VANET Multicast Routing for Congestion Control in Traffic Flow WSN

Hao WU

Nanjing Communications Institute of Technology, Nanjing, Jiangsu, 211100, China

Abstract — Security data fusion is fundamental and essential for intelligent traffic systems. A dynamic data fusion scheme is proposed to guarantee the security and accuracy of sensed data. It is used to collect and manage multi-source traffic messages. It improves K-neighbour nonparametric regression method on the basis of repeated occurrence feature of traffic flow state mode. Double-screening for neighbors is adopted, and identification function based on state mode is introduced and traffic flow of past time frame and traffic flows of related turning at upstream junction and downstream junction are considered in algorithm. This helps to improve capacity prediction of K-neighbour nonparametric regression. Final prediction result is given by the weighted average method of match distance reciprocal on the basis of state mode vector, thus prediction of short-time traffic flow becomes more accurate in real-time.

Keywords- intelligent traffic; wireless sensor network; data fusion; multicast routing; vanet

I. INTRODUCTION

Collection of traffic message is a key technology in intelligent traffic system. Sensor network acquires accurate traffic parameters like vehicle speed, traffic flow and road occupation rate for intelligent traffic system and these parameters are the basis for traffic management. Traditional monitoring sensors like induction coil and camera restrict the expendability of existing system and affect network efficiency but multi-source traffic message data fusion based on WSN can acquire more accurate traffic message than traditional sensor and more effective traffic monitoring and management [1], including electronic charging, parking management, junction traffic guidance, energy conservation and emission reduction, are realized. In the intelligent traffic scene, collected traffic message varies and original data size is huge. Moreover, as vehicle nodes are highly mobile, vehicle sensor network has the topological structure with more changes than that of static sensor network. Although these data fusion schemes have advantages in some aspects, there are still deficiencies because algorithm for intelligent traffic system is too complex and energy consumption of fusion node is too much. Multi-source traffic message data fusion based on application of WSN in intelligent traffic system is studied in the thesis and a dynamic fusion scheme is proposed. The scheme adapts to time-space distribution characteristic [6] of transportation. Data fusion algorithm that design clusters subject to credibility evaluation of the node can reduce energy consumption of fusion node and improve the reality of sensed data.

II. FUSION THEORY BASED ON EVIDENCE REASONING

Evidence reasoning method proposed by Dempster Shafer [7] draws more attention to uncertain factors and unknown factors and is more approximate to human thinking logic and natural decision-making process [8]. A data fusion method is acquired on the basis of Dempster Shafer evidence reasoning theory and is applied to intelligent traffic field.

Dempster Shafer’s theory is mainly applied to target identification and classification. In such fusion method, each sensor obtains local decision and sends them to fusion center for final aggregate decision. The basic concept is described as follows:

Suppose L is finite proposition language, \( \Theta = \{w_1, w_2, ..., w_n\} \) is possible world set and \( w_i = \{i=1,2,......,n\} \) is one explanation of L. A proposition can be expressed as a subset A of \( \Theta \). Set \( \Theta \) has the following features: (1) finiteness. (2) Elements in the set repulse each other.

**Definition 1** Suppose \( \Theta \) is identification frame, function \( m: 2^\Theta \rightarrow [0,1] \) is a basic probability assignment, when and only when (1) \( m(\phi) = 0 \) and \( \sum_{A \in \Theta} m(A) = 1 \). If \( m(A)>0 \), A is a focal element of the function.

**Definition 2** Function Bel: \( 2^\Theta \rightarrow [0,1] \) is called belief function. If Bel(A)= \( \sum_{B \subseteq A} (B) \), Bel(A) here is called belief degree of A.

**Definition 3** Function PI: \( 2^\Theta \rightarrow [0,1] \) is called likelihood function. If PI(A)= \( \sum_{B \cap A \neq \phi} m(B) \), PI(A) here is called likelihood degree of A.

Belief degree Bel(A) refers to the sum of belief degrees of propositions that clearly support the proposition A expressed by A and likelihood degree PI(A) refers to sum of likelihood degrees of propositions that are expressed to potentially support proposition A. Relations PI(A) \( \geq \) Bel(A) and PI(A)=1−Bel(\( \overline{A} \)) are true.

Combination rule: suppose \( m_1 \) and \( m_2 \) are two basic probability assignments on \( \Theta \) and their focal elements are...
HAO WU: VANET MULTICAST ROUTING FOR CONGESTION CONTROL ORIENTED TRAFFIC FLOW WSN

A_1, ..., A_p and B_1, ..., B_q respectively. Bel_1, Bel_2 and Bel_3 = Bel_1 ⊕ Bel_3, are belief function \( \sum_{A_i \neq B_j \neq \phi} m_i(A) \pi m_j(B) < 1 \) induced by \( m_1, m_2 \) and \( m_3 = m_1 \oplus m_2 \) so \( m_3 = m_1 \oplus m_2 \) is defined as:

\[
m_3(\phi) = 0, \quad m_3(A) = K \sum_{A_i \neq B_j \neq A} m_i(A) m_j(B),
\]

\[K = [\sum_{A_i \neq B_j \neq \phi} m_i(A) \pi m_j(B)]^{-1}.
\]

The rule above can be generalized to multiple m functions and Bel functions and applied to integrate opinions of multiple experts. In Dempster Shafer’s theory, a piece of evidence can determine a basic probability assignment and further determine a belief function. Therefore, Dempster Shafer combination rule is called evidence combination rule.

III. DATA FUSION SCHEME

A. Basic Idea

Collection and management of traffic message are based on data meshing system. Traffic network is meshed subject to specific principle and method [9, 10]. Cluster head node in the cluster area acquires sampling data of other members and fuses them into non-redundant data set. Then the data package would be transmitted to aggregate point/base station by making time stamp and geological position on GPS. Base station distinguished fused data from different cluster areas according to timestamp and geological position and upload then to application layer. Application layer uses credibility evaluation and evidence function reliability distribution to calculate fusion result and makes the final decision. Flow of fusion scheme is as shown in Figure 1. The scheme improve the accuracy of multi-source traffic message data and prolong the survival time of network.

B. Data Fusion Process

Cluster head node in the cluster area periodically sends inquiry message MSGREQ to other member nodes to synchronize time. In a sampling term \( t_i \), member node \( A_i \) (vehicle or facility on road side) sends response message REP, to cluster head and formalized description of the message model is as follows:

\[m_i = MSG \oplus REP \parallel D_i \parallel SN_i \parallel data \parallel Pos \parallel sensid \parallel Token\]

\( D_i \) : preamble of \( i \), that reflects data type sensed by node \( A_i \) and infers actual source data. It is different from preamble of other data packages.

\( SN_i \) :the identity of sensor node \( A_i \)

\( data = E(d_{ans}, K_{A_i, BS}) \), where, \( K_{A_i, BS} \) is the symmetric key between \( A_i \) and BS to protect the privacy of sensed data. \( d_{ans} \) is the actual data sensed by \( A_i \), including degree of support for the data value.

\( Pos \) : traffic mesh coordinate \((x, y)\), to describe geological position of message source and position the target [11].

\( sensid \) : identifier of current data package, differing from different messages sent by the same sensor node. In \( sensid = F(sensid') \mod M \), expression, \( F \) is monotone increasing function and \( sensid' \) is the identifier of last data package.

\( Token \) : message verification domain,

\[Token = SIG(D_i \parallel SN_i \parallel data \parallel Pos \parallel sensid, K_{A_i, TP})\]

\( SIG \) is the digital signature of node \( A_i \), \( K_{A_i, TP} \) is private key for verification of \( A_i \) and they are both issued by verification center at network layer.

In time interval \( \Delta t \), the sensed data would be forwarded to cluster head node from the collection position or via intermediate node. In the intelligent traffic environment, sensor node can capture various data simultaneously (like air flow, temperature and humidity). In order to identify data with different attributes, when source node forms preamble, specific separator is inserted between the data with different attributes to differentiate attributes and node types of data. The preamble region value \( D_i \) in data package replaces actual data and executes data fusion. Cluster head node would sort out data packages with the same attribute into same group and fuse them into a new message [12]. In addition, the sensed ranges of members in the same cluster area would overlap due to the distance. Value of \( Pos \) area would be expressed as sensed range in traffic network with circle center of \((x, y)\) and radius of parameter \( R \). In this range, other modes can acquire the same data. Therefore, introduced parameters \( D_i \) and \( Pos \) can avoid data being fused in different groups and forming redundant data set.

Suppose a group of aggregate data sets is composed of \( l \) pieces of different messages, cluster head \( A_{ans} \) fuses data and

![Figure 1. Data fusion scheme](image_url)
Suppose $r_i$ is the maximum number of sources, $m_i = \sum_{j} m_{ij} \cdot \rho_{ij} + \sum_{j} m_{ij} \cdot \rho_{ij}$, where $\rho_{ij}$ is the proportion of the number of sources $i$ in the historical records of the area $S$. If $S$ is constant coefficient, suppose $g = 5$. $\sigma_i$ is the test accuracy of source node $A_i$, which is the proportion of the number $N_{success}$ that test results of $A_i$ are consistent with judgment of application layer in the previous number $N_{total}$.

When there are $n$ data sources, belief function $Bel_i(A)$ of source node $A$ for evidence $A$ can be obtained via equation (1).

$$Bel_i(A) = m_i(A) \cdot \rho_{i} \cdot \rho_{ij} + m_i(A) \cdot \rho_{ij}, A \in \Theta$$

(2)

Suppose $I$ is the maximum subset of event in $U$. Belief functions of $n$ data sources are combined via equation (2) and belief function is updated subject to equation (3).

$$Bel(B_i) = m_i(A_i) \otimes \ldots \otimes m_i(A_k) = k \sum_{A_i \cap \ldots \cap A_k = \emptyset}^{m_i(A_k)} \prod_{i=1}^{k} m_i(A_i)$$

(3)

$$k = \left\lfloor \prod_{i=1}^{n} m_i(A_i) \right\rfloor$$

Given $B_1, \ldots, B_m$ are mutually exclusive, it can be obtained:

$$Pl(I) = 1 - \sum_{A_i \cap \ldots \cap A_k = \emptyset}^{B_k} Bel(B_i)$$
coefficient $w_i$ is more approximate to 1. On the contrary, it would be more approximate to 0. Once evidence function $m(A)$ is close to 0, the evidence is highly likely to be neglected and the support degree of the assuring event declines. Reliability distribution of evidence function is realized on the basis of credibility evaluation, which weakens the impact of low reliability of dishonest evidence on the fusion result.

IV. PERFORMANCE SIMULATION

All the experiment data come from real-time monitoring for the traffic flow on Tibet Road Section of Shanghai City. Considering traffic flow is periodic, monitoring time was from June 6, 2015 to June 13, 2016, including 5 business days and three days of festival. As there were few cars at night, actual use value of traffic data is small, the monitoring time frame was from 7:00am to 20:00pm and timer interval is 2min. 3120 original data samples of traffic flow were obtained. The data of the first 7 days were used to establish historical sample database and the data of the last day served as test data. Daminghu Road Section Diagram is as shown in Figure 1.

Traffic flow prediction method in the thesis includes 5 parameters, including $l$, $n$, $m$, $j$, $b$ and $k$. $l$ is the dimension of state vector. State vector dimension is directly related to prediction accuracy and algorithm efficient. $n$ is the number of dots after passing the first state mode match screening. $k$ is the number of dots after passing the second state mode match screen. $n$ and $k$ directly affect prediction accuracy and algorithm efficiency and excessively big or small $n$ and $k$ would reduce prediction accuracy.

According to the experiment data above, values of parameters $l$, $n$ and $k$ are obtained. It can be seen from Figure 2, when parameter $l$ increases from 2 to 6, prediction error declines significantly; when $l$ increases again, the prediction accuracy remains on the same level because increment of $l$ requires more calculation. Its best match number is around 4. The figure shows, even when the match number increases, the prediction effect would not be improved dramatically and it would may bring forth reverse effect so suppose $l=4$ in the actual application.

Observe Figure 2. On the premise of confirmed state vector and prediction algorithm, when $n$ increases from 40 to 50, prediction error drops significantly; when it increase from 50 to 65, the prediction error rises gradually but the rising speed is slow, which means the best match number $n$ is 50. After the first neighbor number is confirmed, experiment of the second neighbor number was conducted, $k$ is the number of dots that passed the second state mode match screening. As shown in Figure 4, of $k$ is excessively big, the prediction function is excessive smooth and prediction accuracy drops. However, $k$ cannot be excessively small, which would increase the ratio of chance factor and affect prediction accuracy. Suppose $k$ is 9 here.

We compared traditional K-neighbor algorithm and improved K-neighbor algorithm based on state mode, substituted parameters into formulas to predict traffic flow at
next moment and we used Matlab in simulation and obtained better prediction result. Therefore, prediction performance of K-neighbor algorithm based on state mode that is proposed in the thesis is better.

V. CONCLUSIONS

Traditional K-neighbor nonparametric method is improved on the basis of repeated occurrence feature of traffic flow state mode in the thesis. Double-neighbor nonparametric regression method is adopted, identification function based on state mode is introduced in neighbor nonparametric regression method, traffic flow of past time frame and traffic flows of related turning at upstream junction and downstream junction are considered in algorithm so the prediction capacity of K-neighbor nonparametric regression method is improved and final prediction result is given by the weighted average method of match distance reciprocal on the basis of state mode vector. Finally, based on the analysis of measured traffic flow prediction result, prediction of short-time traffic flow obtained by the improved double K-neighbor nonparametric regression method is more accurate and real-time. it is an effective method to predict short-time traffic flow and its prediction result is basis for the traffic guidance and control service of traffic management department and the method is vital for the traffic control and guidance.

ACKNOWLEDGMENTS

The high level talents scientific research project.

REFERENCES