The missing data was induced by the proposed constrained Non-Negative Matrix Factorization (NMF) approach [Paatero and Tapper 1994], which could introduce bias into the final data representations and, thus, may not apply to sparse real-world datasets. The above works are relevant to our study, but may not be directly applicable in our case due to the drawbacks of the utilized data integration approaches.

The problem of wellness profile learning from multiple social networks and sensor data exhibits dual-heterogeneities: each wellness category corresponds to features from multiple sources. Towards this end, the most related work lies in the area of multi-view multi-task learning. Yuan et al. [2012] proposed to treat the problem of multi-source fusion as a multi-task learning for enhancing Alzheimer's Disease prediction. The proposed sparse feature learning approach is relevant to this study regarding its feature selection ability, but could not be directly applied to our case due to its formulation for binary classification. Later, Farseev and Chua [2017] proposed another sparse multi-task learning framework for wellness attributes inference. The proposed

[6]http://www.fitocracy.com
[7]http://myfitnesspal.com

framework is able to handle incomplete data sources and feature selection simultaneously via efficient $\ell_{2,1}$ regularization. However, the framework does not consider the inter-category relationship, which seems to be essential for our task. Aiming to model such a relationship, He and Lawrence [2011] proposed a graph-based iterative framework for multi-view multi-task learning in the context of text classification. Given task pairs, the model projects them to a new latent space based on the sources' similarities. However, such an approach does not allow for generating predictive models based on data samples that do not share any common source, which is an important restriction in the real-world settings. Inspired by the limitations above, Song et al. [2015b] proposed a structure-constrained multi-task learning framework for user interests inference from multi-source data. To model user interest relation, authors included external knowledge in the form of "interest relationship weights", which makes the framework biased towards particular datasets and tasks. Finally, Farseev et al. [2017] proposed the utilization of multi-layer graph clustering for cross-domain venue category recommendation. Specifically, authors introduced an approach for automatic inference of inter-source relationship weights and plugged the obtained weights in multi-layer subspace-regularized spectral clustering objective. However, the method is not applicable to our problem, since it requires user relationship graph to be defined. Due to the reasons above, it is essential to implement a fully-automated multi-source individual wellness profiling approach that would not rely on external knowledge and data completion techniques, while considering the relationships between different wellness categories.

3. DATA FROM SOCIAL MEDIA AND SENSORS

It is well-known that for building a comprehensive individual user profile, it is essential to incorporate multimodal data from various sources that represent users from multiple perspectives [Farseev et al. 2015b; Song et al. 2015b]. At the same time, to build a complete personal wellness profile, it is necessary to utilize information about users' physical health [Corbin et al. 2001; Akbari et al. 2016]. 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]. For example, users often post in microblogs about their recent activities and health updates [Akbari et al. 2016]. This information is related to users' wellness and useful for individual wellness profiling [Akbari et al. 2016]. Second, in our previous work [Farseev et al. 2015b], we observed that visual data plays a significant 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. For example, users may post pictures of food they eat [Sharma and De Choudhury 2015], which may affect users' BMI and can be used by inference models to predict their BMI category and "BMI Trend" attributes. Third, it was reported that data from location-based social networks is useful for obesity estimation at a group level [Mejova et al. 2015], which demonstrates its potential for use in personal BMI category and "BMI Trend" prediction task as well. Finally, data from wearable sensors was found to be valuable for tackling the activity recognition [Lara and Labrador 2013] and health monitoring [Banaee et al. 2013] problems, which shows its potential towards individual wellness profile learning.

Considering the above, in this work we utilize the following social networks:

— Twitter (the largest English-speaking microblog) micro-posts to be used as a textual data source.

— Instagram (the largest mobile Image-sharing service) pictures and its descriptions (comments) for use as image and textual data sources.
— Foursquare (the largest purely location-based social network) check-in records and its corresponding comments (shouts) were adopted as venue semantics, mobility, and textual data sources.
— Endomondo (one of the largest sports activity-tracking social networks) workouts were used as a sensor data source and for ground truth construction (BMI category and "BMI Trend" attributes).

It is worth mentioning that other than providing the so-called, exercise semantics (i.e. sports activity type, as in Fitocracy and MyFitnessPal), 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, and Heart Rate (in cases when a heart rate sensor is set up). The ground truth labels can be derived from publicly accessible Endomondo user profiles' web pages, which often include such personal attributes as Country of Residence, Postal Code, Age (Birthday), Gender, Height and Weight. These attributes are either manually input by Endomondo App users or automatically measured by connected "smart" sensors (i.e. FitBit Aria Smart Scale[8]). The above dictates that Endomondo data goes beyond representing users from one more modality, but bridges the gap between social media-based users' representation and their actual physical activities and condition. The BMI category and "BMI Trend"[9] ground truth extracted is reliable, since Endomondo users are encouraged to provide the correct weight and height values, which are further used by Endomondo App to estimate the calorie spent, BMI, and exercise efficiency. Furthermore, assuming that most of the users are familiar with their height $\pm$ 2 cm (i.e. 165 cm), according to the BMI category classification [WHO 2011], such users will be assigned to the BMI category of "Overweight" in cases when their weight is in the range $68 - 82$ kg. The above means that the BMI category of each user can be labeled correctly in most of the cases. Furthermore, some other recent works confirmed the reliability of self-reported social media wellness attributes [Weber and Achananuparp 2015; Akbari et al. 2016; Park et al. 2016].

3.1. Cross-Network User Identification And Data Collection

To collect data generated by the same individuals from multiple social networks, it is necessary to solve the task of cross-network user identification [Zafarani and Liu 2013]. One approach to solving this issue is the utilization of social network aggregation platforms, such as About.me[10]. About.me enables people to create a public online "name card" that includes a self-described biography and links to the one's social network accounts and personal websites. This information is usually sufficient for further multi-source data collection. However, according to Alexa Internet Survey[11], About.me is ranked as an unpopular website ($9,804$th worldwide position), while some research groups describe About.me users as atypical [Lim et al. 2015]. The tendency of About.me users to leverage on multiple social networks aiming to benefit from having an online content aggregation point could explain their abnormal online behavior [Lim et al. 2015]. Consequently, their generated data could be biased by their online goals

[8]http://www.fitbit.com/sg/aria
[9]The "BMI Trend" was only computed for those users, who have changed their weight in profile during the data collection process
[10]http://about.me/
[11]http://www.alexa.com/siteinfo/about.me Retrieved on 26 April 2017

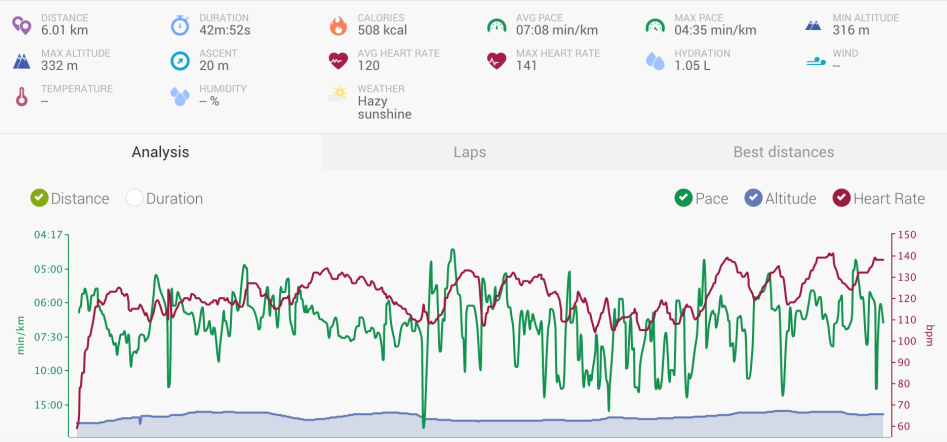


Fig. 2: Publicly available Endomondo workout

and may lead to sub-optimal user profiling performance. Furthermore, the available amounts of appropriate About.me data does not allow for conducting research in Wellness domain at a large scale due to the lack of ground truth.

Motivated by the gaps mentioned above, in this work we propose the new approach for harvesting social multimedia data at a large scale. Specifically, we utilize one social network (i.e. Twitter) as a "sink" for re-post data from other social networks (Instagram, Foursquare, Endomondo), so that the linkage between social media portals can naturally be obtained with close to zero error rate. The data collection process can be described by the following three steps:

— **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[12]. Specifically, we selected those users who have recently posted the hashtags "#Endomondo" or "#endorphines" together with the expression "I just finished" in their Twitter timeline, which allows us to find users who are active in both Endomondo and Twitter.
— **User-generated content collection:** We then started a Twitter "stream"[13] that involve all "seed" users and download the multi-source user-generated content of these users (via URLs to the original posts, provided in tweets). Such a data collection approach makes it possible to avoid the user identification problem, since it fulfills the cross-network account mapping whenever users perform a cross-network activity (i.e. posting an Endomondo image on Twitter). To be confident in cross-network account mapping correctness, we removed all retweets and ensured that all cross-network messages were automatically tweeted via cross-linking functionality of the corresponding social media Apps (Endomondo, Instagram, Swarm).
— **Ground truth collection:** During the Twitter crawling process, we conducted daily monitoring of Endomondo users' accounts and recorded all the BMI updates during the whole data collection period. Users' Weight and Height updates were used to compute their BMI, which was utilized to estimate "BMI Trend" ground truth attribute.

[12]http://dev.twitter.com/rest/public
[13]http://dev.twitter.com/streaming/public

Fig. 3: Publicly available Endomondo user profile

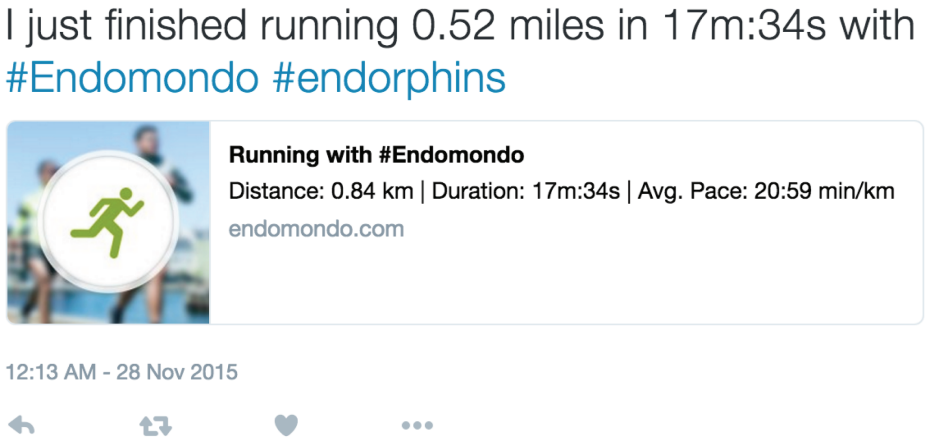


Fig. 4: Twitter repost of an Endomondo workout

Table I lists the number of posts and ground truth records in our collected dataset. We note that all users in the dataset have contributed data to at list two social networks, where one of them is Twitter.

To encourage further research, the collected data was released to public [Farseev and Chua 2017] as NUS-SENSE dataset[14].

3.2. Data Representation

As mentioned earlier, efficient data representation is a significant step in multi-source profile learning pipeline. Its importance is dictated by its ability to bridge the semantic gap between different data modalities by producing mutually-consistent data representations. The extracted features are described below.

[14]http://nussense.farseev.com

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

Data Source	Twitter	Endomondo	Foursquare	Instagram
Total No. of Posts	$1,676,310$	$140,926$	$19,743$	$48,137$
Total No. of Users	$5,375$	$4,205$	609	$2,062$
Average No. of User Posts	311	33	32	23
No. of Age labels	$3,974$	$3,111$	427	$1,525$
No. of Gender labels	$5,375$	$4,025$	609	$2,062$
No. of BMI category labels	$1,052$	870	116	372
No. of "BMI Trend" labels	152	136	18	51

Text Features: In our study, we extracted textual data from the following data sources: *Twitter tweets*, *Instagram image captions*, *Instagram image comments*, and *Foursquare check-in comments*. All textual data was aggregated at the individual user level and filtered/corrected by using Microsoft Office Spell Checker[15]. More specifically, we extracted the following features:

— *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 et al. 2003]. We trained the topic model with $T = 50$ topics (revealed the lowest perplexity score on interval from $T = 1, 5, ..., 200$), with $\alpha = 0.5$, $\beta = 0.1$ (determined empirically). Feature vector was constructed as a distribution of each user among detected latent topics: $\mathbf{u}_{i,LDAText} = (tl_{i,1}, tl_{i,2}, ..., tl_{i,50})$.
— *Writing style features.* As in [Farseev et al. 2015b], we extracted such writing style features as: *number of mistakes per post*, *number of slang words per post*, *average post sentiment*, etc. In total, 14 features were extracted: $\mathbf{u}_{i,HeurText} = (th_{i,1}, th_{i,2}, ..., th_{i,14})$.
— *Lexicon-based features.* We used crowd-sourced lexicon 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 incorporating the Twitter Food Lexicon [Silva et al. 2014]. Lastly, all the food-related features were grouped in one user-dependent vector of size 32: $\mathbf{u}_{i,FoodText} = (tf_{i,1}, tf_{i,2}, ..., tf_{i,32})$.

Venue Semantics Features: Similar to [Farseev et al. 2015b], we represented location data as a distribution of users' check-ins among 764 Foursquare venue categories. Such users' representation describes venue semantics preferences of Foursquare users and can be related to users' food and activity preferences [Mejova et al. 2015]. To overcome the data sparsity problem, we further reduced the data dimensionality by extracting Top 86 principal components [Jolliffe 2002] (preserves 85% of variance) and representing the venue semantics data as: $\mathbf{u}_{i,LocCatPCA} = (lcPCA_{i,1}, lcPCA_{i,2}, ..., lcPCA_{i,86})$.

Mobility and Temporal Features: As stated earlier, user mobility and temporal aspects are important in user profile learning [Farseev et al. 2015b; Noulas et al. 2012]. The mobility features were computed 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 the check-in venue semantics). AOIs were obtained by performing density-based clustering [Sander et al. 1998] over the check-ins of each user (with

[15]http://products.office.com

$\epsilon = 0.05$, the minimum number of points $MinPts = 3$) and consider the convex hull of each cluster as a new AOI. The user's AOIs represent their geographical mobility patterns [Noulas et al. 2012; Qu and Zhang 2013] and is correlated to their wellness lifestyle. We extracted the following 12 mobility and temporal features ($\mathbf{u}_{i,MobTemp} = (mt_{i,1}, mt_{i,2}, ..., mt_{i,12})$):

— *Average number of posts during each of the* 8 *daytime durations*, where each time duration is 3 hours long (i.e. $15 - 18$). The feature is an indicator of user's temporal online activity and related to their wellness [Chahal et al. 2013].
— *Number of areas of interest (AOI)*. The feature reflects the mobility side of user's physical activity. For example, users with more AOIs may maintain more intensive lifestyle. AOIs also represent user's frequent areas of activity (i.e. home, office, university/school) [Qu and Zhang 2013] and may indicate how far users are willing to travel on a daily basis. It, in turn, reflects users' wellness attributes, since it is related to their lifestyle.
— *Median size of user's AOIs*[16], which indicates users' mobility inside each AOI. The feature is an indicator of user's everyday traveling habits inside their main activity areas. It can, thus, indicate the approximate distance that the users are willing to travel on foot.
— *The normalized number of AOI outliers*. The feature shows how often users visit places that are not located inside their AOI, which may reflect the users' "conventionalism" level. Mainly, it shows how often users deviate from their regular mobility patterns, which is related to users' physical activity level.
— *Median distance between AOIs*. This feature reflects users' movement at intra-city/inter-city/international level. Specifically, it shows how often and how far users travel between their activity zones. It could be daily trips to job place, vocation journeys or business trips. Such a feature is related to users' lifestyle and, thus, wellness.

Even though the venue semantics and mobility features were extracted from the same data source, they are different in nature. Specifically, the venue semantics features represent users' geo-independent location preferences [Wang et al. 2015] (i.e. Restaurant, Cinema, Church). At the same time, users' mobility features describe users' movement in space and related to their activity level in a particular geographical region [Noulas et al. 2012].

Visual Features: In order to represent data from visual modality, we computed distribution of each Instagram photo among the 1000 ImageNet visual concepts [Deng et al. 2009]. We then averaged the values of each 1000 visual concept to obtain a user-dependent feature vector of size 1000. The image concept detection was performed by leveraging on GPU implementation [Jia et al. 2014] of a state-of-the-art deep learning system GoogleNet [Szegedy et al. 2015]. Similar to venue semantics features, we extracted Top 150 principal components [Jolliffe 2002] (preserves 85% of variance) from image concepts data: $\mathbf{u}_{i,ImgConPCA} = (vcPCA_{i,1}, vcPCA_{i,2}, ..., vcPCA_{i,150})$.

Sensor Features: One of the most informative data sources for wellness profile learning and, at the same time, the most challenging regarding consistent representation, is the data from sensor devices. The main problem is the necessity to consider the spatial-temporal aspect of sensor data, since the static characteristics of each data sample can be biased toward individual's health and, thus, may not be suitable for further integration with other sources for joint profile learning. To represent sensor data consistently with other data modalities, we incorporated the following feature types:

[16]Where AOI size is defined as the median distance between a center of mass and all points inside AOI

— *Workout statistics.* We computed the following averaged features using all sensor data samples for each user: *Distance (Ascend/Descend)*, *Speed*, *Duration*, *Hydration*. Such features represent users' activity levels and is thus related to one's wellness attributes. In total, 10 features were extracted: $\mathbf{u}_{i,SensStat} = (ss_{i,1}, ss_{i,2}, ..., ss_{i,10})$.
— *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*. It is known that weather may affect one's exercise productivity. It can thus serve as a useful complimentary signal for wellness attribute inference. To make the feature set appropriate for model training, we represented each user as a distribution among 44 weather types from our dataset, computed through all user's workouts: $\mathbf{u}_{i,SensExt} = (se_{i,1}, se_{i,2}, ..., se_{i,44})$.
— *Workout semantics.* We also represented sensor data as a distribution of users' workouts among 96 Endomondo workout categories: $\mathbf{u}_{i,SensWCat} = (sc_{i,1}, sc_{i,2}, ..., sc_{i,96})$. The representation describes users' exercise preferences and also related to their wellness.
— *Frequency domain features.* It is important, to incorporate the temporal aspect of sensor data, since the signal fluctuations during an exercise are often related to one's wellness attributes. However, the user's workouts are not directly comparable due to the differences in the exercise types and duration, or the inconsistency in signal sampling rates. To incorporate the temporal aspect of sensor data without binding to actual signal duration, it is useful to project time series data into the frequency domain. To perform such a transformation, we extracted features for each workout, in three steps as follows: (i) We applied spline interpolation to fill in missing sensor signal measurements (in cases of sampling rate < 1Hz or when it is inconsistent).(ii) We transformed all data points from jth workout with r sensor signal measurements ($\mathbf{w}_{j,u_i} = (d_1, d_2, ..., d_r)$) into relative differences between subsequent sensor signal measurements: $\bar{\mathbf{w}}_{j,u_i} = (d_2 - d_1, d_3 - d_2, ..., d_r - d_{r-1})$. (iii) We applied 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 five sensor signal types, namely, *Altitude, Cadence, Speed, Heart Rate (HR), and Oxygen Consumption (Oxygen)*. We then merged these 5 vectors together and took an average among all workouts to obtain a frequency domain feature vector of size 495 for each user. Such a data representation is no longer inconsistent among different users, since it models the "changes" of sensor signal, rather than the absolute values of data points. To overcome the data sparsity problem, we reduced the frequency-domain data representation dimensionality by extracting its Top 54 principal components [Jolliffe 2002] (preserves 85% of variance) and represented the data as: $\mathbf{u}_{i,SensFreqPCA} = (sfPCA_{i,1}, sfPCA_{i,2}, ..., sfPCA_{i,54})$.

We summarize all the above features and their characteristics in Table II.

4. GLOBAL DATA ANALYSIS

To answer research question (1)[17], we analyzed our data at global level by utilizing Pearson Correlation Coefficient (r) to uncover the relationship between different data representations and wellness attributes. Specifically, we discovered and characterized the relationship between personal wellness attributes (BMI, "BMI Trend") and different data modalities, namely text, images, locations, and sensor measurements. To

[17] *What is the relationship between social media data representations, wearable sensor data representations, and wellness attributes?*

Table II: Features summary

Modality	Features type	Dimension	Red. dimension
Text	Latent topics [Blei et al. 2003]	50	50
	Lexicon-based [Farseev and Chua 2017]	32	32
	Writing style [Farseev et al. 2015b]	14	14
Image	ImageNet concepts [Deng et al. 2009]	1000	150 (PCA)
Location	Venue semantics [Farseev et al. 2015b]	764	86 (PCA)
	Mobility and Temporal	12	12
Sensor	Workout semantics	96	96
	Frequency domain	495	54 (PCA)
	Workout statistics	10	10

do so, we highlighted significant[18] correlations between different data representations and BMI/"BMI Trend" as shown in Figures 5,6, and 7, respectively. It is noted that we consider negative correlation to "BMI Trend" as being favorable since it means losing weight.

4.1. Data from wearable sensors

To study sensor data, we performed correlation analysis between wellness attributes and two types of sensor data representations: distribution among Fourier spectrum bins and workout categories (exercise semantics).

First, it can bee seen that users' BMI is negatively correlated to active sport types ("Dancing", "Running"), while positively correlated to moderately active sport types ("Scuba Diving", "Walking", "Cross-Training") (see Figure 6(b)). The possible reason is that people with higher BMI may select less intensive workout types due to their mobility restrictions and fitness level, while fitter ones often prefer more intensive workouts [Jurgens et al. 2015].

At the same time, there is a significant number of frequency bins from each sensor type that are correlated to users' BMI (Figure 6(a)): 5 "HR" bins, 4 "Altitude" bins, 30 "Cadence" bins, 5 "Oxygen" bins, and 3 "Speed" bins. The prevalence of "Cadence" bins features can be explained by the relatedness of cadence to exercises' tempo, which, in turn, is related to the actual fitness level of users: fitter users may perform sports activity with different intensity as compared to less fit ones. Lastly, the significant correlation to user' "BMI Trend" (Figure 7(a)) could be explained by the connection of exercise intensity and energy spending, which is related to weight loss.

To summarize, *both exercise semantics and frequency bins features are tightly knit to BMI and "BMI Trend"*, which can be seen from the correlation plots. Intuitively, it can be explained by the ability of sensor data to reveal knowledge about users' actual physical status, which is not available from other data sources. Due to the above, it is essential to incorporate sensor data into the learning process.

4.2. Data from social media

The Instagram image concepts "Diaper" and "Sweatshirt" are positively correlated to users' BMI, while "Over Skirt" and "Gown" image concepts reveal negative correlation (Figure 5(a)). The possible reason might be the difference of dressing preferences between users of different BMI: users with overweight problems may take pictures of themselves wearing baggy garments, while people with lower BMI may prefer slinky clothes instead. It is also interesting to note that "Rugby Ball" and "Soccer Ball"

[18]We utilize the significance test with the $\alpha = 0.05$ and data sample sizes are from Table I

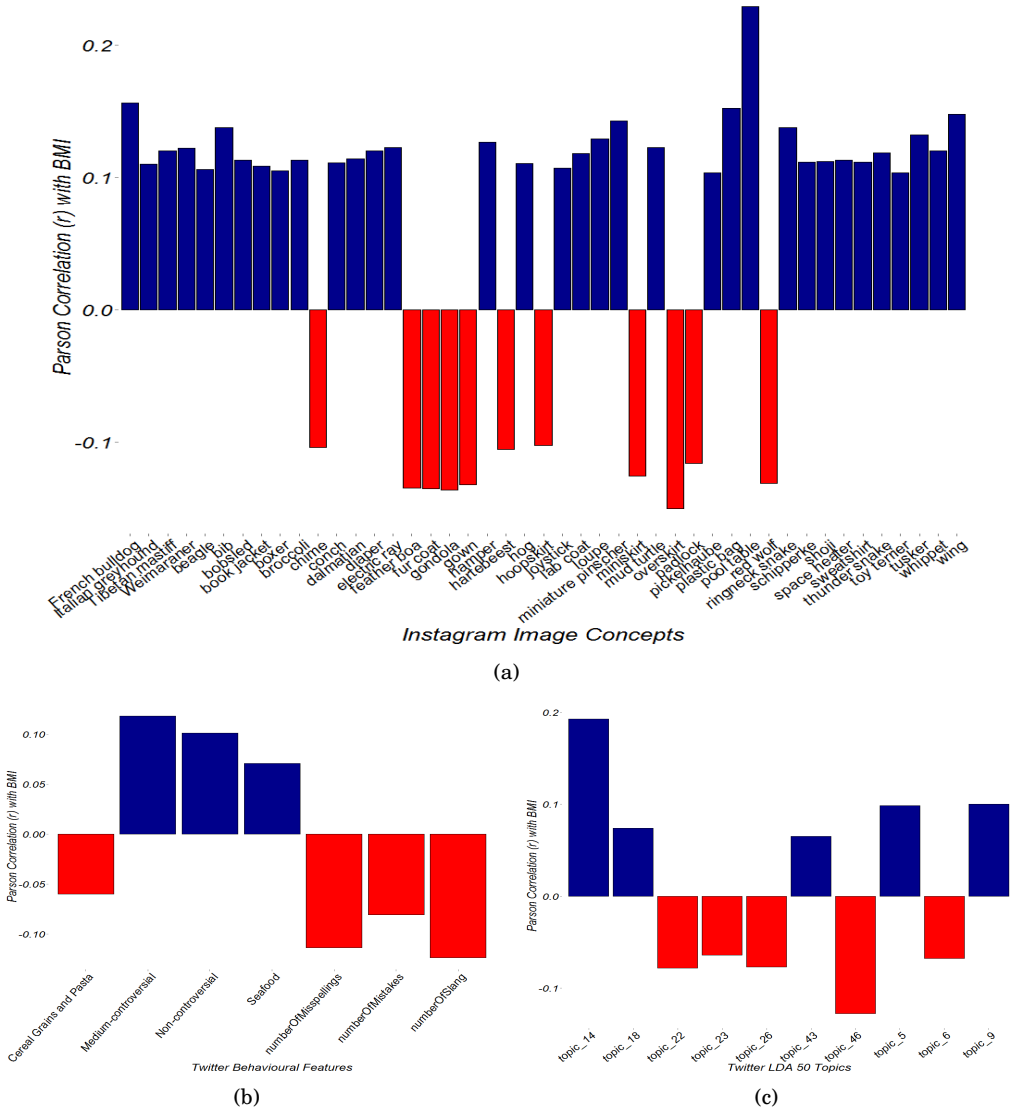
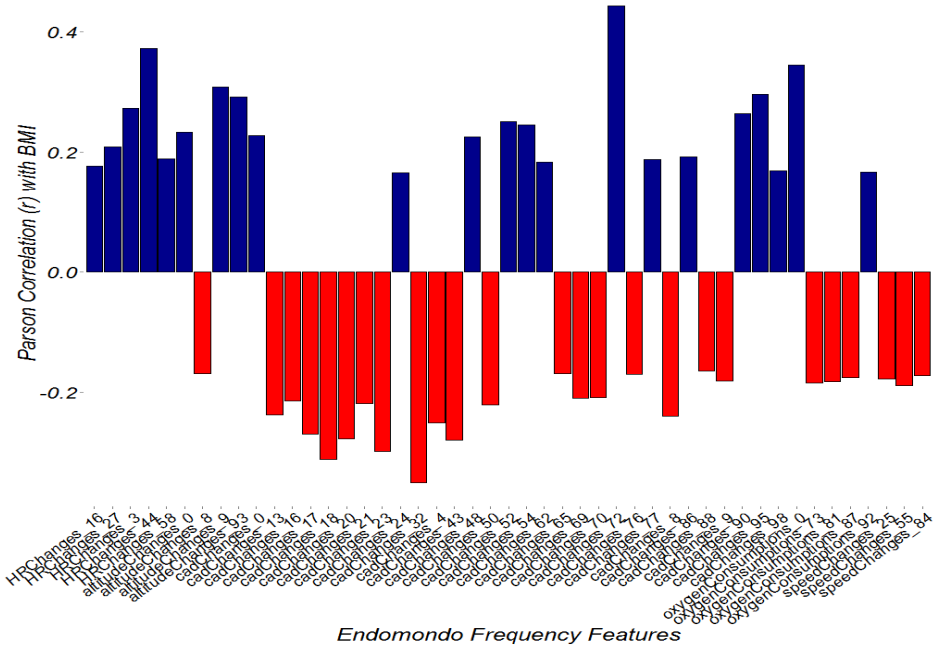
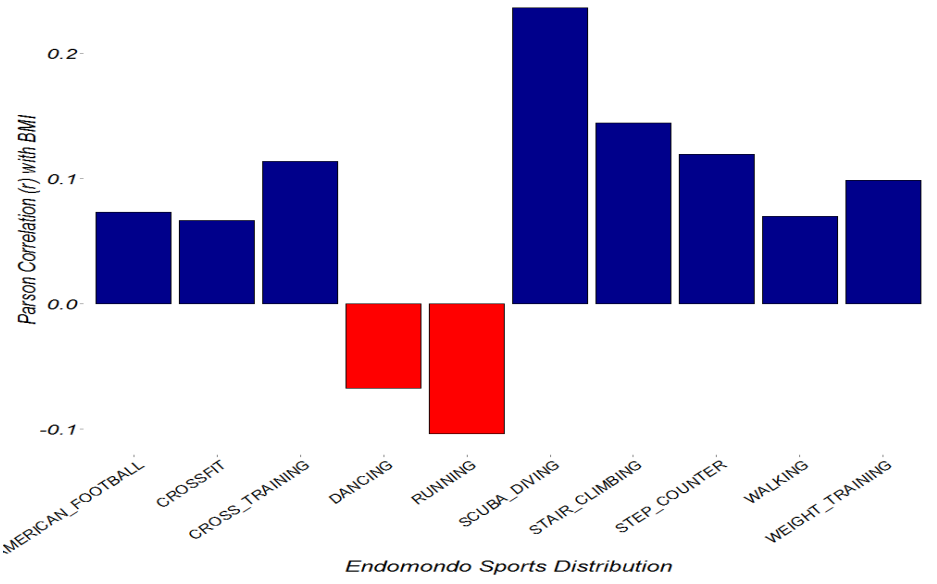


Fig. 5: The image (a) and textual (b, c) features are significantly ($\alpha = 0.05$) correlated to users' BMI data representations. The negative and positive correlations are colored in red and blue, respectively.

show the highest negative correlation level to "BMI Trend". The explanation could be that sports activities (represented by sport-related image concepts) help users to lose weight. Generally, it can be observed that *many Instagram image concepts are significantly correlated to personal wellness attributes*. The possible reason is that these image concepts represent wellness-related aspects of human life, such as diet, sports activities, and dressing preferences. They are useful indicators of users' wellness.

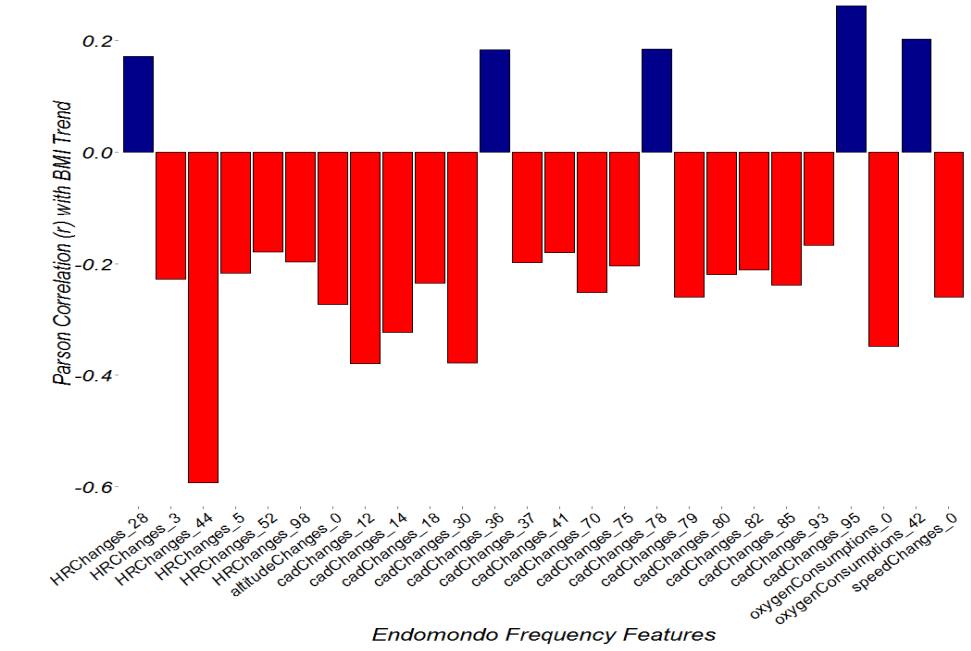


(a)

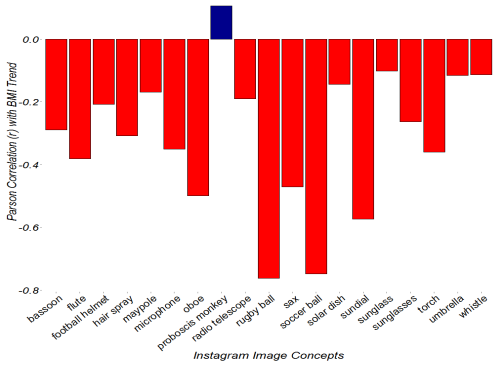


(b)

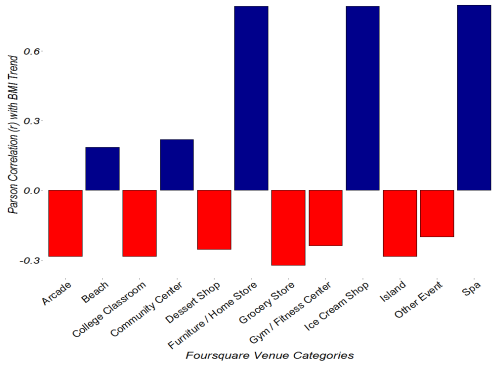
Fig. 6: The sensor features are significantly ($\alpha = 0.05$) correlated to users' BMI data representations. The negative and positive correlations are colored in red and blue, respectively.



(a)



(b)



(c)

Fig. 7: The image (a), location (b), and sensor (c) features are significantly ($\alpha = 0.05$) correlated to users' "BMI Trend" data representations. The negative and positive correlations are colored in red and blue, respectively.

As mentioned in a previous study [Mejova et al. 2015], Foursquare data is correlated to the country-level obesity traits. The above conforms well with our observations (Figure 7(c)), where the venue categories "Ice Cream Shop" and "Fitness Center" are significantly correlated to the tendency of gaining/losing weight, respectively. Since *many venue categories are related to users' dietary habits*, it is necessary to utilize this data in wellness profile learning.

Another interesting dependency can be observed between lexicon-based text features and the BMI category — the *significant correlation between automatically extracted (from user-generated text) food groups* ("Cereal Grains and Pasta", "Seafood"). Such observation is also consistent with previous studies [Mejova et al. 2015; Sharma and De Choudhury 2015; Akbari et al. 2016]. Inspired by these findings, we include text-based features in personal wellness profile learning.

Let's summarize the above observations. First, the global data analysis results are consistent with previous works [Farseev et al. 2015b; Song et al. 2015a] regarding the fact that multiple data sources describe users from different perspectives. Second, many data representations are significantly correlated to BMI and "BMI Trend" personal wellness attributes. At the same time, some data representations are not linearly correlated to them. These means that a feature selection mechanism must be integrated into the model. Third, it can be seen that BMI index negatively/positively correlated to numerous data representations. Essentially, this means that the data representations expose proportional (or inversely proportional) trends to the wellness attributes. This suggests that the adjacent BMI categories may represent users with similar social media behavior, which must be considered at the learning stage. Based on the above investigations, **we positively respond to the research question (1) and conclude that it is reasonable to incorporate all the harvested data sources into our individual wellness profiling framework**.

5. INDIVIDUAL WELLNESS PROFILING

To answer research question (2), it is necessary to develop and evaluate a machine learning solution that will fuse data from multiple social media sources and wearable sensors for wellness attributes inference. Meanwhile, to answer research question (3), it is essential to consider inter-category relatedness at the learning stage. In the following sections, we will address these issues one by one.

5.1. Problem Statement

The problem of multi-source personal user profile learning is not an easy task. First, in many real-world scenarios, multiple data sources are sparse and block-wise incomplete [Song et al. 2015a], which means that some groups of users may contribute content only to certain social networks (Figure 8). Traditionally, the problem can be solved by simply skipping incomplete data samples or by using data completion techniques such as NMF [Paatero and Tapper 1994; Song et al. 2015a]. However, these two approaches may lead to the "curse of dimensionality" problem in the first case, and to noise propagation in the second. To overcome these problems, it is necessary to develop an efficient source fusion technique [Farseev et al. 2015b]. Second, the categories of the wellness attributes are often inter-related. For example, the adjacent BMI categories (i.e. "Obese II" and "Obese III") are more related to each other as compared to those that represent completely different BMI groups (i.e. "Severe Thinness" and "Obese III"). Such relations must be properly modeled.

In this work, we treat the problem of multi-source individual wellness attributes inference as a multi-task learning [Caruana 1997] problem. One significant issue in multi-task learning is how to define and employ a commonality among different tasks [Zhao et al. 2015]. Intuitively, different data source combinations may share common knowledge for predicting wellness attributes, and the "similar" wellness attribute categories must be correspondingly represented inside the inference model. By following this philosophy, we define a multi-task learning task as a unique combination of different sources for a given category (Figure 8), while constrain the overall model by considering the relatedness among adjacent wellness categories.

Table III: Adopted 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)

Notation: In the rest of this paper, we use uppercase boldface letters (i.e. M) to denote matrices, lowercase boldface letters (i.e. v) to denote vectors, lowercase letters (i.e. s) to denote scalars, and uppercase letters (i.e. N) to denote constants. For matrix $\mathbf{M} = (m_j^i)$, $\|\mathbf{M}\| = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{m} \left| m_j^i \right|^2}$ 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 et al. 2009] ($\mathbf{m}^i$ is the ith row of the matrix M). To simplify the reading process, we summarize all defined notations in Table III.

5.2. 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.

Suppose that there is a set of N exclusively labeled data samples, $S \geq 2$ data sources (modalities), and $G \geq 1$ categories of the inference attribute. We divide the dataset into T tasks, where *each task t is represented by the unique combination of available data sources for each attribute category*. The number of features of task t is denoted as D_t; 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 $\hat{T}$ (so that $T = \hat{T} \times G$). 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):

$$t = \{(\mathbf{x}_j^t, y_j^t) \mid j = 1...N_t,\ \mathbf{x}_j^t \in\ R^{D_t},\ y_j^t \in\ \{-1; 1\}\}^{19}.$$

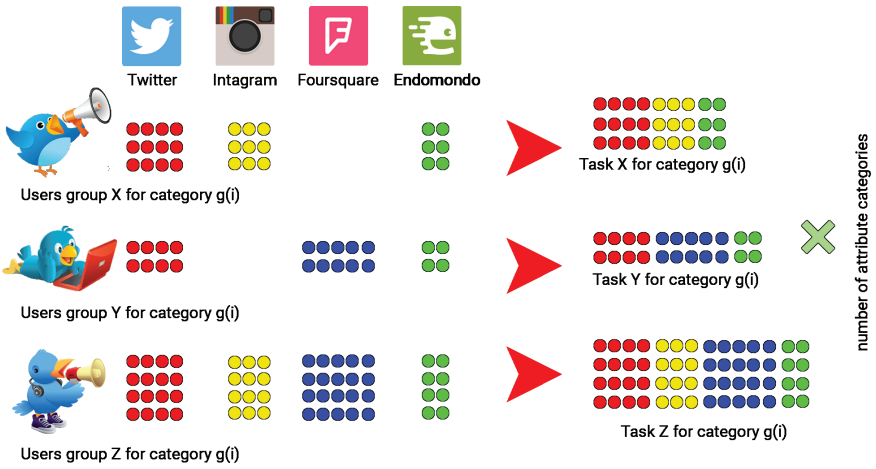


Fig. 8: Incorporating block-wise incomplete data into multi-task learning model.

The prediction for jth data sample of task t is given by the linear model:

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

where $\mathbf{w}^t \in \mathbb{R}^{D_t}$ is the model parameter vector of task t. In case of multi-attribute inference, the prediction for jth data sample of task t is given by:

$$f_{C_{D_t}}(\mathbf{x}_j^{t \in C_{D_t}}) = \arg\max_{t \in C_{D_t}} f_t(\mathbf{x}_j^t; \mathbf{w}^t),$$

where C_{D_t} is the set of D_t-dimensional[20] tasks. All model parameters are denoted as the linear mapping 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 the following optimization problem:

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

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 et al. 2009; Akbari et al. 2016], and λ 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:

$$\Psi(\mathbf{X}, \mathbf{W}, \mathbf{Y}) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{N_t} \sum_{i=1}^{N_t} \log(1 + e^{-y_i^t f_t(\mathbf{x}_i^t; \mathbf{w}^t)}). \tag{2}$$

To incorporate feature selection into the objective [Liu et al. 2009], we define $\Upsilon(\mathbf{W})$ as:

[20]Each inference attribute category g corresponds to $k_g \leq 1$ tasks of dimension D_t.

$$\Upsilon(\mathbf{W}) = \sum_{s=1}^{S} \sum_{f=1}^{F_s} \left\| \mathbf{w}_{\rho(s,f)} \right\|, \tag{3}$$

where F_s is the sth source feature vector dimension, and $\rho(s, f)$ is the index function that denotes all the model parameters of the fth feature from the sth source. The Υ term is the $\ell_{2,1}$ norm [Song et al. 2015a], which leads to a sparse solution (controlled by $\lambda \geq 0$) via constraining all models that involve source s to share a common set of features.

5.3. Modeling Inter-Category Smoothness

The above optimization problem treats all prediction attribute categories equally, while in real world scenario (i.e. detection of the user's BMI category), the attribute categories are often inter-related. For example, the adjacent BMI categories "Obese I" and "Obese II" could be similar to each other, since users from these BMI groups may follow similar lifestyles. To incorporate such relatedness, we additionally regularize the model by "Inter-Category Smoothness" term, which constrains model parameters of the adjacent attribute categories to be close to each other:

$$\Omega(\mathbf{W}) = \sum_{t=1}^{\hat{T}} \sum_{g \in C_{D_t}} \kappa_{g,\xi(t,g)} \left\| \mathbf{w}_g^t - \mathbf{w}_{\xi(t,g)}^t \right\|^2, \tag{4}$$

where $\xi(t, g)$ is the index function that picks up the model parameter ($\mathbf{w}_{g+1}^t$), which corresponds to the attribute category $g + 1$ (adjacent to g) from the combination of data sources C_{D_t}. The coefficient $\kappa_{g,\xi(t,g)}$ is the pre-defined weight of the adjacent categories' relationship. In our study, we did not incorporate weighting into category inter-relatedness chain, due to the lack of prior knowledge:

$$\kappa_{g,\xi(t,g)} = \begin{cases} 1 & g = 1...G-1; \\ 0 & g = G. \end{cases}.$$

However, in the cases when such knowledge is available [Akbari et al. 2016; Song et al. 2015b], the weighting can be naturally introduced.

The final optimization framework that integrates data from multiple social networks and wearable sensors, incorporates feature selection, and consider inter-category relationship, is defined as follows:

$$\Gamma(\mathbf{W}) = \arg\min_{\mathbf{W}} \Psi(\mathbf{X}, \mathbf{W}, \mathbf{Y}) + \lambda \Upsilon(\mathbf{W}) + \mu \Omega(\mathbf{W}), \tag{5}$$

where $\lambda \geq 0$ and $\mu \geq 0$ are the model parameters that could be obtained, for example, by performing a grid search.

5.4. Optimization

The objective function $\Gamma(\mathbf{W})$ is not smooth, since it consists of smooth terms (Ψ, Ω) and non-smooth term (Υ). 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 [Liu et al. 2009], we reformulate the non-smooth problem from Eq. (5) as:

$$f(\mathbf{W}) = \arg\min_{\mathbf{W}\in\mathbf{Z}} \Psi(\mathbf{X}, \mathbf{W}, \mathbf{Y}) + \mu\Omega(\mathbf{W})$$
$$s.t.\ \mathbf{Z} = \left\{ \mathbf{W} \mid \|\mathbf{W}\|_{2,1} \leq z \right\}, \tag{6}$$

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] Sec. 4.2).

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). The detailed description of the Nesterov's optimization approach can be found in previous works [Liu et al. 2009; Yuan et al. 2012; Akbari et al. 2016].

5.5. On time complexity:

Since $\Psi(\mathbf{X}, \mathbf{W}, \mathbf{Y})$, it costs $\mathcal{O}(D_{max}N))$ floating point operations (flops) for evaluating the function value and gradient of the objective function in Equation 6 at each iteration [Liu et al. 2009], where D_{max} and N are maximum possible dimensionality of the task and number of data samples, respectively. At the same time, it was shown [Liu et al. 2009] that the Euclidean projection, which is necessary for Nesterov's optimization, can be computed in a time of $\mathcal{O}(NT)$, where T is the total number of multi-task learning "tasks". Considering that Nesterov's approach requires $\mathcal{O}(\frac{1}{\sqrt{\epsilon}})$ iterations for achieving an accuracy of ϵ [Nesterov 2013], the overall time complexity is $\mathcal{O}(\frac{1}{\sqrt{\epsilon}}(D_{max}N + NT))$.

5.6. On data balancing:

To tackle the data imbalance problem at the training stage, we randomly and uniformly selected equal number of negative and positive samples for each binary classification task. In other words, to train task t (where tth feature dimension is D_t) we used all its corresponding positive data samples and the same number of uniformly sampled negative data samples, which were taken from other D_t-dimensional tasks.

6. EVALUATION

In this section, we present the personal wellness profiling evaluation results to answer the remaining research questions. First, we compare performance of our proposed framework for cases when it was trained based on different data sources and its

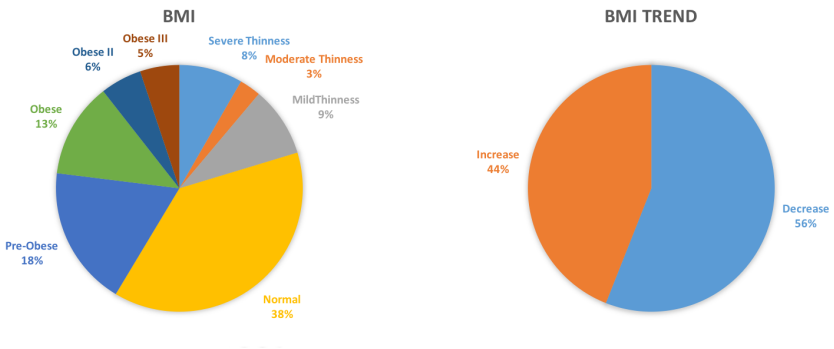


Fig. 9: Distribution of users among different BMI Categories and "BMI Trends"

combinations, which helps to answer the research question (2)[21]. Second, we perform evaluation against baselines, to respond the research question (3)[22].

6.1. Evaluation Metrics

We explicitly evaluate the performance of our proposed "WellMTL" framework[23] 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).

The distribution of users among different BMI Categories and "BMI Trends" is presented in Figure 9. It can be seen that users are quite well distributed in all BMI categories. The highest percent of users (38%) belong to "Normal" BMI category, and the smallest percent of users (3%) belong to "Moderate Thinness" BMI group. The above suggests that there is enough 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 datasets evaluation problem [Farseev et al. 2015b]. To avoid the prevalence of popular BMI categories in evaluation, we use "Macro-Recall" (R_{Mac}), "Macro-Precision" (P_{Mac}), and "Macro-F_1" ($F_{1,Mac}$) metrics, which are the averaged "Precision", "Recall", and "F_1" measures across all categories.

6.2. Evaluation Against Individual Data Sources And Source Combinations

To study the corresponding performance of individual data sources and its combinations, we evaluated "WellMTL" 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. It also should be noted that we did not evaluate the "BMI Trend" prediction performance on independent sources since there are only two users in the test set, who have contributed data in all modalities and have been labeled with the "BMI Trend" attribute.

First, we examine the contribution of different data sources towards user profile learning and source integration ability. An interesting observation comes from the data

[21] *Is it possible to improve the performance of BMI category and "BMI Trend" inference by joint modeling of social media and sensor data?*

[22] *Is it possible to improve the performance of BMI category and "BMI Trend" inference by incorporating inter-category relatedness into the learning process?*

[23] $\alpha = 0.1, \mu = 0.1, \kappa_{g,\xi(t,g)} = \begin{cases} 1 & g = 1...G-1; \\ 0 & g = G. \end{cases}$.

Table IV: Evaluation of the "WellMTL" framework trained on independent data sources and its combinations. Evaluation metrics: Macro Precision, Macro Recall, Macro F_1

Data Source Combination	BMI category prediction	
	R_{Mac}/P_{Mac}	$F_{1,Mac}$
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**

source combination results (see Table IV), where the combinations "Sensors + Text" and "Sensors + Venue Semantics & Mobility" return the best performance and seem to be the most powerful among other bi-source combinations. More impressive results can be gained from triplet combinations, where the combination "Visual + Sensors + Venue Semantics & Mobility" perform 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".

Let's now describe the single-source evaluation results (Table IV). It is interesting to note that in the case of learning from independent data sources for the BMI category inference task, 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 of all, the superiority of Text data over other modalities can be explained by its quantitative dominance (see Table I). At the same time, the Sensor data holds the 2nd position, which again highlights its importance. Finally, being trained on visual data, "WellMTL" performs the worst among all other data sources, which is consistent with current data source combination results. However, it does not conform well with our previous study [Farseev et al. 2015b], where Visual data was essential for Age and Gender inference. One possible explanation of the Visual data poor performance is the high level of noise in users' Instagram photos. At the same time, 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 attributes inference [Farseev et al. 2015b], but less effective to the more narrow problem of individual wellness profile learning. The last observation can also be explained by the results obtained in Buraya et al. [2017], where visual data was observed to play an auxiliary role in the task of relationship status prediction. In con-

Table V: Evaluation of the "WellMTL" framework trained on all data sources against baselines. Evaluation metrics: Macro Precision, Macro Recall, Macro F_1

Method	BMI category		"BMI Trend"	
	R_{Mac}/P_{Mac}	$F_{1,Mac}$	R_{Mac}/P_{Mac}	$F_{1,Mac}$
MSESHC [Farseev et al. 2015b]	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 [Yuan et al. 2012]	0.171/0.174	0.172	0.649/0.649	0.649
$aMTFL_2$ [Liu et al. 2009]	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**
"WellMTL"	0.221/0.229	**0.225**	Ω is not applicable	

clusion, we would like to highlight and suggest further usage of Text and Sensor data sources as the strongest contributors towards individual wellness profile learning.

6.3. Evaluation Against Baselines

To compare "WellMTL" framework against state-of-the-art user profiling approaches and answer the research questions (2) and (3), we evaluate it against the following baselines. We note that we did not include baselines that require prior knowledge for model guidance [Akbari et al. 2016; Song et al. 2015b] due to the unavailability of such knowledge. It also worth noting that the inter-category relationship regularization (Ω) is not applicable to "BMI Trend" classification since it is a binary personal wellness attribute.

— **RandomForest** — strong baseline for the user profile learning [Farseev et al. 2015b]. We combined features by applying an early fusion strategy and filled in the missing values by NMF. The number of trees was selected based on 10-fold cross validation and equals to 105 and 25 for the "BMI category" and "BMI Trend" inference, respectively.
— **aMTFL$_2$** [Liu et al. 2009] — the $\ell_{2,1}$ norm regularized multi-task learning with the least squares lost and $\alpha = 0.5$.
— **iMSF** [Yuan et al. 2012] — the sparse $\ell_{2,1}$ norm regularized multi-source multi-task learning, with $\alpha = 0.4$;
— **MSESHC** — weighted ensemble proposed in [Farseev et al. 2015b]. The modality weights s were learned by Stochastic Hill Climbing (SHC): 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** [Farseev and Chua 2017] — multi-source multi-task learning framework "TweetFit" (equivalent to **"WellMTL"** framework, trained without inter-category relatedness regularization (Equation 1)).
— **"WellMTL"** — our proposed framework trained on all data sources (Eq. 5), where $\alpha = 0.1$, $\mu = 0.1$, $\kappa_{g,\xi(t,g)} = \begin{cases} 1 & g = 1...G-1; \\ 0 & g = G. \end{cases}$.

The evaluation results are presented in Table V. One can notice that the results achieved by "WellMTL" are higher as compared to all the baselines (Table V), which enables us to give a **positive answer to research question (2)**. Specifically, we conclude that it is possible to improve the individual wellness profiling performance by integrating multiple data sources. Moreover, we note that the incorporation of inter-category smoothness (Ω) into model further improves the performance by 1.4%, as

compared to the model without inter-category smoothness regularization ("$TweetFit$"), which gives **positive response to the research question (3)**. The possible reason is that "WellMTL" considers inter-category relatedness for adjacent BMI categories, which constrains the model parameters of the corresponding adjacent BMI categories to be "close" to each other. Consequently, similar users are treated similarly, which allows for increasing the inference performance. At last, "WellMTL" outperforms other state-of-the-art approaches from the multi-task learning family as well as non-linear baselines, such as Random Forest and MSESHC [Farseev et al. 2015b]. This result demonstrates the framework's ability to integrate data from wearable sensors and social media for wellness profile learning.

6.4. Error Analysis

From the previous section, it can be seen that "WellMTL" framework can achieve better prediction performance as compared to different data source combinations and other user profiling baselines. However, its current performance may not be sufficient for its use in real-world scenario. To understand the reasons behind low BMI category prediction performance, we present in Table VI the evaluation results of "WellMTL" framework trained on all data sources with respect to each BMI category. We use Precision, Recall, and F_1 as evaluation metrics and color different BMI performance levels in Table VI as follows: green — good performance; blue — reasonable performance; brown — lower performance; red — low performance.

From the Table, it can be seen that two BMI categories (Normal, Pre-Obese) can be more accurately predicted as compared to the rest. The higher performance of these two categories prediction can be attributed by the large number of samples in these categories, which increases the classifiers' generalization ability. Accurate detection of Pre-Obese category is of crucial importance in public wellness applications since it helps to find those social media users with "obesity risk". This allows us to offer information support and assistance to such users, before they are deteriorated to other obesity groups.

Another interesting observation comes from the BMI categories with comparably "lower performance" (brown color) cover that all intermediate levels of either obesity or thinness. It suggests that the neighboring wellness categories inside one wellness type of overweight/underweight are not always easily separable in the social media space. This can be explained by the similar online behavior of users with similar wellness type. Such lower prediction performance also inspires us to develop new better social media-separatable obesity categorization schemes in our future works.

Last but not least, we would like to highlight the comparably good performance of Severe Thinness category and poor performance of Obese III category. Even though both categories are extreme cases, the Severe Thinness category is often common among certain categories of well-trained amateurs and sportsmen [Martinsen et al. 2010]. This implies that many users that belong to Severe Thinness category can be easier distinguished from others by the way they perform their workouts. On the contrary, Obese III may tend to have very few samples (as users may not wanted to reveal their severe obesity status to public) and thus found to behave similar to other obese cases (i.e. Obese I, Obese II) in Social Media. They are, thus, more challenging to classify.

6.5. Feature Importance Analysis

In Section 6.2, we've discussed the contribution of each data source and source combination towards BMI category inference. However, to boost the prediction performance, it is often important to know which particular feature groups are useful for predicting different BMI categories. To gain such understanding, in this section, we perform feature importance analysis by eliminating features in two steps:

Table VI: Performance of "WellMTL" framework trained on all data sources with respect to each BMI category.

BMI Category	% in Dataset	Precision	Recall	F-Measure
Severe Thinness	8%	0.333	0.25	0.286
Moderate Thinness	3%	0.167	0.2	0.182
Mild Thinness	9%	0.222	0.138	0.17
Normal	38%	0.333	0.543	0.413
Pre-Obese	18%	0.371	0.277	0.317
Obese I	13%	0.125	0.176	0.146
Obese II	6%	0.182	0.111	0.138
Obese III	5%	0.095	0.082	0.088

— **Correlation filtering:** For each feature, we computed its linear correlation with every other feature. We kept the feature with the largest number of correlated features and filtered out all the correlated features. This procedure was repeated until there are no more features with correlation higher than the threshold of $\sigma = 0.2$ was found.

— **Backward feature elimination:** Each feature elimination iteration was performed as follows: first, we trained binary "WellMTL"[24] for each BMI category on k input features; second, we removed one input feature at a time and trained "WellMTL" for each BMI category on $k-1$ input features k times; third, we removed the input feature whose removal has produced the smallest increase in the error rate. The total number of feature elimination iterations was: $D*(D+1)*G$, where $D = 230$ and $G = 8$ are respectively the data dimensionality after correlation filtering and the number of BMI categories.

Table VII describes the usage frequency of different feature types for predicting the BMI categories after the feature elimination procedure. Such frequency helps to gain an insight into feature importance for predicting each BMI category. In other words, the listed feature types are the minimal subset of features that perform the prediction of every BMI category with the lowest error rate. The most frequently used feature type is colored in red, while the feature types that are used by all the BMI inference models are colored in blue.

First, it is noted that *all the proposed feature types were utilized by the model for individual wellness profiling*. This is consistent with our source combination evaluation results and highlights the necessity of using multi-source multi-modal data for the task of individual wellness profiling.

Second, from Table VII, it can be seen that frequency domain features (Furrier Transform-based features extracted from sequential workout data) and workout statistics features are of crucial importance for predicting all BMI categories, while workout semantics features (distribution over exercise types) are weak indicators of one's wellness profile. The first observation conforms well with our initial assumptions and, once again, highlights the *importance of sequential sensor data for wellness profiling*. On the contrary, former observation suggests that the types of users' workouts are indeed not very useful for BMI prediction, which also limits the use of exercise tracking platforms that provide only the workout description but not the actual sensor measurement data sequences (i.e. MyFitnessPal, Fitocracy) [Weber and Achananuparp 2015; Jurgens et al. 2015; Park et al. 2016]. The above observation can be explained by the

[24] where $\alpha = 0.1$, $\mu = 0$, $\kappa_{g,\xi(t,g)} = 0; g = 1...G$;

Table VII: Usage frequency of different feature types for predicting BMI categories.

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

strong relation of sensor measurements (intensity, speed, heart rate) to users' actual health status, which may not be directly related to an exercise type.

Third, we would like to highlight the *importance of temporal, and venue semantics features*, which were widely used by all the prediction models. Such observation is consistent with our source combination analysis results (see Section 6.2) as well as the correlation analysis results (see Section 4). The importance of temporal and venue semantics data was also mentioned in previous studies [Noulas et al. 2012; Farseev et al. 2015b; Mejova et al. 2015] and can be explained by the major difference between people from different BMI groups in terms of location preferences and time schedules.

Lastly, we would like to bring to attention the infrequent use of text-based features, which could be due to the low representativeness of textual data itself, as compared to multi-modal data [Buraya et al. 2017]. Another possible explanation is the high level of noise in Twitter [Samborskii et al. 2016].

To summarize, we note that the results of feature importance analysis are consistent with our global data analysis results (see Section 4). For example, the previously-observed significant correlation between food groups and wellness attributes is well aligned with the utilization of lexicon-based features for predicting the 6 BMI categories. At the same time, the significant correlation of Instagram image concepts and venue semantics is consistent with its extensive usage by all the predictive models. Finally, the significant correlation of multiple frequency bins has resulted in the dominance of frequency domain features in the final models as compared to other feature types. All the above, once again, highlights the significance of utilizing all data modalities and sources for wellness profile learning and supports our answer to the research question (1).

7. ON RELATIONSHIP BETWEEN DATA SOURCES

In past works [Jain and Jalali 2014; Weber and Achananuparp 2015; Farseev et al. 2016], the research community highlighted the importance to study the relationship between different social media sources and sensors at the inter-source (cross-modal) level. To gain more insights into such a relationship, in this section, we additionally perform correlation analysis at the inter-source level. To do so, we extracted LDA topics from each data source and modality [Vorontsov et al. 2015], where the number of topics K is dictated by the lowest perplexity values and equals to 50, 9, 7, 6, for Twitter, Instagram, Endomondo, and Foursquare data sources, respectively. For all data types, we treat all user data as a document (one document corresponds to one user) and each data unit (a word in a Twitter message, Instagram image concept, Endomondo workout type, Foursquare venue category) as a "word". After obtaining the latent topics, we computed Pearson's r between all the extracted topics, user de-

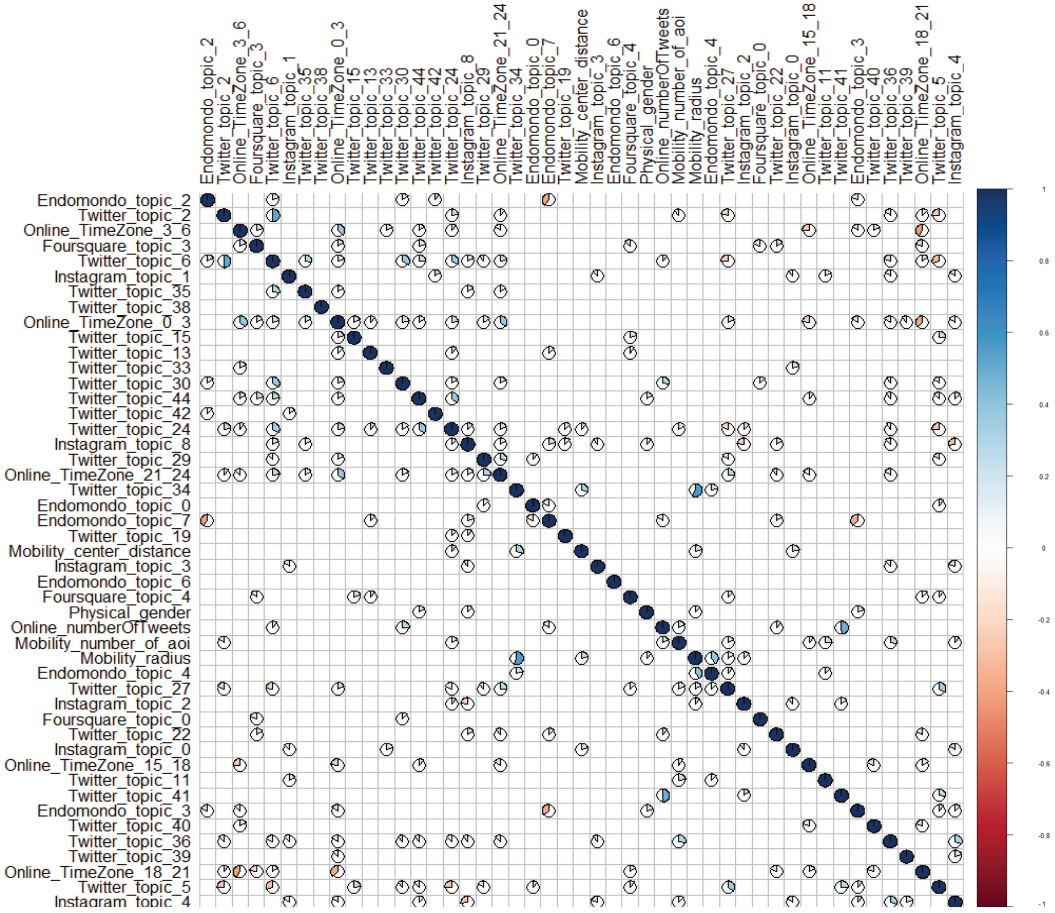


Fig. 10: Significantly ($\alpha = 0.05$) inter-source correlated data representations (matrix 1 out of 4). The negative and positive correlations are colored in red and blue, respectively.

mographic attributes, temporal and mobility features. The results are presented in Figures 10,11,12,13, where inter-source and intra-source high order relations are represented in a correlation matrix. For better readability, we name some of the extracted topics and list them in Table VIII. To achieve better visibility, we reorder the correlation matrix according to the angular order of the eigenvectors [Friendly 2002] and highlight only significantly correlated pairs ($\alpha = 0.05$) in it. Some of the strongest significant correlations are discussed below.

Inter-Source correlation: Except the abundance of intra-source correlations, due to obvious reasons, the strong correlation can also be observed at the inter-source level. For example, topic 7 from Endomondo (intensive sports categories) is positively correlated to topic 8 from Instagram (represented by beach/leisure image concepts), but negatively correlated to topic 28 from Twitter (one of the sport-related textual topics). Such a relation could be hypothesized by the difference of information that users tend to uncover in different social venues [Song et al. 2015a]: users who prefer more intensive workouts may not post their sports activities on Twitter due to the nature of their sports life, but may like to take more pictures of leisurely occasions.

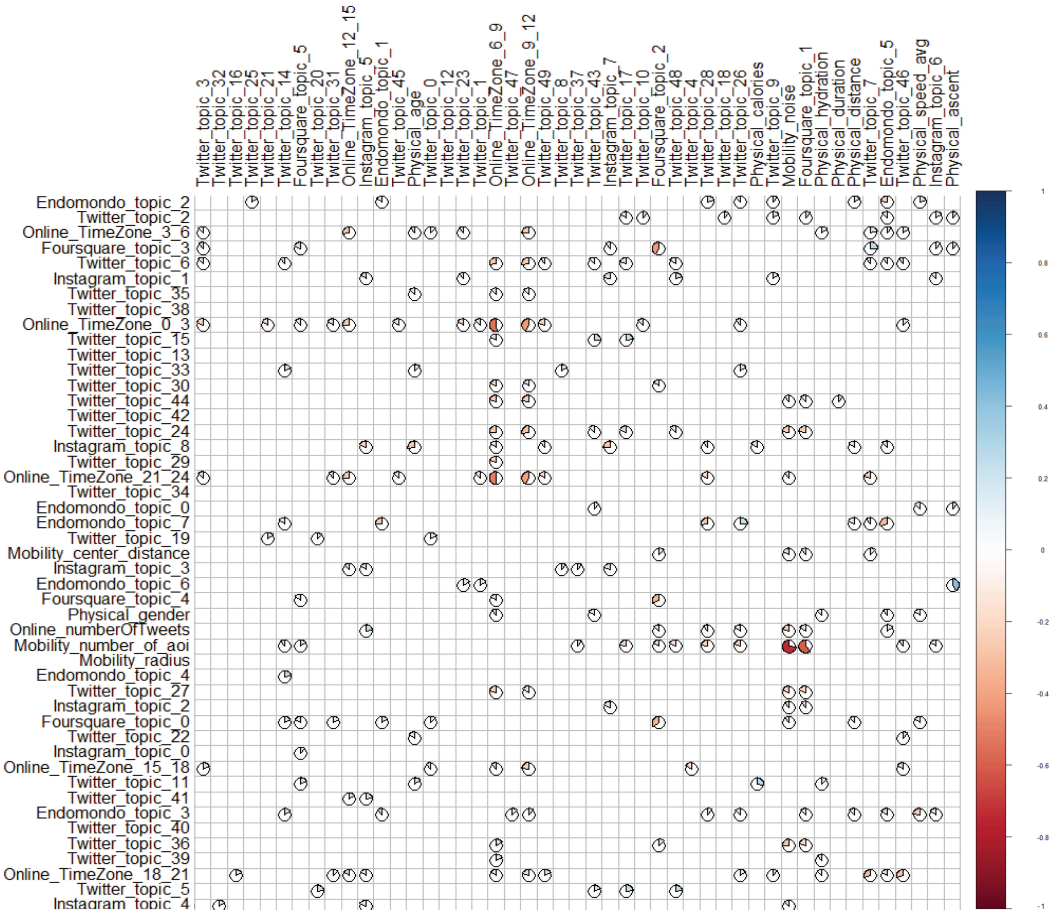


Fig. 11: Significantly ($\alpha = 0.05$) inter-source correlated data representations (matrix 2 out of 4). The negative and positive correlations are colored in red and blue, respectively.

Temporal dependencies: We observed that most of the important $3h$ time durations ($0-3$, $6-9$, $9-12$, $18-21$, $21-24$) are significantly correlated to latent topics from all data sources: topic 28 from Twitter and topic 3 from Endomondo are negatively correlated to time duration $18-21$, but positively correlated to time duration $9-12$ [Noulas et al. 2012].

Mobility dependencies: From the correlation matrix, it can be seen that many of user mobility representations are significantly correlated to other data modalities. For example, the median size of users' AOIs (the size of the area, where users are active: "Home", "Work", etc. [Qu and Zhang 2013]) is significantly positively correlated to topic 34 from Twitter and to the Endomondo's topic 4 (related to walking). The possible reason is the difference of users' lifestyle: users, who, on average, travel by foot more (i.e. during the lunch break inside Work AOI) may have larger AOI size, as compared to those who use public/private transport.

Online-to-Offline dependencies: The encouraging observations can be made from the correlation between sensor data statistics and latent topics. Specifically, the average speed is negatively correlated to topic 0 from Foursquare and topic 14 from Twit-

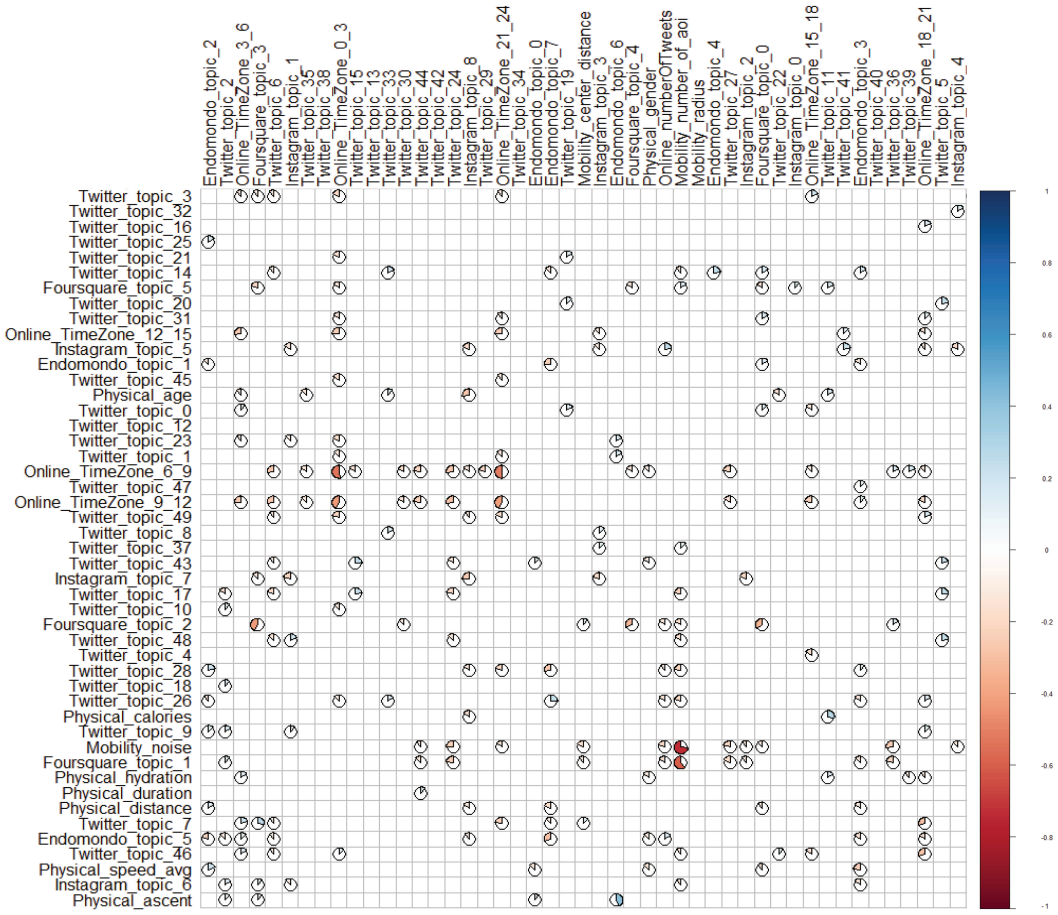


Fig. 12: Significantly ($\alpha = 0.05$) inter-source correlated data representations (matrix 3 out of 4). The negative and positive correlations are colored in red and blue, respectively.

ter, but positively correlated to topic 6 from Instagram. We note that the negatively correlated topics are both related to fitness sports activities, where topic 6 from Instagram is related to biking and cycling. The possible explanation could be that people who perform stationary fitness activities are, on average, attain lower speed during their exercises as compared to those who use bikes for workouts and transportation. It also supports the idea that different data modalities are inter-correlated and mutually complement each other [Farseev et al. 2015a; Farseev et al. 2015b; Song et al. 2015a], which allows us to discover users' activities in much wider perspective as compared to modeling each modality independently.

Aiming to bring attention to the problem of inter-source integration, we listed some interesting examples of inter-source data relationships above. These examples show that even completely different, at first sight, data modalities, such as Sensor Data and Data from Image Sharing Social Networks, are, indeed, significantly inter-correlated to each other and, thus, must be properly integrated for joint learning purposes. Such correlation may negatively affect linear models like in Equation 5 and may lead to sub-optimal performance. However, considering that we applied $\ell_{2,1}$ regularization as

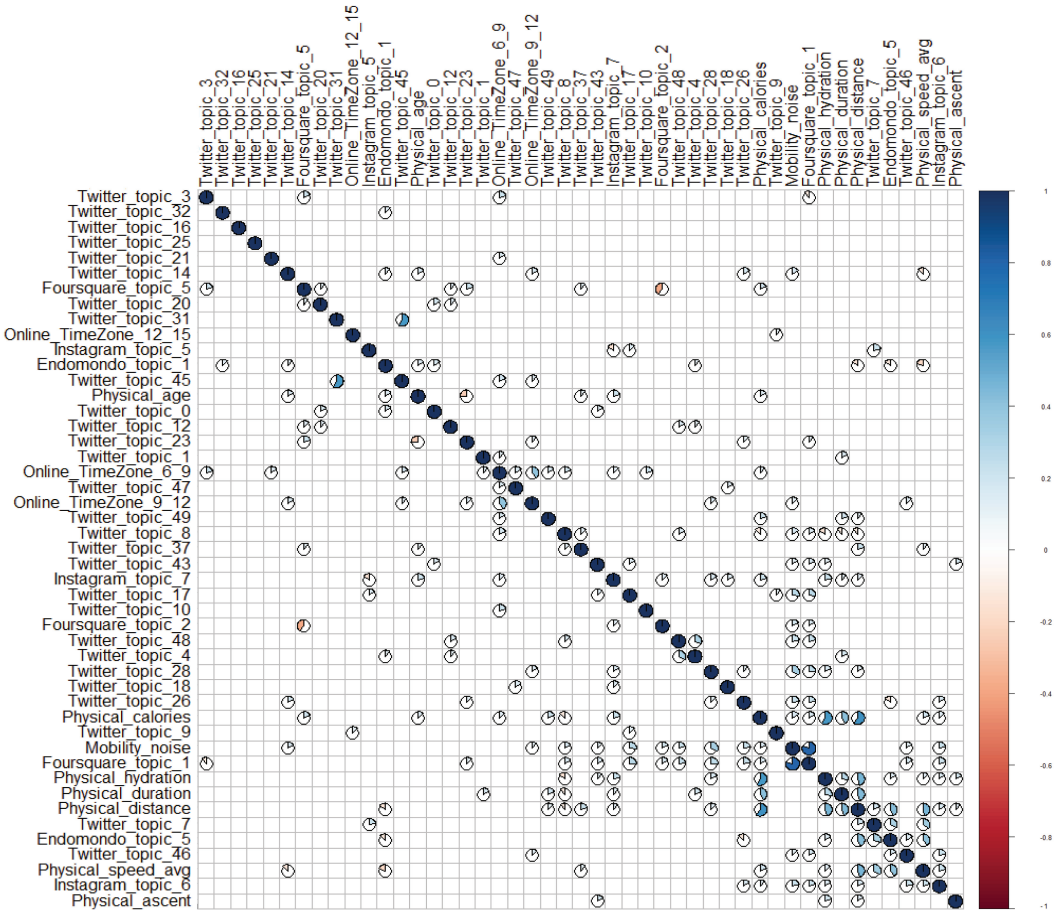


Fig. 13: Significantly ($\alpha = 0.05$) inter-source correlated data representations (matrix 4 out of 4). The negative and positive correlations are colored in red and blue, respectively.

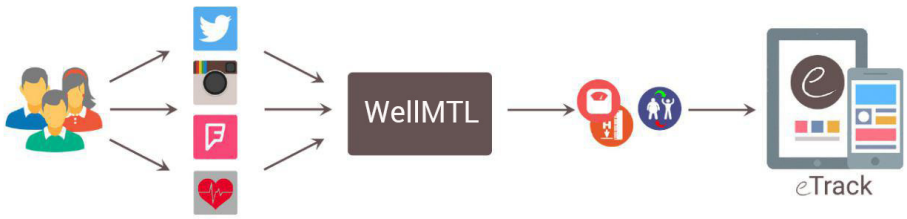


Fig. 14: "WellMTL" framework integrated into the eTrack mobile App

feature selection mechanism (see Equation 1), we do not expect the linear correlation between features to affect the final performance significantly.

Table VIII: Terms distribution among LDA topics extracted from different data sources

Topic ID	Topic Name	Topic Terms (Words)
Text latent topics		
14	Sport	mile, do, walk, fit, train, bit, step, travel, 00s, weight
28	Sport	finish, kiss, cycl, walk, 1h, bike, mountain, 2h, 3h
34	Happy	follow, pleas, list, sunshin, happiest, love, girl
Foursquare venue categories latent topics		
0	Sport	Gym, Home, Office, Train Station, Bakery, Athletics
1	Shop	Restaurant, Shop, Store, College, Place, School, Bar
Instagram image concepts latent topics		
6	Biking	mountain, bike, unicycle, tricycle, moped, helmet
8	Leisure	bathing, sunglasses, tie, wig, sunscreen, bow, maillot
Endomondo workout categories latent topics		
3	H.Int.	walking, skating, gymnastics, cricket, baseball
4	M.Int.	step counter, walking, tr. walking, am. football
7	H.Int.	running, boxing, snow shoeing

8. APPLICATIONS

To demonstrate the applicability of the "WellMTL" framework to a real-world scenario, it is integrated into the personal social media analytics App "eTrack"[25]. eTrack encourages its users to be engaged in multiple Social Media and provides lifestyle-related suggestions, which helps users to rich their healthy lifestyle goals. The structure of the App is presented in Figure 14. From the figure, it can be seen that the data from multiple social networks is aggregated by the "WellMTL" framework to predict "BMI Trend" of eTrack users. This attribute then further utilised by the App to perform user lifestyle assessment. For example, if the "BMI Trend" is predicted as "Increase" (based on the recently generated user's data), the App will suggest to reduce the calorie intake and recommend to be engaged in sports activities of the preferred type [He et al. 2016]. Additionally, users of multiple social networks are no longer required to switch between different Apps to see recent posts and events, since the aggregated multi-source timeline is automatically filtered and presented inside eTrack user interface. To the best of our knowledge, eTrack is one of the first lifestyle assessment Apps that is based on multi-source multimodal Big Data analytics.

9. LIMITATIONS AND FUTURE WORK

Although "WellMTL" outperforms the baselines, the achieved BMI prediction performance may not be sufficient for using in real-world applications. This also highlights BMI category inference as a challenging problem. The BMI prediction performance improvement could possibly be achieved by applying inter-category weights ($\kappa_{g,\xi(t,g)}$) to Ω regularization term (Equation 4) or by introducing inter-source weights into the multi-source learning objective. The inter-category and inter-source weights could be either learned during the optimization process or could be automatically inferred from the data [Akbari et al. 2016]. At the same time, the conducted error analysis suggests that it could be useful to apply a different BMI categorization schema in which similar BMI groups (i.g. Obese II and Obese III) would belong to the same inference category. Such categorization will still be useful for most applications, while able to achieve higher BMI Category prediction performance. Lastly, we would like to highlight the data sparsity problem as one of the crucial issues, which has yet to be resolved. Specif-

[25] http://etrack.bbridge.net

ically, we bring to attention the skew distribution of BMI categories and the insufficient number of positive samples for training some BMI category inference models (i.e. Moderate Thinness, Obese III, Obese II). The problem could potentially be resolved by applying data oversampling techniques, or by incorporation of external health-related knowledge into the multi-source learning process.

Except for BMI category and "BMI Trend" prediction, the inference of other individual and group wellness attributes can be approached in future works. For example, the so-called wellness-related hashtags could be used for users' wellness status prediction, wellness-related life events detection [Akbari et al. 2016], and chronic disease tendency estimation. Specifically, the #bgnow hashtag could be utilized for diabetes-related social-media content identification and individual wellness timeline construction [Akbari et al. 2016]. At the same time, such hashtags as #foodporn or #asthma could be employed in food consumption-related research [Mejova et al. 2015] and asthma inference-related research [Pongpaichet et al. 2013], respectively. In a real-world scenario, such research must be conducted based on multi-source multi-modal data [Farseev et al. 2015b; Farseev et al. 2016; Farseev and Chua 2017], which was not broadly utilized yet and thus expected to be widely adopted in future. Lastly, we would like to highlight that "NUS-SENSE" dataset naturally supports all the above research directions by providing not just personal wellness and demographics ground truth, but also the wellness-related hashtags as well as users' sports preferences and demographic attributes.

10. CONCLUSIONS

In this work, we presented one of the first studies on joint individual wellness profile learning from sensor and social media data. To uncover the relationship between different data representations and wellness attributes, we performed an extensive correlation analysis of different data representations at both inter-source and intra-source levels. Inspired by the data analysis results, we trained the "WellMTL" framework to predict BMI category and "BMI Trend" wellness attributes based on data from multiple social networks and wearable sensors. The performance of the framework was enhanced by collective data sources fusion, the introduction of sparsity into data representations, and the novel idea of a multi-source multi-task learning via efficient inter-category smoothness regularization. To demonstrate the advance of our framework over other state-of-the-art baselines as well as gain deeper insight into wellness profiling performance, we conducted an extensive quantitative evaluation, classification error analysis, and feature importance analysis. Our experimental results demonstrate the crucial importance of joint learning from multiple social media and sensor data for individual wellness profiling, which encourages further research in this direction. The successful integration of the "WellMTL" framework into personal social media analytics App "eTrack" confirms its applicability to real-world scenarios.

ACKNOWLEDGMENTS

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

The Microsoft Windows Azure Cloud and Microsoft MSDN subscription were provided by "bBridge"[26] project.

The authors gratefully acknowledge the assistance of Mr. Daron Benjamin Loo and Mr. Ivan Samborskii for assisting with preparation of the manuscript.

[26] http://bbridge.net

REFERENCES

Sofiane Abbar, Yelena Mejova, and Ingmar Weber. 2014. You tweet what you eat: Studying food consumption through twitter. (2014).

Mohammad Akbari, Xia Hu, Nie Liqiang, and Tat-Seng Chua. 2016. From tweets to wellness: Wellness event detection from twitter streams.. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. AAAI.

Hadi Banaee, Mobyen Uddin Ahmed, and Amy Loutfi. 2013. Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. *Sensors* (2013).

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* (2003).

Ron Bracewell. 1965. The Fourier Transform and it's Applications. *New York* (1965).

Kseniya Buraya, Aleksandr Farseev, Andrey Filchenkov, and Tat-Seng Chua. 2017. Towards User Personality Profiling from Multiple Social Networks. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. AAAI.

Rich Caruana. 1997. Multitask learning. *Machine learning* (1997).

H Chahal, C Fung, S Kuhle, and PJ Veugelers. 2013. Availability and night-time use of electronic entertainment and communication devices are associated with short sleep duration and obesity among Canadian children. *Pediatric obesity* (2013).

Wen-ying Sylvia Chou, Abby Prestin, and Stephen Kunath. 2014. Obesity in social media: a mixed methods analysis. *Translational behavioral medicine* 4, 3 (2014), 314–323.

Charles B Corbin, Gregory Welk, William R Corbin, and Karen Welk. 2001. *Concepts of fitness and wellness*. McGraw-Hill.

Aron Culotta. 2014. Estimating county health statistics with Twitter. In *Proceedings of the Conference on Human Factors in Computing Systems (SIGCHI)*. ACM.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.

Emma M Eggleston and Elissa R Weitzman. 2014. Innovative uses of electronic health records and social media for public health surveillance. *Current diabetes reports* (2014).

Theodoros Evgeniou and Massimiliano Pontil. 2004. Regularized multi–task learning. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 109–117.

Aleksandr Farseev, Mohammad Akbari, Ivan Samborskii, and Tat-Seng Chua. 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 (2016), 4.

Aleksandr Farseev and Tat-Seng Chua. 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.

Aleksandr Farseev, Denis Kotkov, Alexander Semenov, Jari Veijalainen, and Tat-Seng Chua. 2015a. Cross-Social Network Collaborative Recommendation. In *Proceedings of the ACM International Conference on Web Science*. ACM.

Aleksandr Farseev, Liqiang Nie, Mohammad Akbari, and Tat-Seng Chua. 2015b. Harvesting Multiple Sources for User Profile Learning: a Big Data Study. In *Proceedings of the ACM International Conference on Multimedia Retrieval*. ACM.

Aleksandr Farseev, Ivan Samborskii, and Tat-Seng Chua. 2016. bBridge: A Big Social Multimedia Analytics Platform. In *Proceedings of the International Conference on Multimedia*.

Aleksandr Farseev, Ivan Samborskii, Andrey Filchenkov, and Tat-Seng Chua. 2017. 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*. ACM.

Alison E Field, Eugenie H Coakley, Aviva Must, Jennifer L Spadano, Nan Laird, William H Dietz, Eric Rimm, and Graham A Colditz. 2001. Impact of overweight on the risk of developing common chronic diseases during a 10-year period. *Archives of internal medicine* (2001).

Andrey A Filchenkov, Artur A Azarov, and Maxim V Abramov. 2014. What is more predictable in social media: Election outcome or protest action?. In *Proceedings of the 2014 Conference on Electronic Governance and Open Society: Challenges in Eurasia*. ACM, 157–161.

Daniel Fried, Mihai Surdeanu, Stephen Kobourov, Melanie Hingle, and David Bell. 2014. Analyzing the language of food on social media. In *Proceedings of the International Conference on Big Data*. IEEE.

Michael Friendly. 2002. Corrgrams: Exploratory displays for correlation matrices. *The American Statistician* (2002).

GlobalWebIndex. 2016. GWI Social report. http://insight.globalwebindex.net/social. (2016).

Jingrui He and Rick Lawrence. 2011. A graph-based framework for multi-task multi-view learning. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. 25–32.

Xiangnan He, Hanwang Zhang, Min-Yen Kan, and Tat-Seng Chua. 2016. Fast matrix factorization for online recommendation with implicit feedback. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 549–558.

Ramesh Jain and Laleh Jalali. 2014. Objective Self. *MultiMedia, IEEE* (2014).

Laleh Jalali and Ramesh Jain. 2015. Bringing deep causality to multimedia data streams. In *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, 221–230.

Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, and Trevor Darrell. 2014. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the International Conference on Multimedia*. ACM.

Ian Jolliffe. 2002. *Principal component analysis*. Wiley Library.

David Jurgens, James McCorriston, and Derek Ruths. 2015. An Analysis of Exercising Behavior in Online Populations. In *Proceedings of the Int. Conference on Web and Social Media*. AAAI.

Oscar D Lara and Miguel A Labrador. 2013. A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys & Tutorials* (2013).

Bang Hui Lim, Dongyuan Lu, Tao Chen, and Min-Yen Kan. 2015. # mytweet via Instagram: Exploring user behaviour across multiple social networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2015 IEEE/ACM International Conference on*. IEEE, 113–120.

Jun Liu, Shuiwang Ji, and Jieping Ye. 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.

Marianne Martinsen, Solfrid Bratland-Sanda, Audun Kristian Eriksson, and Jorunn Sundgot-Borgen. 2010. Dieting to win or to be thin? A study of dieting and disordered eating among adolescent elite athletes and non-athlete controls. *British Journal of Sports Medicine* (2010).

Yelena Mejova, Hamed Haddadi, Anastasios Noulas, and Ingmar Weber. 2015. # FoodPorn: Obesity Patterns in Culinary Interactions. In *Proceedings of the 5th International Conference on Digital Health*. ACM.

Yelena Mejova, Amy X Zhang, Nicholas Diakopoulos, and Carlos Castillo. 2014. Controversy and Sentiment in Online News. *arXiv preprint arXiv:1409.8152* (2014).

Recovery Rehabilitation Center Morningside. 2013. Technology Addiction: Warning Signs of A Cell Phone Addict. www.medicaldaily.com/technology-addiction-warning-signs-cell-phone-addict-247344. (2013).

Yurii Nesterov. 2013. *Introductory lectures on convex optimization: A basic course*. Vol. 87. Springer Science & Business Media.

Anastasios Noulas, Salvatore Scellato, Neal Lathia, and Cecilia Mascolo. 2012. Mining user mobility features for next place prediction in location-based services. In *Proceedings of the 12th International Conference on Data Mining (ICDM)*. IEEE.

Pentti Paatero and Unto Tapper. 1994. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* (1994).

Kunwoo Park, Ingmar Weber, Meeyoung Cha, and Chul Lee. 2016. Persistent sharing of fitness app status on twitter. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM.

Siripen Pongpaichet, Vivek K Singh, Mingyan Gao, and Ramesh Jain. 2013. EventShop: recognizing situations in web data streams. In *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 1359–1368.

Yan Qu and Jun Zhang. 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.

Ivan Samborskii, Andrey Filchenkov, Georgiy Korneev, and Aleksandr Farseev. 2016. Person, Organization, or Personage: Towards User Account Type Prediction in Microblogs. (2016).

Jörg Sander, Martin Ester, Hans-Peter Kriegel, and Xiaowei Xu. 1998. Density-based clustering in spatial databases: The algorithm gdbscan and its applications. *Data mining and knowledge discovery* (1998).

Sanket S Sharma and Munmun De Choudhury. 2015. Measuring and Characterizing Nutritional Information of Food and Ingestion Content in Instagram. In *Proceedings of the 24th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee.

Thiago H. Silva, Pedro O. S. Vaz de Melo, Jussara M. Almeida, Mirco Musolesi, and Antonio A. F. Loureiro. 2014. You Are What You Eat (and Drink): Identifying Cultural Boundaries by Analyzing Food and Drink Habits in Foursquare. In *Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM*.

Xuemeng Song, Liqiang Nie, Luming Zhang, Mohammad Akbari, and Tat-Seng Chua. 2015a. Multiple social network learning and its application in volunteerism tendency prediction. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM.

Xuemeng Song, Liqiang Nie, Luming Zhang, Maofu Liu, and Tat-Seng Chua. 2015b. Interest inference via structure-constrained multi-source multi-task learning. In *Proceedings of the International Joint Conference on Artificial Intelligence*. ACM.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1–9.

Konstantin Vorontsov, Anna Potapenko, and Alexander Plavin. 2015. Additive regularization of topic models for topic selection and sparse factorization. In *Statistical Learning and Data Sciences*. Springer.

Xiangyu Wang, Yi-Liang Zhao, Liqiang Nie, Yue Gao, Weizhi Nie, Zheng-Jun Zha, and Tat-Seng Chua. 2015. Semantic-based location recommendation with multimodal venue semantics. *IEEE Transactions on Multimedia* (2015).

Ingmar Weber and Palakorn Achananuparp. 2015. Insights from machine-learned diet success prediction. *arXiv preprint arXiv:1510.04802* (2015).

World Health Organization WHO. 2011. Global database on body mass index. 2011. *Global Database on Body Mass Index* (2011).

Lei Yuan, Yalin Wang, Paul M Thompson, Vaibhav A Narayan, and Jieping Ye. 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)*.

Reza Zafarani and Huan Liu. 2013. Connecting users across social media sites: a behavioral-modeling approach. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 41–49.

Liang Zhao, Qian Sun, Jieping Ye, Feng Chen, Chang-Tien Lu, and Naren Ramakrishnan. 2015. Multi-task learning for spatio-temporal event forecasting. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1503–1512.