Social Sensing Enhanced Time Ruler for Real-Time Bus Service

Jin Liu, Caixin Gong, Juan Li, Jian Li
State Key Lab. of Software Engineering, Computer School
Wuhan University
Wuhan, China
{jinliu, cxgong, ljwuse2012, lj1992}@whu.edu.cn

Xiaohui Cui*
International School of Software
Wuhan University
Wuhan, China
*Corresponding author email: cuixhui@gmail.com

***Abstract*—Social sensing helps realize the real-time prediction of bus routes by perceiving social events and evaluating their influence on the time when a bus passes through a road section. This paper proposes a social sensing enhanced real-time bus service that integrates sensing ability and social networks to better understand and measure social events that influence vehicle velocity. The establishment of the real-time service revolves around the *PT* service quality attributions *PEA*s and the road condition attributions *PRCA*s. The processes that collect bus relevant events and further categorize them into *PEA* events or *PRCA* events were proposed respectively. A method of scoring *PEA*s for travelers according to social events was discussed. An artificial neural network based prediction model was proposed to estimate the bus arrival time by analyzing *PRCA* events.**

Keywords—social sensing; real-time bus service; bus route; arrival time prediction

I. Introduction

Considering serious social problems caused by the rapid growth of private cars and the well-developed public transport infrastructure in most large cities, the public transport PT, especially the bus transport, promises once again a way of getting out of the dilemma such as heavy traffic cost, traffic congestion and poor air quality [1, 2]. Recent studies indicate that the PT buses would be a better mode of travel for city people if a real-time bus service can be provided, where quality attributes of PT buses such as price, convenience and environmental friendliness can be persisted at a high level by tapping the potentials of the PT bus system [3, 4].

The time ruler places a critical role in real time bus service. To understand the time ruler, one should first examine a simple but typical case of a real-time bus service: at the beginning a traveler tells the service where he would go. And the service provides him potential bus routes based on the result of the real-time bus route analysis and the estimation how the traveler go to the pointed bus stop to meet the bus route. Then the traveler chooses a bus route according to his or her preferences such as speed, ease of transfer, comfort, safety, or even aesthetics [5]. After that, the traveller catches the bus at an appointed bus stop in time and the bus route in practice meets the expectation of the traveler. This illustrative case is similar to a common mathematical problem that most of us meet in the daily life: if the traveler goes to the pointed bus stop to meet the bus route at the appointed stop with a speed x and the initial bus at the bus route rush toward the appointed stop with a speed y, when will they meet at the appointed bus stop? A time ruler helps answer such kind of question by analyzing the actual case and estimating the speed x and speed y. Without loss of generality, the case of the time ruler can be reduced to focusing on the speed y. If the time ruler helps the traveler know the accurate arrival time of the bus, uncertain and excessively long waiting time can be avoided, which encourage him to take bus again. Traditionally the establishment of the time ruler depends on conventional ways such as bus schedule that only provides the departure time and departing interval of buses. Under the best conditions, a few bus companies provide GPS devices to improve the prediction of the time ruler. Even these efforts cannot deal with most unpredictable events such as traffic jam, bad weather and temporary road-work that disturb the bus schedules and destroys the traveler's confidence in taking buses. These uncertain events may frequently be published and broadcasted on social networks as "social events" [6, 7, 8].

This paper focuses on enhancing the accuracy of time rulers on bus routes by taking advantage of the in-time traffic information collected from social networks. The primitive idea is that bus timetables and features of different commuting time still act as principal contributors to time rulers, and the in-time traffic events through social sensing can be adopted to improve accuracy of time rulers for real-time bus service. Even individual people can be regarded as social sensors. On this basis, social sensing emerges, where social events happened in specific scenes both from the real-world and the virtual network world were recorded and broadcasted through the social network. The relationship, activities and influence of events were discussed and evaluated [9, 10, 11]. In practice, the whole path of a bus is divided into road sections according to their features. The estimation of the arrival time depends on synthesis of the local time rulers on road sections. The estimation of a real-time bus route relies on the smooth connection of these local time rulers to form a macro ruler.

Based on the social sensing enhanced time ruler, the bus arrival time prediction system is consisted of three major components: (1) Sharing social information sources: authoritative social communities allow people publish,

978-1-4799-3012-8/14 $31.00 © 2014 IEEE
DOI 10.1109/SKG.2013.38

broadcast, share, discuss and estimate in-time traffic events that occurs around road sections, where individual people act as social sensors to sense and report in-time traffic events and the surrounding environment to a backend server. (2) Querying users: querying the bus route and the arrival time of the bus to a appointed stop with mobile web services; (3) Backend server: collecting the in-time reported information from the sharing social information sources, and analyzing such information so as to locate the position that traffic events occur, monitor the bus routes, determine the time ruler and predict the bus arrival time [12].

The proposed social participation on bus arrival time prediction faces at least two challenges. (1) Detection of in-time traffic events that influence road conditions: the work for social sensing enhanced time rulers only focuses on identifying events that began or occurred just previous to the present time and have continuing effects on road conditions. It is necessary to know the effect of relevant social events on bus routes by making clear how they influence road conditions. Without accurate event detection, irrelevant information would be collected for the estimation of bus routes, leading to inaccuracy time prediction and useless activity or energy waste. (2) Identification of the effect of relevant social events on bus routes: it is necessary to identify events that influence a specific road section and the degree of impact of these events on this road section from massive reports contributed by volunteers on social network. This identification cannot work well without the indication of these participants.

II. Main Tasks

Task 1 (Essential problem formulation) Given a traveler (*TR* for short) with a movement velocity x and the target bus (*TB* for short) with a cruising speed y, the essential problem of the social sensing enhanced time ruler for real-time bus service is the time estimation in two aspects: the first is the time when the traveler meets the bus at a pointed bus stop (*PBS* for short), denoted as T1, and the second is the time when the bus reaches the destination of the journey (*DES* for short), denoted as T2, where the distance between the initial position of the traveler and the position of *PBS* is denoted as $L_{TR\text{-}PBS}$, the distance between the initial position of *TB* to the position of *DES* is $L_{TB\text{-}DES}$, and the distance between the initial position of *TB* to *PBS* is $L_{TB\text{-}PBS}$.

Task 2 (Sensing of perceived *PT* service quality attribution) For a target bus, if one of its perceived *PT* service quality attributions is PEA_i, then the gathering of in-time PEA_i from a set of social networks is denoted as GA_{PEAi}, the processing of PEA_i is denoted as PR_{PEAi}, and the analyzing of PEA_i is denoted as AN_{PEAi}. The sensing of perceived *PT* service quality attributions is a process $SPEA(PEA_i<GA_{PEAi}, PR_{PEAi}, AN_{PEAi}>|i=1,\ldots, n)$ under the specific standard of features, qualities and estimation, denoted as *SPEA*.

Task 3 (Sensing of road conditions) For a physical road condition attribution $PRCA_i$ on the cruising route of a target bus, the gathering of in-time $PRCA_i$ from a set of social networks is denoted as GA_{PRCAi}, the processing of $PRCA_i$ is denoted as PR_{PRCAi}, and the analyzing of $PRCA_i$ is denoted as AN_{PRCAi}. The sensing of road conditions is a process $SRC(PRCA_i<GA_{PRCAi}, PR_{PRCAi}, AN_{PRCAi}> |i=1,\ldots, n)$ under the feature standard, qualities and estimation, denoted as *SRC*.

III. Prepare Your Paper Before Styling

The social sensing enhanced real-time bus service that devotes to satisfying travelers' requirement and bringing them favorable service experiences is mainly consisted of three components: the traveler, the mobile client and the real-time bus server, as illustrated on the lower part of Fig.1.

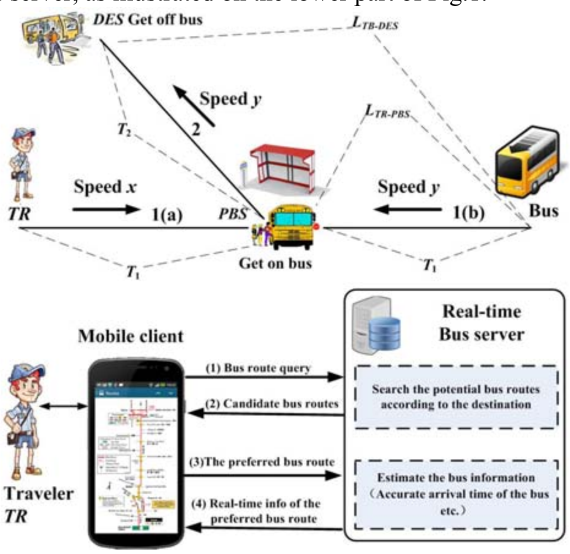


Fig. 1 Components of the social sensing enhanced real-time bus service

- **Traveler**: The traveler query the mobile client the potential bus routes for a specific destination with his preference to the mobile client application on his phone. The fundamental requirement of the traveler is to catch the bus that will arrive to his destination in time.
- **Mobile client:** The mobile client plays the role of interaction interface that bridges travelers and the real-time bus server. The mobile client collects travelers' queries and forwards them to the real-time bus server. The client also forwards the results from the server to travelers.
- **Real-time bus server:** The main task of the server is to find the potential bus routes according to the user's destination. Travelers pick up the optimal bus routes from the potential ones on the basis of their preferences such as vehicle speed, ease of transfer, comfort, safety, or even aesthetics [3].

The time ruler that makes the real-time bus service takes effect depends on the speed estimation of travelers and target buses. For the case in section 1, a traveler walks towards a pointed bus stop, takes on the target bus and takes off the bus at the destination, which was illustrated on the upper side of the speed of travelers can be formally modeled as follows:

Definition (Speed of travel) The speed of traveler can be represented as $X(x_1, x_1, x_3, x_4)$, where x_1 is the age factor that takes value as {"adult", "aged", "child"}; x_2 is the health factor as {"disabled", "non-disabled"}; x_3 is the factor of traffic environment as {"downtown", "suburb"}; and x_4 is the factor of emergency situation as {"run", "walk"}.

Definition (Road section) Roads in a universe of discourse for the social enhanced real-time bus service have been manually divided into a road section set $RoadSection\{roadsection_x|(x, i, j\in(1, N), roadsection_i\cap roadsection_j)\}$, where $i\neq j$ and $N=|RoadSection|$. The division of road sections is up to their distinctive physical characteristics. A bus route is a sub set of *RoadSection*.

Taking the public bus system as a controlled system, the limitation of the maximum speed of bus is v_{max}. Generally the traffic roads were divided into several sections in advance according to the characteristics such as cross fork, sharp bend, square, flyover and so forth. For a specific road section, the bus speed varies in a controlled interval $[0, v_{max}]$. According to task 1, the speed of target bus can be modeled as follows.

Definition (Speed of target bus) A context is a set of facts or circumstances that surround a situation or event. For any specific context such as the period of commuting time in the rush-hour, a drench of rain or temporary road cover, a constant speed for the bus on a given road section can be refined as its "average" speed on the road on the road section by analyzing its historical on this context. Thus for the whole bus route, the bus speed on the entire bus route can be estimated by combining deduced bus speeds on all road sections. Under the similar context, the cruise of the bus on the same road section can be considered as a relevant stable process. During the cruising period, the bus speed v at time t can be denoted as a relatively fixed function $v = f(t)$ that is a piecewise function refined from the statistics of the historical cruising behavior as to the bus. In most cases, actual speed may fluctuate around constant-value $f(t)$.

IV. Social sensing of perceived PT service quality attribution

A. Perceived PT Service Quality Attribution

Many attributes have been proposed to help define *PT* quality [3, 13]. Some of them can be defined as perceived attributes such as comfort, safety and convenience [3]. Comfort attribute is one kind of important bus *PT* attribute. A comfort attribute with high rating can attract travelers to ride buses and result in high bus ridership. "The crowd index", "comfortable temperature" and "equipment for the disables" are three well-known comfort attributes for bus service. For the attribution "the crowd index", many travelers are attentive to the overcrowding of a bus. For the attribution "comfortable temperature", most travelers prefer taking a bus with a working air-conditioner in the hot summer or the severe winter. For the attribution "equipment for the disables", buses that provide disable service obviously attract disable travelers to take them on. Other comfort attributions such as noise, seat availability or smooth cruise that are relevant to safety are intended omitted due to their relatively limited impact on taking buses in our work. For perceived attributes as to safety, safety attributes not only includes the common traffic accident rate, but also involves robbery and theft rate as well as the situation of sexual harassment on the bus. Female travelers mind the gravity of sexual harassment on the bus. The experience of the bus driver is an important safety attribution. According to [3], convenience is kind of attributes about the ease of use for the *PT* service and how well the *PT* service enables one's ease of mobility. Moreover, 4 perceived attributes were ultimately selected for the interaction between the social sensing enhanced real-time bus service and travelers, as illustrated in the top half part of Fig. 2.

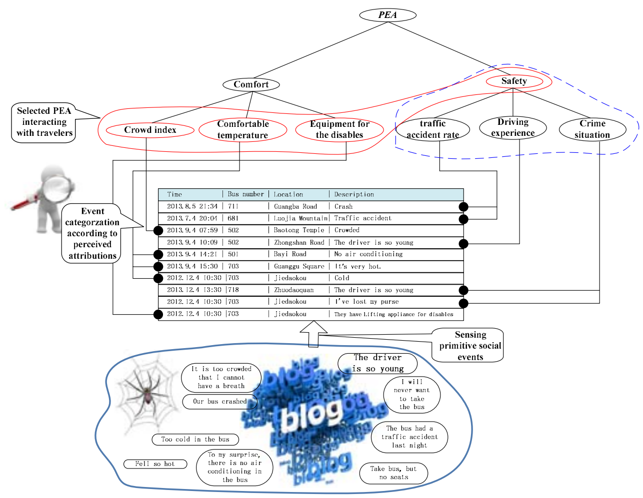


Fig. 2 The developed PEAs and category of PEA events

B. Social Sensing PEA Events

Since direct sensing and gathering of *PEA*s is often difficult and expensive, it is possible to gather *PEA*s from a set of social networks such as microblog. The general process gathering an in-time PEA_i from a set of social networks GA_{PEAi} is as follows: authors in a social media network play a role of "sensors" by sharing their experiences of buses such as "the bus is so cold and dirty", as demonstrated in the bottom part of Fig. 2. Web crawlers detect primitive social events that will be further classified into several *PEA* categories according to their contents. After that, according to the synthesis of these *PEA* labeled events, each perceived attribute PEA_i was scored for the interaction with travelers.

Since the time-effectiveness of *PEA* labeled events is important for real-time bus service, web crawlers collect social media authored within one month. A typical social media network is twitter [14]. Tweets such as "the bus 11 is so hot and dirty", "oh my god, it's boring and crowd in bus 109" and "it is so comfort in bus 583" are about the *PT* service quality. The *PEA* event can be defined as a quadruple $event_{PEA}$<*time*, *location*, *bus_id*, *event_description*>, where *time* is the release time of the event; *location* is the location of the event; *bus_id* is the unique identifier of the bus; *event_description* is the content of the event. The set of $event_{PEA}$ can be denoted as $Event_{PEA}$.

A primitive social event can be categorized into a *PEA* event type according to its content. A *PEA* can label a set of bus events that are congruent with the *PEA*. A primitive social event was intently omitted if it has not been categorized into a specific *PEA*. To categorize primitive social events into the *PEA*s, an iterative approach based on support vector machine SVM classifies events into six *PEA*s as leave node of the *PEA* tree in Fig.2. The SVM based iterative event classification involves several steps: The first step is to collect and record bus relevant social events. Then the bus relevant events were grouped into several sets $event_{si}$ according to bus identifiers, $i \in \{1, 2, \ldots, n\}$. Each event group $event_{si}$ that was denoted as $event_{si}$<*time*, *location*, *bus_id*, *event_decription*> corresponds to a specific bus identifier *bus_id*. The pre-process of each $event_{si}$ involves document parsing, gerunds extracting, stop-word processing, word restoring and word frequency counting. After that, a term frequency statistics result for each event can be obtained, denoted as *TFS4E*. Further, a training set and a testing set were specified from *TFS4E*. The *KF-IDF-DF* (keyword frequency-inverse document frequency-domain frequency) vector space of event content was also constructed upon *TFS4E*. *KF-IDF-DF* is an enhancement of the typical *TF-IDF* (term frequency–inverse document frequency) that is a numerical statistic which reflects how important a word is to a document in a collection or corpus [15, 16]. Compared with *TF-IDF*, *KF-IDF-DF* takes the importance of a term in characterizing a domain into consideration to improve the accuracy of the classification [15, 16]. Technically, *KF-IRF* (keyword frequency-inverse repository frequency) calculates the rank of keywords among the vocabulary that was associated to a given classification [16]. With *KF-IRF*, *KF-IDF-DF* can magnify the *TF-IDF* value of a keyword of a document if this keyword highly characterizes the classification that the document belongs to.

Six *PEA* event categories PEA_{Eventi} can be trained according to the training set, i=1, 2, …, 6. With the SVM classifier, an event that belongs to *TFS4E* was classified into a specific PEA_{Eventi}. The classification process continues until each event has been successfully classified into PEA_{Eventi} or casted away for its failure in the classification. An event with the successful *PEA* classification can be denoted as a *PEA* labeled event. For a *PEA* labeled event, the terms that describe the event content were ranked with *KF-IRF* algorithm to generate a term rank list *RTL* for the event. Finally, the *PEA* labeled event can be scored according with *RTL* and a pre-defined rating criteria. The pseudo-code of classification process can be programed as follows.

```
Process. PEA_event_classification(event_si)
eventset   TFS4E;            //Term frequency statistics for event
eventset   Trainset;         // Training set
eventset   Testset;          //Test set
termlist    RTL;             //Ranked term list

foreach(event in event_si)
   //Preprocess event content and generate TFS4E
   TFS4E=preprocess(event_si);
end foreach
//Identify the training set and the test set
identify(TFS4E,Trainset.count,Testset.count);
//Execute PEA event classification until all events have been processed
while(processed_event_number(TFS4E)<count(TFS4E))
   foreach (event in TFS4E)
   //Constructing vector space for event content
         build_vector_space(event);
   end foreach
   SVM(Trainset,Testset);     //Classify events with SVM classifier
end while
foreach (term in TFS4E)
   RTL=kf-irf(term);          //Generate RTL with kf-irf algorithm
end foreach
foreach(event in TFS4E)
   //Score PEA event with RTL and a pre-defined rating criteria
   score(event, RTL);
end foreach
```

To estimate the rank for a domain, the evaluation value can be scored from 1 to 5, where the score 5 stands for the degree "very high", 4 for "high", 3 for "middle", 2 for "low" and 1 for "very low".

As indicated in Fig.2 the leaf nodes of the *PEA* event tree include "crowd index", "comfortable temperature", "equipment for the disables", "traffic accident rate", "driving experience" and "crime situation". Three out of six *PEA* leaf nodes, i.e., "crowd index", "comfortable temperature", "equipment for the disables" were employed in direct interaction with travelers. The other three *PEA* leaf nodes as "equipment for the disables", "traffic accident rate", "driving experience and crime situation" were involved in indirect interaction with travelers by being synthesized into PEA "safety". To achieve the indirect interaction with travelers, an *PEA* merge operation was adopted to smooth the complex relationship between *PEA* "safety" and the others three *PEA*s, as highlighted in the right side of Fig. 2 with a closed line. The merge operation of *PEA*s "traffic accident rate", "driving experience" and "crime situation" can be formulized as $Y=\alpha x_1+\beta x_2+\gamma x_3$, where x_1 represents the score of "traffic accident rate"; x_2 represents the score of "driving experience score", x_3 represents the score of "crime situation"; α, β and γ are their respective weights, $0\leq \alpha, \beta, \gamma \leq 1$, $\alpha+\beta+\gamma=1$.

In this demonstration, there are 5 *PEA* events with the same bus id "502": e_1< "Aug 3, 2013, 15:30", "Zhongshan Road", "502", "The driver is essentially a child">; e_2<"Sept 7, 2013, 12:30", "Luoyu Road", "502", "Bus drove at a steady rate">; e_3<"Aug 12, 2013, 9:30", "Luojia Road Stop", "502", "The driver is skillful">; e_4<"Aug 12, 2013, 11:30", "Mountain Ave", "502", "bus 502 pitched and rolled and passengers were sick">; e_5<"Sept 3, 2013, 20:30", "Luojia Road Stop", "502", "The driver slammed on the brakes but failed to stop">.

The time of all these events ranges within the recent two months. The top 5 terms in *RTL* of *PEA* "driving experiences" , that is "driver", "skill", "shake", "passengers" and "safety" were assigned weight from 5 to 1.

Denote the direct score of *PEA* event as x and the direct score of the term rank in the corresponding *RTL* vocabulary as y. To calculate the synthesis of x and y, it is necessary to achieve the normalization of x and y. That is $\frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$ and $\frac{y_i - y_{\min}}{y_{\max} - y_{\min}}$ so that the synthesis of x and y can be $\alpha \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} + \beta \frac{y_i - y_{\min}}{y_{\max} - y_{\min}}$, where empirical parameter α takes the value 0.9 and β takes the value 0.5. This scoring approach may lead to the highest score of an event such as "*The driving skill of the man is so excellent*!" is 1.4. Conversely, the worse score of an event such as "*The driving skill of the man is so terrible!*" is 0.5. Thus, a reasonable event score should range from 0.5 to 1.4. An event score out of this range will be intended omitted.

Denote the value of score of a *PEA* event as *Z*, the normalization of Z that makes Z ranges from 0 to 5 is $\frac{z-0}{5-0}=\frac{w-0.5}{1.4-0.5}$. This paper takes the formula to calculate the value of score of a *PEA* event: Z=$\frac{50}{9}$×(w-0.5).

For e_1, "*driver*" ranks top 1 takes its weight as 5. "*child*" is a near-synonym of "awkward" so its experienced score is 2. Accordingly, the normalization of "*driver*" is 1 and the normalization of "*child*" is 0.25. Accordingly, the score of e_1 is (0.9 ×1+0.5×0.25-0.5)×50/9 ≈ 2.9. Similarly, the scores of e_2, e_3, e_4 and e_5, are 3.6, 5, 4.3 and 2.2 respectively.

Scores of events that corresponds to a specific *PEA* should be synthesized into the score of a *PEA* $score_{ei}$ at the working time of the real-time bus service. And the elapsed time of a *PEA* event has to be considered for the time interval between the service period of the social enhanced real-time bus service and the time posting the social event, denoted as Δt_{e_i}. The weight of e_i contributes to the synthesis of the score of a *PEA* can be as $w_{e_i} = \frac{1}{\Delta t_{e_i}} / \sum_{i=1}^{n}\frac{1}{\Delta t_{e_i}}$. Thus the synthesis of the score of a *PEA* is $\sum_{i=1}^{n}(w_{e_i} \times score_{e_i})$. For the case with 5 *PEA* "driving experience" events e_1, e_2, e_3, e_4 and e_5, suppose the serving day of the service period of the social enhanced real-time bus service is Sept 13, 2013, the intervals between the service period and the time posting e_i are 40, 6, 32, 32 and 10 days, *i*=1, 2,.., 5. The weights of of e_i are 0.205, 1.694, 0.441, 0.3794 and 0.6211 respectively. The score of *PEA* "driving experience" contributed by these five events is 0.205×2.9+ 1.69×3.6 + 0.441×5 + 0.38 × 4.3+0.621×2.2≈3.34. Considering *PEA* "driving experience" does not directly interact with travelers and *PEA* "traffic accident rate" and *PEA* "crime situation" was not discussed in this specific case, the score of *PEA* "safety" default to taking the score of the score of *PEA* "driving experience" as 3.34.

V. Social sensing of road condition attribution

This section introduces an artificial neural network ANN based prediction model to estimate the bus arrival time by perceiving and analyze social events of physical road condition attribution *PRCA* that affects vehicle velocity.

A. Physical road condition attributions

The physical road condition attributes PRCAs that have a close impact on vehicle velocity can help evaluate PT quality. Obviously, *PRCA* "Physical road condition" belongs to Physical road condition in the strict sense. Extend this limit, "Traffic environment", "Change of traffic rule" and "Specific trip demand" can also be treated as *PRCA* in a broad sense because all these attributes have capacities to work with "Physical road condition" and make buses drop their speed, stop abruptly, or even change their routes. In this way, a *PRCA* tree can be established to represent the system of physical road condition attributes in general.

To explain the system of *PRCA*s, *PRCA* "Physical road condition" contains two kinds of situation "Physical damage" and "Road reconstruction". *PRCA* "Physical damage" can further be divided into two child attributes *PRCA* "Direct physical damage" and *PRCA* "Indirect physical damage".

PRCA "Traffic environment" includes 36 attribute nodes. The better the "Traffic environment" of the current road section, the faster rate buses pass through the current road section. *PRCA* "Traffic environment" can further be divided into 6 categories: "Traffic accident", "Relatively decrease of access", "Occupying-road construction", "Sudden increase of traffic flow", "Traffic checking" and "Weather" due to the variety of situations. *PRCA* "Traffic accident" needs to consider two cases: "Individual accident" involving a single injured person and "Group accident" covering many injured people. *PRCA* "Group accident" can gets further broken down into *PRCA* "Fatal group accident" and *PRCA* "Normal group accident". *PRCA* "Relatively decrease of access" is consisted of two child *PRCAs* as "Temporal decrease" and "Regular decrease". *PRCA* "Sudden increase of traffic flow" contains two child *PRCA*s "Heavy flow caused by district change" and "Places likely to attract flow". *PRCA* "Business entertainment sports centers" can be separated into three main categories "Business", "Entertainment" and "Sports". *PRCA* "Traffic checking" concludes children "Public security check" and "Public Hygiene check". *PRCA* "Weather" includes children "Season weather" and "Temporary weather". *PRCA* "Season weather" involves "Winter", "Rainy season" and "Summer". *PRCA* "Summer" is divided into "Normal summer" and "Summer typhoon". *PRCA* "Temporary weather" that is unlimited to season factors concludes "Fog", "Dust blowing" and "No-seasonal temporary weather". *PRCA* "Change of traffic rule" includes children "No left turn", "No round turn", "Traffic control" and "Toll" that can cause the slowness of buses. PRCA "Specific trip demand" that influences the vehicle velocity includes children "Demands of specific group" and "Holiday or festival".

B. Social sensing PRCA event and prediction model of bus arrival time

The process of sensing *PRCA* events from social networks is similar to the sensing process of *PEA* events. But strictly time limit is very critical for a *PRCA* event because an in-time *PRCA* event can effectively contribute to the prediction of the bus speed and the bus arrival time. For an example, a tweet "there was a serious crash in Naijing Road" has been collected by web crawlers and classified into a specific *PRCA* category according to its content. This indicates that the content of the tweet has been identified as a *PRCA* event. The maximum time difference between the posting time of the event and the real-time bus service should be ranged from half an hour to three hours according to the time validity of events. The *PRCA* event can be defined as a *eventPRCA*<time, location, event_description>, where time is the release time of the event; location is the location the event happened; event_description is the content of the event. The set of *eventPRCA* can be denoted as *EventPRCA*.

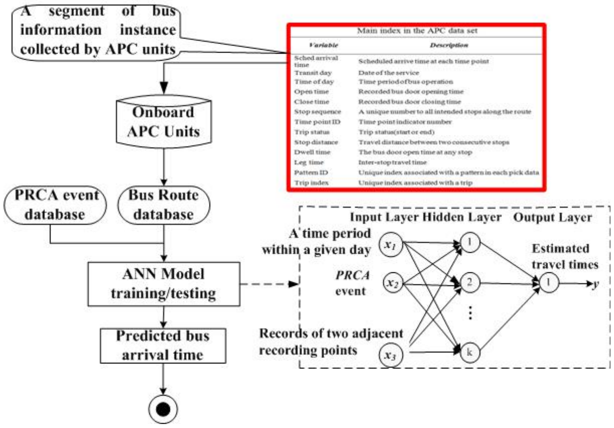


Fig. 3 The working flow of ANN based time prediction model

An ANN model with the social sensing of road conditions was adopted to predict the bus arrival time. The well-trained ANN can analyze complex relationships between the dependent variables such as the output “bus arrival time” and a set of explanatory variables such as “the road conditions”. The left side of Fig. 3 describes the working flow of the proposed ANN prediction model. The automatic passenger counter *APC* as equipment installed on the bus can report the time that the bus arrive to a specific bus stop [17]. The *APC* units have been equipped by many buses to monitor bus state at ACP sensing points along their routes. The ANN model for trip query is established according to the characteristics of the historical trip data recorded by *APC* units and the gathered in-time *PRCA* events from the social networks. The training of the ANN model was continues and its update rate depends on the update frequency of collected data and event. The output of model is the bus travel time from two adjacent time nodes. With this ANN based time prediction model, the bus arrival time to an appointed bus stop can be estimated.

A bus information segment collected by APC units as the input of the working flow of ANN based time prediction model is illustrated in the upper right hand corner of Fig. 3. The APC data was processed according to different time periods. The workdays from Monday to Friday (MF) and the weekend from Saturday to Sunday (WE) were divided into two categories. each day in a category was further divided into 8 time sections with time interval 3 hours, i.e., the early morning (EM), the morning peak (MP), the late morning(LM), the mid-day (MD), the early afternoon (EA), the afternoon peak (AP), the evening (E), and the late night (LN).

Generally, the APC mechanism divides a given road section into several fixed recording points. The ANN model takes records of two adjacent recording points as one of its input parameters in the current iteration. Thus, three input parameters can be identified to establish the ANN model: that are a 3-hour time period within a given day (x_1), *PRCA* event(x_2), and the records of two adjacent recording points (x_3). A typical value of x_1 is <MF, EM>. The output of the ANN based time prediction model at each iteration is the estimated travel time for the next recording point. In this way, the bus arrival time can be calculated by synthesizing the time at previous recording point and the estimated travel time interval to the next recording point at each iteration.

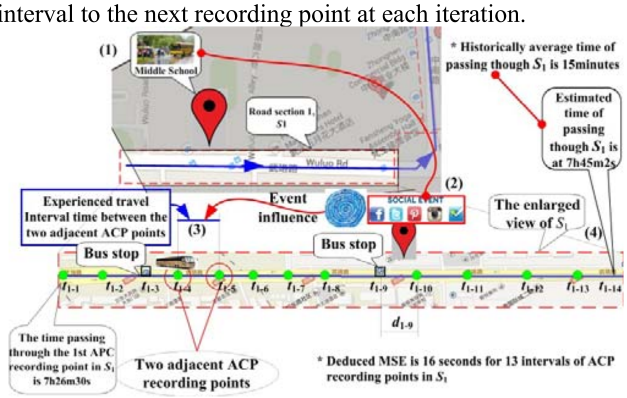


Fig. 4 A simple case of social sensing of *PRCA* events

For an ANN model, there is a hidden layer to adjust the basis of the ANN model according to a back-propagation algorithm [18]. The hidden layer contains the learning process to find a weight matrix that minimizes the mean squared error, denoted as MSE. MSE is defined as the average squared error between the result of prediction and the actual value, which can be represented as a formula MSE= $\frac{\sum_{j=0}^{P}\sum_{i=1}^{N}(d_{ij}-y_{ij})^2}{NP}$, where P is the number of output neurons; N is the number of samples in the data set; d_{ij} is the desired output for sample i at neuron j and y_{ij} is the ANN output for sample i at neuron j.

Take the route of a bus 801 for an example, the route was divided into several road sections as illustrated in Fig. 4. The example focuses on dicussing the road section S_1 that has 13 ACP recording points denoted as $t_{1\text{-}1}, t_{1\text{-}2}, ..., t_{1\text{-}13}$. The distance between two adjacent points can be denoted as $d_{1\text{-}j}$, j =1, 2, …, 12. For the demonstrating case, the 1st ACP recording point recorded that the bus entered S_1 at 7:26:30. The input of the ANN model at the recording point $t_{1\text{-}9}$ is reported as <<MF, LM>, *PRCA* event<“Sept 16, 2013, 7:32”, “Wuluo Road”, “There are many students in the road” >, “$d_{1\text{-}9}$”>, as indicated in Fig. 4. The performance of the proposed ANN model was evaluated by MSE that was programmed with MathWorks Matlab [19]. MSE was deduced as 16 second delay for 13 intervals of ACP recording points in S_1. Thus the prediction of the time when the bus 801 left S_1 was 7:45:2.

VI. Case Study

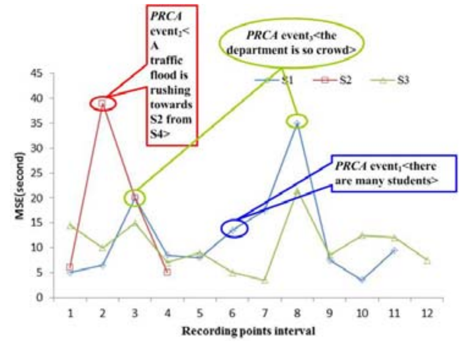
To systematically demonstrate the work of social sensing enhanced real-time bus service, a complex case involves 4 road sections was represented. A traveler at the Soyi park submitted a real-time bus route request to go to the Hongshan hotel at 15:45, September 15, 2013. The traveler is a young office worker. The social enhanced real-time bus service provided him two bus routes. The first bus route is that the traveler walks to the bus stop at the Bus Station "Dadongmen" within 3 minutes, then takes the bus 578, gets off the bus at the station "Hongshan Road" and walks to "Hongshan hotel" at "Hongshan square" within 3 minutes. The second bus route is that the traveler walks to the bus stop at the Bus Station "Dadongmen" within 3 minutes, then takes the bus 577, gets off the bus at the bus station "Zhongnan road 2", and walks to "Hongshan hotel" at "Hongshan square" within 8 minutes.

Fig. 5 A complex case of social sensing enhanced real-time bus service

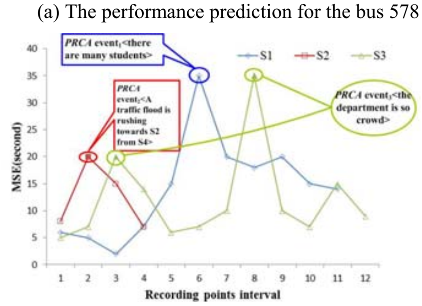
The routes that the buses get past mainly involve three connected road sections S_1, S_2 and S_3 in Fig. 5. This picture also marked a road section S_4 that bus 587 or 577 does not pass through. There are 4 bus stops as "Dadongmen", "Middle school 15", "Zhongnan road 2" and "Hongshan road" on the bus routes. S_1 can easily get into traffic jam when crowded students go to school in the period <MF, LM>. When *PRCA* events such as "crowded students go to school" were detected in time, the chance of the bus arrival time being delayed becomes very high. S_2 is a T-shaped road junction that most drivers slow down their passing velocity. The bus passing time is easy to be postponed because S_2 is vulnerable to increasing traffic flows. The traffic environment of S_3 is subject to change because a primary school and two commercial department stores were close near to the street.

The first step of the bus service provides the traveler a visual interface to represent his travel preferences. An important part of this interface is PEA preferences. The *PEA* values of these buses were recalculated periodically according to the contents of *PEA* events continuously gathered from the several well-known traffic social networks. The in-time *PEA* event was restricted within one month. For the recommended bus 578, its *PEA* scores are *PEA* "Crowded index" 4.15, *PEA* "Comfortable temperature" 2.98, *PEA* "Equipment for the disables" 0 and *PEA* "Safety" 3.95. For the recommended bus 577, its *PEA* scores are *PEA* "Crowded index" 3.15, *PEA* "Comfortable temperature" 4.15, *PEA* "Equipment for the disables" 0 and *PEA* "Safety" 3.20. The fact that the score value of *PEA* "Equipment for the disables" for both buses is 0 implies that no social events indicate there is any equipment for the disables on the buses. Since the bus 578's *PEA* "Crowded index" 4.15 is greater than the bus 577's *PEA* "Crowded index" 3.15, the traveler decided to take the bus 577. Since the traveler can take either the bus 578 or the bus 577 at the bus station "Dadongmen", the bus service indicates that it takes the traveler three minutes to walk from Soyi park to the bus station "Dadongmen".

For this route query, there are three *PRCA* event has been detected as *PRCA* event as the input parameter of the ANN based arrival time prediction model: e_1<"Sept 16, 2013, 15:32", "Wuluo Road", "there are many students going home">, "d_{S1-9}">, *PRCA* event e_2<"Sept 16, 2013, 15:30", "Wuluo Road", "A traffic flood is rushing towards the Zhongnan t-shape junction (S_2) from Baotong temple (located at S_4)" >, "d_{S4-2}"> and *PRCA* event e_3<"Sept 16, 2013, 15:27", "Wuluo Road", "the department is so crowded" >, "d_{S3-2}">. These in-time *PRCA* events collected at different periods on these road sections are various.



(a) The performance prediction for the bus 578



(b) The performance prediction for the bus 577

Fig. 6 The performance prediction for the bus

For the bus 578, according to our experimental knowledge, the historical average time for a bus passing through S_1, S_2 and S_3 at this period <MF, LM> is 15, 2 and 18 minutes respectively. According to the prediction of the ANN prediction model, e_1 leads to 61.5 seconds time delay with respect to the average travel time through S_1. e_2 leads to 70 seconds time delay with respect to the average travel time through S_2. e_3 leads to 85.5 seconds delay with respect to the average travel time through S_3. According to the error analysis in Fig. 6(a), the maximum error is 39 seconds and the minimum error is 3.5. The average MSE values are 12.2 seconds on S_1, 21.3 seconds on S_2 and 10.5 seconds on S_3. The arrival time to the target location is consisted of three parts. The first is the walking time from Soyi park to "Dadongmen" bus stop is within 3 minutes. The second is the travel time through S_1, S_2 and S_3, which is the sum of the average time pulsing the time delay in road sections respectively: (15minutes +61.5 seconds) +(2 minutes +70 seconds) +(2 minutes +85.5seconds)=38 minutes and 37 seconds. The third is the walking time from the bus stop that the traveler takes off to the target location Hongshan hotel is 3 minutes. Thus, the final prediction that the traveler arrives to his target location is at 16:29:37.

For the bus 577, the historical average time for the bus passing through S_1, S_2 and S_3 at <MF, LM> is 15, 2 and 8 minutes respectively. According to the prediction of the ANN model, e_1 leads to 70 seconds time delay with respect to the average travel time through S_1. e_2 leads to 50 seconds time delay with respect to the average travel time through S_2. e_3 leads to 114 seconds delay with respect to the average travel time through S_3. According to the error analysis in Fig. 6(b), the maximum error is 35 seconds and the minimum error is 2. The average MSE values are 14.2 seconds on S_1, 12.5 seconds on S_2 and 12.1 seconds on S_3. The arrival time to the target location is consisted of three parts. The first is the walking time from Soyi park to "Dadongmen" bus stop is within 3 minutes. The second is the travel time passing through S_1, S_2 and S_3. That is the sum of the average time pulsing the time delay in road sections respectively: (15minutes +70 seconds) + (2 minutes +50 seconds) +(8 minutes +114 seconds)=28 minutes and 54 seconds. The third is the walking time from the bus stop that the traveler takes off to the target location is 8 minutes. Thus, the prediction that the traveler arrives to his target location is at 16:25:54.

VII. Conclusions

In our work, the detection of in-time traffic events that influence road conditions was achieved by analyzing features of road sections and in-time social events. Simple leaning and reasoning ability of the experimental knowledge of the influence of events to a specific road section was trained to estimate the influence of a reported social event on specific road sections. Compared with conventional ways, there are several advantages of the social participation approach for enhancing the prediction of bus arrival time. First, the integration of sensing ability and social networks provides a better understanding and measurement of the aggregate behavior of a community that concerns traffic events together with the external environment where the community functions. By doing so, a better understanding of road conditions in a city would be possible. Massive individuals collectively report events of interest via network. These events can influence the speed of a bus passing through some road sections. Second, if the events can be perceived and evaluated, their influence on the time in which a bus passes through a road section could be more precisely evaluated. Third, if the accuracy of the time ruler can be kept, it is possible to provide more complex service of bus prediction such as bus transfer, which keeps high arrival rate on time.

Acknowledgments

This work was supported by the grants of National Natural Science Foundation of China (61070013, U1135005), NSFC Research Fund for International Young Scientists (61250110541), Programme of Introducing Talents of Discipline to Universities (No. B07037).

References

[1] Rémy Prud'homme, Martin Koning, Luc Lenormand, Anne Fehr, "Pub lic transport congestion costs: the case of the Paris subway," Transport Policy, 21, 2012, pp. 101–109.

[2] Chiara M. Travisi, Roberto Camagni, Peter Nijkamp, "Impacts of urba n sprawl and commuting: a modelling study for Italy," Journal of Tran sport Geography, 18(3), 2010, pp.382–392..

[3] John Zimmerman, Anthony Tomasic, Charles Garrod, Daisy Yoo, Cha ya Hiruncharoenvate, Rafae Aziz, Nikhil Ravi Thiruvengadam, Yun H uang, Aaron Steinfeldl, "Field trial of tiramisu: crowd-sourcing bus arr ival times to spur co-design," SIGCHI Conf. on Human Factors in Co mputing Systems (CHI'11), 2011, pp. 1677-1686.

[4] Lauren Redman, Margareta Friman, Tommy G¨arling, TerryHartig, "Q uality attributes of public transport that attract car users: a research revi ew," Transport Policy, 25(2013)119–127.

[5] Daisy Yoo, John Zimmerman, Aaron Steinfeld, Anthony Tomasic, "U nderstanding the space for co-design in riders' interactions with a trans it service," Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'10), 2010, pp. 1797-1806.

[6] N. Wanichayapong, W. Pruthipunyaskul, W. Pattara-atikom, P. Chaov alit, "Social-based traffic information extraction and classification," Int .Conf. on ITS Telecommunications (ITST'11), 2011, pp.107-112.

[7] H. Zhuge, "Semantic linking through spaces for cyber-physical-socio intelligence: a methodology", Artificial Intelligence, 175, 2011, pp. 988-1019.

[8] H. Zhuge, "The knowledge grid -toward cyber-physical society", World Scientific Publishing Co., Singapore, 2012. 2nd Edition.

[9] H. Zhuge, "Communities and emerging semantics in semantic link network: discovery and learning", IEEE Transactions on Knowledge and Data Engineering, 21(6), 2009, pp. 785-799.

[10] Xiangfeng Luo, Zheng Xu, Jie Yu, Xue Chen, "Building association link network for semantic link on web resources", IEEE T. Automation Science and Engineering 8(3), 2011, pp. 482-494.

[11] Zheng Xu, Xiangfeng Luo, Jie Yu, Weimin Xu, "Measuring semantic similarity between words by removing noise and redundancy in web snippets", Concurrency and Computation: Practice and Experience, 23(18), 2011, pp.2496-2510.

[12] Zheng Xu, Xiangfeng Luo, Jie Yu, Weimin Xu, "Mining Web search engines for query suggestion", Concurrency and Computation: Practice and Experience, 23(10), 2011, pp. 1101-1113.

[13] C. Jakobsson Bergstad, A. Gamble, O. Hagman, M. Polk, T. G¨ arling, "Affective-symbolic and instrumental-independence psychological m otives mediating effects of socio-demographic variables on daily car us e," Journal of Transport Geography, 19(1), 2011, pp.33–38.

[14] Raymondus Kosala, Erwin Adi, Steven, "Harvesting real time traffic i nformation from twitter," Procedia Engineering, 50, 2012, 1-11.

[15] Zheng Li, "An on-demand service discovery approach based on mined domain knowledge",IEEE 8th World Congress on Services (SERVICE S), 2012, pp.24-29.

[16] Jian Wang, Jia Zhang, Patrick C.K. Hung, Zheng Li, Jianxiao Liu, Keq ing He, "Leveraging fragmental semantic data to enhance services disc overy", Int. Con. on High PerformanceComputing and Communications, 2011, pp.687-694.

[17] J. Patnaik, S. Chien, A. Bladikas."Estimation of bus arrival times using APC data", Journal of Public Transportation, 7(1), 2004, pp. 1-20.

[18] Helen Zhang, David Jin and X.J. Zhao. "Application of ANN back-propagation for fracture design parameters of middle carbon steel in extra-low cycle bend torsion loading," 345, 2013, pp. 272-276.

[19] Tamás Tettamanti, István Varga, "Development of road traffic control by using integrated VISSIM-MATLAB simulation environment", Civil Engineering, 56(1), 2012, pp.43-49.