Physical-Cyber-Social Similarity Analysis in Smart Cities

Nazli Farajidavar
Institute for Communication Systems
University of Surrey, UK GU2 7XH
Email: n.davar@surrey.ac.uk

Şefki Kolozali
Institute for Communication Systems
University of Surrey, UK GU2 7XH
Email: s.kolozali@surrey.ac.uk

Payam Barnaghi
Institute for Communication Systems
University of Surrey, UK GU2 7XH
Email: p.barnaghi@surrey.ac.uk

***Abstract*—The inter-departmental interactions and coordination of resources are two essential components for realising a smart city platform. In this study, we investigated citizens' role in enhancing and facilitating the delivery of services by merging three key aspects of the smart city research field, namely Internet of People, Internet of Things and Web of Data. To this end, we developed a hybrid approach to extract meaningful information and to find physical-cyber-social similarity in smart cities. The three specific data sources used in this study were Twitter, road traffic disruptions collected from Transport for London API, and events parsed from Time Out London. With the proposed hybrid approach, we found that 49.5% of the Twitter traffic comments are reported approximately five hours prior to the authority's official records. Moreover, we discovered that amongst the pre-scheduled sociocultural events topics; transportation, cultural and social event topics are 31.75% more likely to influence the distribution of the Twitter comments than sport, weather and crime topics.**

I. Introduction

Smart City frameworks endeavour to provide authorities and citizens with real-time information and assistance in the decision-making and resource allocation processes in their daily lives. On this purpose, online monitoring and cognition of city events are not only critical for municipal authorities to leverage the management of resources and reduce the costs, but also for citizens to promote their life quality.

With the recent advancements in information communication technologies, it is now possible to collect (near) real-time observations and detect critical events in cities using social media, sensors, and the Web of Data. Social media networks such as Twitter offer a platform for citizens to share their opinion in near real time, which can serve as an informal service quality assessment based on population probes. Previous studies [1] demonstrated citizens' social media contribution in the detection of phenomena in cities. Such textual information can further be enriched with textual data sources on the Web and at city departments.

Internet of Things is another key aspect that enables smart systems to monitor physical or environmental conditions [2], [3]. It allows us to deploy sensor nodes in different locations to detect the critical events. For example, when road sensors report the decreased vehicle speed on the road for a long time, we can induce that there is a traffic congestion on a particular road.

In this study, we have developed a hybrid system for real-time sensing in cities through the utilisation of complementary dynamic data sources, namely Twitter, London Road disruption reports from traffic sensors; and Time Out London and Wikipedia to obtain the prescheduled events and words associations, respectively. The proposed data processing chain involves data wrappers, Natural Language Processing (NLP) component, and correlation analysis. In our experiments, we used three different data wrappers: Twitter stream API boosted with Google Translate API for the collection of tweets and their translation, Transport for London API for the collection of road sensor data, and a parser to collect scheduled events from Time Out London website. For the NLP unit, we used three sub-components: a Conditional Random Field (CRF) for Name Entity Recognition (NER), a Convolutional Neural Network (CNN) Part of Speech (PoS) tagging, and a multi-view event extraction which combines the information extracted from the previous sub-components to obtain a sentence-level inference for event classification. Subsequently, we conducted a similarity analysis on the data collected from social media, road sensors, and Web of Data, and discover the associations between incidents in near real-time.

The research contributions are four-fold and can be summarised as follows: i) Automated real-time data collection wrappers for Twitter and city sensors; ii) A near real-time NLP component for classifying Twitter data; iii) A correlation analysis for detecting the dependencies between Twitter stream and city sensors and web driven data records; iv) A web interface for displaying and visualising the city's event highlights.

The rest of the paper is organised as follows. In section II, we give an overview of the Natural Language Processing techniques used for tweet analysis. In section III, we describe the methodology we utilised in this study. In Section IV, we present our experimental results. Finally, in the section V, we present our conclusion and note on further difficulties of the research problem and outline our future work.

II. Related Work

There have been several work on the event prediction from textual data. In [4], the authors conducted their experiments based on the assumption that various event data sources are readily available in cities, such as sensor data (*e.g.*, loop

detectors) and formal report of events (*e.g*., eventful[1]). On one hand, the utilisation of such a formal data source can serve as a ground truth to validate an automated event extraction system. On the other hand, such data source may not be available in all cities. Therefore, we need the alternative and complementary data sources for event extraction in cities.

Meanwhile, Social Networks, such as online news articles, Twitter and Facebook, offer large data that can be exploited to serve as an alternate to formal reports. Textual data sources can be categorised in two different structural types: (i) formal and (ii) informal. While the former refers to the grammatical texts such as news documents, the latter addresses the user-generated content with no overt structure that might contain a lot of slang and non-standard abbreviations and notations such as Twitter.

In the context of formal text analysis, Liu *et al*. [5] proposed to alleviate information overload in daily news by extracting key entities and significant events from news documents. They utilised a bipartite graph that is induced based on the entities and their associations to documents using mutual reinforcement principle, capturing salient entities and associated documents, to rank the news events. Meanwhile Okamoto et. al. [6] extracted local events from blog entries, Tanev et.al. [7] used lightweight patterns to extract events from textual news for the global health crisis and Grishman [8] extracted events using finite-state pattern matching on the tokenized input text to detect infectious disease outbreak. The event schema that is used in this study consisted of date range, geo-location, disease name, organism type and the number of people affected by the disease, and the organism survival information.

Recently, there have been a few studies on event extraction from informal text [9], [10], [1]. In [9], the authors proposed to use temporal (*i.e*. volume changes), social (*i.e*. replies, broadcast), topical (*i.e*. coherence of clusters), and Twitter-centric (*i.e*. multi-word hashtags) features to train a classifier which performed better than the baseline. Ritter *et al*. [10] demonstrated that building a calendar of significant events is feasible using an unsupervised approach to process tweets and extract event types such as sports, concerts, protests, politics, TV, and religion. The approach utilised Latent Dirichlet Allocations (LDA) method to model each entity with a mixture of event types and each event type with a mixture of entities. Following the same trend, Wang *et al*. [11] developed a model to predict hit-and-run crimes. They used a latent topic model for events [12] and the generalised linear regression model is then applied to capture the association between event topics and crimes through training on an auxiliary dataset. Lampos and Cristianini [13] applied an optimised feature selection approach coupled with regression to estimate the intensity of events based on event markers. They evaluated the model based on the ground truth collected from rain gauges dataset. In another study, Lampos and Cristianini extended their study to discover regional flu rates from the Twitter and compared the outcome with the official data from Health Protection Agency [2]. They found a strong correlation between the HPA reports and Twitter patterns.

Similar to the assumption in [4], Anantharam *et al*. [1] recently developed an automatic data annotation unit to obtain ground truth by using officially reported traffic events [3] and location [4] knowledge-bases. The authors then used this annotated data to train a CRF-based event extraction model to capture long-term word dependencies for Twitter analysis. While their proposed approach for the preparation of the ground-truth data has shown a good word-tagging performance, the proposed CRF-based event extraction had some limitations: it was only for traffic event extraction. Because the automatic annotation unit was trained with the officially reported ground-truth traffic events of a limited period, the model performed poorly in the prediction of future incidents. Besides, although the location terms have been extracted, they were not utilised to associate locations with extracted events. Instead, the authors assumed that the tweet's geo-location tag (the location where the users tweet the events) is same as the event locations, which does not necessarily agree with the actual event's location.

These limitations motivated our work. Inspired by the ground-truth annotation approach of [1], we propose a hybrid approach that involves social media, traffic sensor data as well as Web of Data to detect the patterns in the cities and to evaluate the detection performance. Our generic approach comprises of an event-class dependent dictionary of terms and phrases, which then used for training a CRF model to recognise class specific name entities. The model further boosted with a CNN feature extraction [14] for extracting part of speech tags and the two views are later combined through a multi-view learning framework. Unlike Anantharam [1]*et al*.'s model, our proposed model can be easily generalised to future events and used for other cities.

III. Methodology

Our hybrid approach composed of two main tasks: i) the application of Natural Language Processing (NLP) on Twitter data streams and ii) similarity analysis on Twitter, road sensor data, and prescheduled events collected from websites. Figure 1 depicts the proposed hybrid approach. We developed three data wrappers namely, Twitter stream API, Transport for London API, and Time Out London website. In parallel, we used the Google translate API to automatically detect the source language and translate the tweets in English to facilitate the text analysis.

Figure 2 shows the data processing units in the NLP component which is composed of three sub-components: a semantic embedding subspace learning, a syntactic embedding subspace learning, and a multi-view event extraction. Given a tweet, $\mathbf{x}$, the NLP component annotates it based on the following event class set: $C = \{$TransportationEvent,

[1]http://eventful.com/

[2]http://www.hpa.org.uk/

[3]http://511.org

[4]https://www.openstreetmap.org/

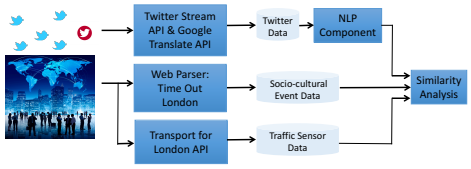

Figure 1. The proposed hybrid Physical-Cyber-Social pipeline

EnvironmentalEvent, CulturalEvent, SocialEvent, SportEvent, FoodEvent, CriminalEvent}.

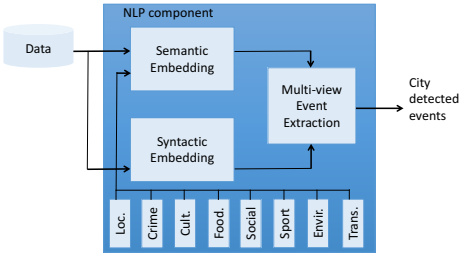

Figure 2. NLP component: detailed event detection pipeline

A. *Twitter NLP Component*

To assign event tags to tweets, we have assumed that each tweet contains only one sentence. Considering the 140 character limit of a tweet, this assumption sounded plausible. We then decomposed sentences into semantic and syntactic embeddings. The former deals with the meaning of the words in the sentence and the latter address its grammar structure. We denote these two embeddings with ϕ and θ, respectively. The combination of both of the embeddings provides an explicit insight into the full meaning of the sentece (tweet). Subsequently, we formulated the event extraction as a multi-view learning, where each embedding contributes to a distinct view of the training data. We used Restricted Boltzmann Machine (RBM) in our multi-view learning framework. Doing this, we obtained a learning mechanism based on a compatibility weight between the semantic and syntactic embeddings. An illustration of the proposed model is shown in Fig. 3.

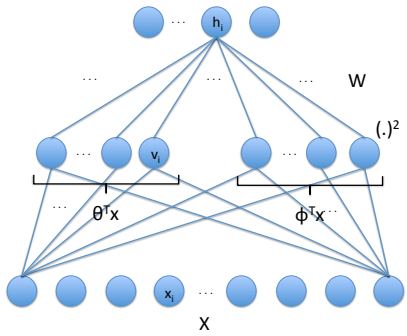

Figure 3. Modeling the multi-view learning through an RBM energy model applied to the concatenation of the semantic and syntactic embeddings of a sentence.

The RBM is a Markov Random Field associated with a bipartite undirected graph. In the Bernoulli RBM, our focus in this work, the visible and hidden variables are assumed to take values (v, h) in $[0, 1]$. Each value encodes the probability that the specific feature would be active. Perceiving the RBM as an energy model, the activity of a hidden unit can be written as:

$$\mathbf{h} = \mathbf{W^T}(\mathbf{F^T v}) \odot (\mathbf{F^T v}) \quad (1)$$

where $\odot$ is element-wise product and the columns of $\mathbf{F}$ contain subspace projection filters that are learned along with $\mathbf{W}$ from data [15].

For our analysis, we have modelled the visible layer of the RBM as the concatenation of two embeddings $\mathbf{F^T v} = \theta^\mathbf{T}\mathbf{x} + \phi^\mathbf{T}\mathbf{x}$ where θ denotes the part of the filter $\mathbf{F}$ in Eq.(1) that is applied to input sentence $\mathbf{x}$ to extract its syntactic embedding and ϕ denote the part of the filter $\mathbf{F}$ in Eq.(1) that extracts the semantic embedding of the sentence that can be perceived as a projection of the label space $\mathcal{Y}$.

Substituting the view concatenations in Eq.(1), hidden unit activities take the form:

$$\mathbf{h} = \mathbf{W^T}(\theta^\mathbf{T}\mathbf{x} + \phi^\mathbf{T}\mathbf{x})^\mathbf{2} = \mathbf{2W^T}(\theta^\mathbf{T}\mathbf{x})(\phi^\mathbf{T}\mathbf{x}) + \mathbf{W^T}(\theta^\mathbf{T}\mathbf{x})^\mathbf{2} + \mathbf{W^T}(\phi^\mathbf{T}\mathbf{x})^\mathbf{2} \quad (2)$$

To estimate the learning parameters, weight matrices θ, ϕ and $\mathbf{W}$, one can train a multilayer neural network on a large number of training instances. However, since the annotation of a large corpus of tweets can be time-consuming, and the trained model can be time and space dependent, we have alternatively estimated the embedding matrices θ, ϕ offline and independently using more comprehensive data. As such, we applied a PoS tagging, which has been trained on encyclopaedia corpus to extract a reasonable syntactic embedding $(\theta^T\mathbf{x})$ for tweets. We have adopted the CNN model proposed in [14], which had been trained on the entire English Wikipedia [5]. To formulate the semantic embedding (ϕ) extraction, we used a CRF formulation for the extraction of name entities for conditional events. Table I lists the general phrases, short reports and location terms that are extracted from official websites (listed in Tab. I) to built class conditional corpora. The conditional corpora are then used to train CRF [6] models for each event class. The proposed approach will, in fact, deal with the challenge of open domain city nature of events and city-dependent location terms and references which will vary from one city to another.

Having the subspace projection matrices θ and ϕ, we now just need to learn the weight matrix $\mathbf{W}$ from in-domain data (tweet instances). Given the fact that the activity of the second layer in the proposed architecture (see Fig. 3) will be the concatenation of the projected views, the last layer can be trained by simply maximising the log-likelihoods over the training set. Given a sentence $\mathbf{x}$, the network with energy function $E(v, h) = \sum_i \sum_j w_{ij} v_i h_j + \sum_i b_i v_i + \sum_j k_j h_j$ and parameter set $\Theta = (\mathbf{W}, \mathbf{b}, \mathbf{k})$ computes a score $s^c_\Theta(\mathbf{x})$ for

[5] Available at: http://download.wikimedia.org

[6] we have used the Alias-i 2008 LingPipe 4.1.0. implementation of CRF available at http://alias-i.com/ lingpipe.

Event Class	Phrase/Report Source
Crime	http://www.shouselaw.com/crimes-a-z.html
Cultural event	http://en.wikipedia.org/wiki/Category:Cultural_events
Food	http://www.foodterms.com/encyclopedia
Location	https://www.openstreetmap.org
Social event	http://en.wikipedia.org/wiki/Category:Social_events
Sport	sport corpus of [10]
Weather	http://www.erh.noaa.gov/er/box/glossary.html
Transportation	http://511.org

Table I. City event classes and their corresponding corpus resources

each event label $c \in C$. In order to transform these scores into a conditional probability distribution of labels given the sentence and the set of network parameters, one can apply a softmax operation on the scores:

$$p(c|\mathbf{x}, \Theta) = \frac{e^{s^c_\Theta(\mathbf{x})}}{\sum_{\forall i \in C} s^i_\Theta(\mathbf{x})} \tag{3}$$

Taking the log from two sides of the Eq.(3):

$$\log p(c|\mathbf{x}, \Theta) = s^c_\Theta(\mathbf{x}) - \log(\sum_{\forall i \in C} s^i_\Theta(\mathbf{x})) \tag{4}$$

We then used the gradient decent theorem to minimise the negative log probability with respect to Θ. The back-propagation algorithm is a natural choice to efficiently compute gradients of the network architecture as stated in [14].

An example presenting the performance of such a multi-view inference is in the case of tweets such as *"seeing someone being given a parking ticket"* where individual words *"parking"* and *"ticket"* can belong to Transportation and Cultural event categories respectively. However, considering the word grammar roles can help in resolving this confusion and assigning the tweet to Transportation category.

B. Similarity Analysis Graph Representation

For the similarity analysis, we narrowed down our focus to the event classes which we had access to in the authority event records namely, the Transport for London API [7] and Time Out London website [8]. To do this, we collected officially registered traffic reports from London open data store portal and parsed the Time Out London webpage to get a list of pre-scheduled sociocultural events taking place in London along with their time stamp and locations.

Two graph structure have been considered to represent the spatial distribution of the traffic and sociocultural records. Let assume that, $G^T = \{n^T_1, n^T_2, ..., n^T_{|T|}\}$ and $G^{SC} = \{n^{SC}_1, n^{SC}_2, ..., n^{SC}_{|SC|}\}$ representing the traffic and sociocultural record graphs respectively where $|.|$ denotes the cardinality of each record set. The edge values in these graphs are associated with the pair-wise Euclidean distances between the nodes which are in 3D coordinates and are denoted with feature vectors $\mathbf{n_i} = (\mathbf{x_i}, \mathbf{y_i}, \mathbf{z_i}, \mathbf{t_i}, \mathbf{e_i})$, The first three variables, x, y, z, represent the spatial coordinate of a point in Cartesian space and t_i, e_i are the event time stamp and event type, respectively. We employed these two graphs for detecting the nearest node to each of the automatically annotated Twitter events. However, to perform this similarity measure, we have considered the spatio-spectral topography of London city.

As one can note, the spectral structure of the city infers a spectral weighting of computed distances. Meaning that, the farther we move away from the city centre, the distance turns to be inversely prominent. Taking this into account, given that seven target graphs each representing the twitter events of each class and simplifying the location vector of each node with $P_i = (x_i, y_i, z_i)$, we then formulated the graph dissimilarities with respect to authority graphs as follows:

$$\hat{D}_c = \Sigma_{i=1:|c|} \min(d(P^{G^c}_i, P^{G^T}_j)/\lambda_j) \tag{5}$$

where $d(.)$ represents the Euclidean distance between two points. The points superscripts, G^c, G^T, shows the graph memberships, and $|c|$ denotes the cardinality of tweets, which are classified as event type c and with P_{centre} being the Cartesian conversion of the city centre geo-coordinates [9]. The parameter λ is formulated as $\lambda_j = d(P^T_j, P_{centre})$. In practice, the value of the parameter imposes higher weight to close-to-centre events compared to off centre events.

IV. Experiments and Results

Our evaluation objectives are three-fold: i) we first evaluated the performance of the proposed CRF NER model in extracting city events from Twitter. We compared our approach with state-of-the-art baseline approach presented in [1], [10] on $SanFrancisco$ data; ii) we then investigated the performance boost of the proposed MV-RBM approach for sentence inference rather than just word tagging by testing the model on a locally collected dataset from London; iii) we finally performed a similarity analysis to see how well the Twitter extracted events are matching with authority reports.

A. Dataset

We conducted our experiments on textual Twitter data collected from two geographically different locations: San Francisco Bay [10] area (data collected by [1]) and our locally collected London datasets, $London_1$ and $London_2$.

The $London_1$ is composed of 3000 tweets collected between $15^{th} - 31^{th}$ May 2015 and manually cleaned and annotated for training and testing the MV-RBM model. The manual annotation results undergo a second investigation for ensuring their consistency and validity. We have asked a group of technical users, who work in the field of smart cities to peer-review the validation of the annotations.

The $London_2$ dataset is collected on 3^{ed} of February 2016 and is of size $1.1MB$. In IV-C, we used this dataset to examine the Twitter extracted event similarity with the road sensor data and the prescheduled events that are parsed from the Web.

[7] http://data.london.gov.uk/

[8] www.timeout.com/london/thing-to-do

[9] Following flicker, (-0.127, 51.507) is considered as the (longitude, latitude) pair describing the city centre

[10] Please see [1] for a detailed description of this dataset

Traffic and sociocultural reports corpora which were used for constructing the authority graphs of section III-B, are also locally collected on the same date for synchronism.

B. Performance Evaluation I

Table II shows a comparison between our proposed CRF-based NER using universal dictionaries and the two state-of-the-art approaches: M_1 [10] and M_2 [1]. We have used the $SanFrancisco$ dataset for this comparison as the baseline approaches had been evaluated using this dataset.

We found that our CRF-NER model performed equally well as the best performing baseline model despite using generic city-independent dictionaries, recalling for 95% vs 93% precision on Location terms detection and 84% vs 86% precision on event detections. This means unlike M_2 [1] approach, we did not provide domain (city) specific prior knowledge for training our CRF model. Instead, we used the CRF-NER tagging approach that is more flexible due to benefiting from a more generic set of conditional class terms. This enabled us to remove any geographical bias. However, our model performed as well as the model, which is specifically trained for a controlled domain (San Francisco Bay area) with access to official traffic reports. Overall, we developed a model that is more flexible and adaptable, while producing results that were as good as the baseline approach. Therefore, our approach can be used for other cities with potentially varying event distribution.

C. Performance Evaluation II

Table III shows the evaluation results on $London_1$ and $London_2$ datasets. For the $London_1$ dataset, the performance is reported regarding one versus all class accuracies. The multi-view approach shows 5% improvement in the performance compared to the single view model that classified tweets according to the CRF NER word tagging. We obtained a lower performance on *"Social", "Cultural"* and *"Food"* topics compared to other topics. This can be explained by two reasons: class conditional dictionary term similarities in *"Social"* and *"Cultural"* categories and incomplete class conditional dictionaries in *"Food"* category. However, there were some terms and phrases which contributed in more than one dictionary. This has made the event extraction process more challenging in such scenarios.

The effect of the dictionary problem could be reduced by increasing the training data from the event classes that provide less accurate results. The multi-view training step can in fact provide more flexibility in expanding class-specific dictionaries. However, adding more terms will require the CRF model to be trained with a larger size training data and will cause higher time and computational complexity.

D. Similarity Analysis

We measured the similarity of the extracted Twitter events against the road sensor data and scheduled sociocultural events that are parsed from the Web. The $London_2$ dataset is used for this experiment.

The comparison results are reported in terms of classes average distance (represented with μ) from their nearest authority/web record along with its variance (represented by σ) and a similarity measure. To compute the similarities from the average distance values, we have first rescaled the averaged distances by the within-class maximum average distance (sport class distance) and then subtracted the result from one. Doing so, the similarity values are confined to be between 0 and 1 where closer to one values will guarantee higher similarity and values close to zero show a higher level of dissimilarity. Third row results in Table III show different Twitter event distributions compared against the ground-truth traffic sensor data, where latitude longitude pairs are converted to Cartesian values and Euclidean distance used as the metric. The results showed that the smallest average distance of 0.74, considering the variance intervals, correspond to the Traffic events, which is a proof of an acceptable tweet classification performance. Additionally a comparison of Twitter traffic report times with their nearest neighbour authority traffic record time-stamp, shows that 49.5% of the Twitter traffic alerts are reported on average 297.5 minutes. This is approximately 5 hours in prior to the authority's official reports. This finding highlights the advantage of utilising the social media, particularly Twitter driven knowledge in facilitating and speeding up the city management and potentially smoother task-handling.

Fourth row results in Table III show different Twitter class distributions compared with the ground-truth sociocultural data, which was parsed form Time Out London [11] computed in the same way to traffic similarities. In our experiments, we discovered higher similarity measurements for traffic, cultural and social tweet event classes with 1.73, 2.95 and 2.00 respective average distance values. This, in fact, proves that the popular and well-advertised cultural activities are potential high-traffic zones. It is important to point out that the Time Out London sociocultural event set does not categorise social events, such as special events held in pubs and restaurants, from cultural events (*i.e.* exhibitions and ceremonies) while our proposed pipeline labels these events differently.

E. Web Interface

To facilitate the visualisation of the extracted city events, we developed a Web interface. It displays the results on a Google map in near real-time. The interface is composed of; i) Google map canvas layer on which the processed and annotated Tweets are displayed with their class-identical icons; ii) a live London traffic layer from Google traffic API - code coloured paths on the map; iii) a bar chart panel which presents the class distribution histogram of daily Tweets; and iv) a panel for displaying Twitter timeline. The map data is being updated every 60 seconds by appending the past minute's Tweets to existing ones up to a 60-minutes time window. In practice, the whole data will be updated on hourly basis. Clicking on each event a dialogue box is shown on the map which reveals

[11] http://www.timeout.com/london/things-to-do/things-to-do-in-london-today

	Other			Location			Events		
	M_1	M_2	CRF NER	M_1	M_2	CRF NER	M_1	M_2	CRF NER
Recall	0.93	0.98	0.98	0.75	0.85	0.92	0.24	0.88	0.85
Precision	0.93	0.98	0.97	0.82	0.93	0.95	0.11	0.86	0.84

Table II. Comparisons of our CRF NER tagging vs. the baseline M_1 and M_2 methods on San Francisco dataset. Note that our approach does not take any domain (city) prior knowledge into account except for location terms.

	Dataset		Crime	Cultural	Food	Social	Sport	Weather	Trans.
CRF tagging	$London_1$		0.53	0.36	0.35	0.16	0.52	0.80	0.67
MV-RBM	$London_1$		**0.60**	**0.37**	**0.48**	**0.21**	**0.54**	0.80	**0.69**
Road Rep.	$London_2$	$\mu_D \pm \sigma$	1.00 $\pm$ 8.79	1.00 $\pm$ 3.67	0.95 $\pm$ 5.14	0.90 $\pm$ 3.74	1.69 $\pm$ 8.72	1.72 $\pm$ 8.21	**0.74 $\pm$ 2.34**
		Similarity	0.44	0.44	0.47	0.50	0.00	0.04	**0.59**
TimeOut Rep.	$London_2$	$\mu_D \pm \sigma$	3.0 $\pm$ 31.27	**2.95 $\pm$ 7.95**	3.60 $\pm$ 45.45	**2.00 $\pm$ 7.49**	5.26 $\pm$ 31.17	3.77 $\pm$ 17.46	**1.73 $\pm$ 5.03**
		Similarity	0.41	**0.44**	0.32	**0.60**	0.00	0.28	**0.67**

Table III. Multi-view learning evaluation and similarity analysis: Numerical results of multi-view learning evaluation compared against CRF event classification. Third row: different Twitter class distributions compared against the ground-truth traffic sensor driven data distribution, Fourth row: different Twitter class distributions compared against the ground-truth sociocultural data parsed form TimeOut London.

the underlying Tweet content along with its time-stamp. The twitter user id and the names are anonymised for privacy purpose. The web interface [12] utilises javascript and HTML coding to read the data results saved in a CSV data structure format and displays the tweets on the map in near real time with less than 60 seconds latency.

V. Conclusions

In this study, we have investigated the citizen's rule in enhancing and facilitating the cities' service delivery. We utilised the Twitter streams as a platform demonstrating the Internet of People for extracting and analysing city events. We proposed a multi-view deep learning model with an averaged 5% performance boost for event detection and developed a live stream analysis interface to present the analysis results. We have also performed similarity analysis which showed how authority driven traffic reports and pre-scheduled sociocultural events can affect the traffic pattern and how citizens project on them through social media. The evaluations showed that 49.5% of the Twitter traffic comments are reported approximately five hours prior to authority's official records and the pre-scheduled sociocultural events have observed influencing the distribution of the twitter comments on traffic, cultural and social classes. The study highlights the possibility of utilising social media as human probes for realising a real-time Physical-Cyber-Social platform for ranking, completing and potentially speeding up the city service deliveries. The proposed model serves as a proof of concept and improvements can be made in several stages of the pipeline. For example, the multi-view learning component can be updated with an online learning model which is more adaptive and provides performance feedback to the rest of the pipeline.

Acknowledgment

This work has been carried out in the scope of the European Commissions Seventh Framework Programme funded project CityPulse (FP7-609035).

References

[1] P. Anantharam, P. Barnaghi, K. Thirunarayan, and A. P. Sheth, "Extracting city traffic events from social streams," in *ACM Transactions on Intelligent Systems and Technology*, vol. 6, no. 4, 2015.

[2] D. Puiu and etal, "Citypulse: Large scale data analytics framework for smart cities," *IEEE Access*, Dec 2015.

[3] S. Kolozali, D. Puschmann, M. Bermudez-Edo, and P. Barnaghi, "On the effect of adaptive and non-adaptive analysis of time-series sensory data," *IEEE Internet of Things*, 2016.

[4] D. Mladenić and A. Moraru, "Complex event processing and data mining for smart cities," in *Proc. of the 15th International Conference , Multiconference Information Society*, 2012.

[5] M. Liu, Y. Liu, L. Xiang, X. Chen, and Q. Yang, "Extracting key entities and significant events from online daily news," in *Intelligent Data Engineering and Automated Learning - IDEAL 2008, 9th International Conference, South Korea, 2008*, 2008, pp. 201–209.

[6] M. Okamoto and M. Kikuchi, "Discovering volatile events in your neighborhood: Local-area topic extraction from blog entries." in *AIRS*, ser. Lecture Notes in Computer Science, vol. 5839. Springer, 2009, pp. 181–192.

[7] H. Tanev, J. Piskorski, and M. Atkinson, "Real-time news event extraction for global crisis monitoring." in *NLDB*, ser. Lecture Notes in Computer Science, E. Kapetanios, V. Sugumaran, and M. Spiliopoulou, Eds., vol. 5039. Springer, 2008, pp. 207–218.

[8] R. Grishman, S. Huttunen, and R. Yangarber, "Real-time event extraction for infectious disease outbreaks," in *Proceedings of the Second International Conference on Human Language Technology Research*, ser. HLT '02. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2002, pp. 366–369.

[9] H. Becker, M. Naaman, and L. Gravano, "Beyond trending topics: Real-world event identification on twitter," in *Proceedings of the Fifth International Conference on Weblogs and Social Media, Spain, 2011*, L. A. Adamic, R. A. Baeza-Yates, and S. Counts, Eds. The AAAI Press, 2011.

[10] A. Ritter, Mausam, O. Etzioni, and S. Clark, "Open domain event extraction from twitter," in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD. USA: ACM, 2012, pp. 1104–1112.

[11] X. Wang, M. S. Gerber, and D. E. Brown, "Automatic crime prediction using events extracted from twitter posts," in *Social Computing, Behavioral-Cultural Modeling and Prediction*. Springer, 2012, pp. 231–238.

[12] L. Màrquez, X. Carreras, K. C. Litkowski, and S. Stevenson, "Semantic role labeling: An introduction to the special issue," *Comput. Linguist.*, vol. 34, no. 2, pp. 145–159, Jun. 2008.

[13] V. Lampos and N. Cristianini, "Nowcasting events from the social web with statistical learning," *ACM Trans. Intell. Syst. Technol.*, vol. 3, no. 4, pp. 72:1–72:22, Sep. 2012.

[14] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.*, vol. 12, pp. 2493–2537, Nov 2011.

[15] R. Memisevic, "On multi-view feature learning," *CoRR*, vol. abs/1206.4609, 2012.

[12] The interface is accessible at http://iot.ee.surrey.ac.uk/citypulse-social/.