Optimal Selection of Traffic Sensors: An Information-Theoretic Framework*

Ruzanna Mat Jusoh and Konstantinos Ampountolas

Abstract— **This paper presents an information-theoretic framework for the optimal selection of sensors across a traffic network. For the selection of sensors a set covering integer programming (IP) problem is developed. A measure of correlation between random variables, reflecting a variable of interest, is introduced as a "distance" metric to provide sufficient coverage and information accuracy. The ultimate goal is to select sensors that are most informative about unsensed locations. The Kullback-Leibler divergence (relative entropy) is used to measure the dissimilarity between probability mass functions corresponding to different solutions of the IP program. Efficient model selection is a trade-off between the Kullback-Leibler divergence and the optimal cost of the IP program. The proposed framework is applied to the problem of developing sparse-measurement traffic flow models with empirical inductive loop-detector data of one week from a central business district with about sixty sensors. Results demonstrate that the obtained sparse-measurement rival models are able to preserve the shape and main features of the full-measurement traffic flow models.**

I. INTRODUCTION

Location science has a long history in single criterion location problems characterized by the (maximum or average) distance, or some measure more or less functionally related to distance (e.g., average travel time, demand satisfaction). The general problem is to locate new facilities or sensors to optimize some objective or cover a spatial area of interest or satisfy some demand points. Basic facility or sensor location models include: set covering, maximal covering, p-center, p-median, fixed charge, hub, and maxisum [1].

These models work well in spatial problems where distance as a metric is well defined. However, geometric assumptions related to the metric are too strong in case of monitoring spatio-temporal phenomena, such as traffic flow in urban road networks by traffic detectors. Traffic congestion propagates upstream in the network to random locations (see e.g. [2]) and traffic sensors make noisy measurements about the nearby regions, and this spatial sensing area is not usually characterized by a regular disk. For example, sensors located at arbitrary regions of the network can provide similar information and thus should be excluded. On the other hand, combining data from multiple sensors can give good predictions. For the network of Glasgow city [2], it has been demonstrated that signal control of an intersection is affected by the traffic conditions even of relatively distant links (irrespective of whether the intersection is located centrally or at the network boundaries). Moreover, the spatio-temporal distribution of congestion in traffic networks affects the shape of aggregated models used for traffic monitoring [3], [4]. Thus the notion of combination of data from multiple sensors is of fundamental importance and cannot be easily characterized by existing spatial models relying on the measure of distance. Except transport networks, the sensor selection problem arises in various other applications, including robotics, parametric identification of structural systems [5], wireless networks [6], and others [7].

This work presents a novel methodology for the optimal selection of sensors across a transport network, including an information theoretic framework for efficient model selection. For the optimal selection of sensors a set covering IP problem is developed, which is NP-hard (even with only polynomially many constraints [8]). Though polynomial time algorithms with a constant-factor approximation guarantee can be developed. A measure of correlation (based on mutual information) between random variables, reflecting a variable of interest, is introduced as a "distance" metric to provide sufficient coverage and information accuracy. The ultimate goal is to select sensors that are most informative about unsensed locations. The problem of finding the configuration that maximises mutual information is NP-complete [9]. In this work, the Kullback-Leibler divergence is used to measure the dissimilarity between probability mass functions corresponding to different solutions of the IP program (unlike other works without making any assumption on the measurement model and its distribution), and thus to quantify the approximation error between different group of sensors. Efficient model selection is a trade-off between the Kullback-Leibler divergence and the optimal cost of the IP program. The effectiveness of the proposed framework is demonstrated with the use of empirical data.

II. INFORMATION THEORY

Let X be a discrete random variable that is completely defined in a finite set $\mathcal{X} = \{0, 1, 2, \ldots\}$. The value $p_X(x) = \mathbf{P}(X = x)$ is the probability that the variable X takes the value x. Then $p(x)$ defines a probability mass function (pmf) for the discrete random variable X with support $\mathcal{X}$. All random variables and distributions are considered discrete.

The (Shannon) entropy of a pmf is a non-negative measure of the amount of "uncertainty" in the distribution [10].

Definition 1 (Entropy): The *entropy* $H[p(x)]$ of a distribution $p(x)$ is defined by (when the sum exists)

$$H[p(x)] \triangleq -\sum_{x\in\mathcal{X}} p(x)\log p(x) = -\mathbf{E}[\log p(X)]\,. \quad (1)$$

*The first author is supported by the Malaysian Government and the National Defense University (NDUM) of Malaysia.

R. M. Jusoh and K. Ampountolas are with the School of Engineering, University of Glasgow, Glasgow G12 8QQ, United Kingdom. E-mail: r.mat-jusoh.1@research.gla.ac.uk and konstantinos.ampountolas@glasgow.ac.uk

Note that $H[p(x)]$ and $H[X]$ are equivalent. The operator $\mathbf{E}_p[\cdot]$ denotes expectation, where X is drawn according to the distribution $p(x)$. Here entropy is measured in *bits* (base 2 logarithm) and $0 \log 0$ is defined to be 0.

The *relative entropy* or the *Kullback-Leibler divergence* (KL-div) of two distributions is denoted as $\Delta(p \,\|\, q)$. It is a measure of the inefficiency of assuming that the distribution is q when the true (reference) distribution is p. It arises as an expected logarithm of the likelihood ratio [11].

Definition 2 (Relative Entropy; KL Divergence): Let $p(x)$, $q(x)$, $x \in \mathcal{X}$, be two pmf with $q(x) \ll p(x)$. The relative entropy of $q(x)$ with respect to $p(x)$ (reference distribution), or the Kullback-Leibler divergence of $q(x)$ from $p(x)$ is defined as

$$\Delta(p \,\|\, q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)} = \mathbf{E}_p\left[\log \frac{p(X)}{q(X)}\right]. \quad (2)$$

Here $0 \log 0/p$ and $0 \log 0/0$ are defined to be 0, while $p \log p/0$ is defined to be ∞. The KL-div is not symmetric under interchange of the distributions p and q ($\Delta(p \,\|\, q) \neq \Delta(q \,\|\, p)$), and it does not obey the triangle inequality, i.e., it is not a classic *distance measure*.

Convergence of probability distributions, $p^n \to p$, means point-wise convergence, that is, $p^n(x) \to p(x)$ for each $x \in \mathcal{X}$. A key property of KL-div is that it is non-negative and zero if and only if two distributions are equal.

Lemma 1: Let $p(x)$, $q(x)$, $x \in \mathcal{X}$, be two pmf. Then $\Delta(p \,\|\, q) \geq 0$, with equality iff $p(x) = q(x)$ for all $x \in \mathcal{X}$.

In general the KL-div is unbounded from above, since we can find distributions that are arbitrarily close in total variation but with arbitrarily high relative entropy. Pinsker's inequality gives a lower bound on the relative entropy in terms of the total variation distance. It suggests that convergence in relative entropy, $\Delta(p \,\|\, q_n) \to 0$ as $n \to \infty$, where q_n is a sequence of rival distributions, implies convergence in the total variation ℓ_1 metric.

Consider two random variables X defined in a finite set $\mathcal{X}$ and Y defined in a finite set $\mathcal{Y}$ with marginal pmf $p(x)$ and $p(y)$, respectively; a joint pmf $p(x, y)$ and a conditional pmf $p(x \,|\, y)$. Similarly to the definition of the entropy of a single variable, we define the joint entropy of a pair of random variables.

Definition 3 (Joint Entropy): The joint entropy $H[X, Y]$ of two discrete variables X and Y with a joint pmf $p(x, y)$ is defined as

$$H[X, Y] \triangleq -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y) = -\mathbf{E}[\log p(X, Y)].$$

It represents the amount of info needed on average to determine the value of two random variables. Mutual information is the reduction in uncertainty of a random variable conditional on the knowledge of another random variable.

Definition 4 (Mutual Information): The mutual information between two random variables, X and Y, is the divergence of the product of their marginal distributions from their actual joint distribution:

$$I\,[X; Y] \triangleq \Delta(p(x, y) \,\|\, p(x)\, p(y)) = \mathop{\mathbf{E}}_{p(x,y)}\left[\log \frac{p(X, Y)}{p(X)\, p(Y)}\right].$$

The mutual information $I\,[X; Y]$ is symmetric in X and Y and always non-negative and is equal to zero if and only if X and Y are independent.

The following proposition provides a measure of information correlation between two random variables with the use of the conditional, joint entropy and mutual information.

Proposition 1 (A measure of information correlation): The quantity

$$\varrho(X, Y) = \frac{I\,[X; Y]}{H[X, Y]}, \quad (3)$$

is a measure of information correlation between the random variables X and Y, where $0 \leq \varrho \leq 1$. Furthermore, if X and Y are identically distributed, but not necessarily independent then the measure of information correlation is given by

$$\varrho(X, Y) = \frac{I\,[X; Y]}{H[X]}. \quad (4)$$

The measure of information correlation ϱ given by (3) or (4) is zero if and only if X and Y are independent, while ϱ is one if and only if X and Y have a one-to-one relationship.

Proof: Omitted due to space limitations. ■

III. PARSIMONIOUS TRAFFIC FLOW MODELS AND OPTIMAL SELECTION OF SENSORS

A. Motivation

Network-wide traffic models, like the so-called macroscopic fundamental diagram (MFD) of urban road networks, have been found to be particularly useful for monitoring traffic congestion in urban areas [12], [13], [14]. Recent simulation studies have confirmed that a sparse-measurement model [15], which involves a small number of sensors and corresponding measurements, can be used for the monitoring and perimeter control of congested urban areas [16]. In this direction, [17] proposed a quasi-optimal strategy for link selection. This work develops a rigorous framework for building sparse-measurement models, which are in principle less costly. These models should preserve the main features of a full-measurement model, e.g., capacity, critical density.

A full-measurement model can be constructed by flow-occupancy measurements of n inductive-loop detectors placed at appropriate network locations, a mid-block detector is usually placed in each link of the network. A sparse-measurement diagram can be then constructed by selecting only a number of $k < n$ detectors. Clearly, different levels of network coverage in terms of selected detectors k can provide different levels of accuracy. In principle, this is a combinatorial problem where the number of possible combinations is given by, $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. Depending on the size of the network and number of sensors, checking these combinations would be overwhelming and practical impossible. As an example, the San Francisco network (financial district and south of market) in [16] includes around $n = 400$ links. If $k = 100$ links selected (25%) then the number of sparse-measurement MFD to be examined is 2.24×10^{96}. Note that the total number of elementary particles in the universe is around 10^{80} (Eddington number).

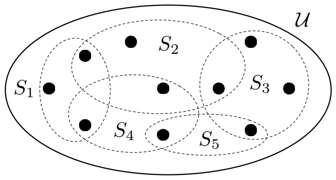


Fig. 1. The set covering problem.

To overcome this difficulty, this work proposes to formulate the network coverage problem as a sensor ("facility") location problem. The idea here is to select a subset of links from a given candidate set, and place in each of these links a "sensor" that will provide flow-occupancy measurements to construct the sparse-measurement traffic flow model. This is combinatorial optimization problem and can be formulated as an integer programming problem where a 0-1 decision variable associated with selecting any given link for sensor selection, at a given cost. This problem is NP-hard even with only polynomially many constraints, see e.g. [8].

B. Optimal Selection of Sensors

For the selection of sensors a set covering model is employed where the number of sensors p to be selected is not known in advance (unlike to p-center or p-median models). A formal definition of the set covering problem is as follows:

Definition 5 (Set Covering Problem): Let $\mathcal{U}$ be a finite set of cardinality n and let $\mathcal{S} = \{S_1, S_2, \ldots, S_m\}$ be a family of subsets of $\mathcal{U}$, whose union equals the universe, $\mathcal{U} = \bigcup_{j=1}^{m} S_j$. Find a minimum-cardinality subfamily $\mathcal{C} \subseteq \mathcal{S}$ that covers the universe set $\mathcal{U}$, i.e., the union of all sets in $\mathcal{C}$ is $\mathcal{U}$.

Clearly, such a cover exists if and only if the union of all sets in $\mathcal{S}$ is $\mathcal{U}$, and we assume this for the rest of the paper. For instance, a network with n sensors and universe $\mathcal{U} = \{1, 2, \ldots, n\}$ can be covered by $\mathcal{S} = \{S_1, S_2, \ldots, S_n\}$, where each $S_j = \{j\}$ is a singleton. Of course the cardinality of $\mathcal{S}$ in this case is n (i.e., maximum), given that all sensors are selected. On the other hand, multiple sensors can be assigned to a single set S_j if they are correlated and provide the same amount of information or coverage, and thus the cardinality of $\mathcal{S}$ can be accordingly reduced (see Fig. 1). The ultimate objective is to find a minimum-cardinality subfamily $\mathcal{C}$ of $\mathcal{S}$ that covers the universe set $\mathcal{U}$. The cardinality $k \le m$ of $\mathcal{C}$ is free and will be specified by the optimization.

The integer programming (IP) problem formulation reads:

$$
\begin{aligned}
\min_{\boldsymbol{\chi}} \quad & f(\boldsymbol{\chi}) = \sum_{j=1}^{m} w_j \chi_j \\
\text{subject to:} \quad & \sum_{j=1}^{m} c_{ij}\chi_j \ge 1, \quad \forall\, i \in \mathcal{U}, i \in S_j, \\
& \chi_j \in \{0,1\}, \quad \forall\, j \in \{1, \ldots, m\},
\end{aligned} \tag{5}
$$

where $\mathcal{U}$ is the universe set, w_j is the fixed cost of assigning a sensor to set S_j, c_{ij} is a binary covering constant that takes the value 1 if element of $i \in \mathcal{U}$ is covered by set S_j (within "distance" δ) and the value 0 otherwise. Finally, a variable χ_j is introduced for every set S_j, with the intended meaning that $\chi_j = 1$ when S_j is selected, and $\chi_j = 0$ otherwise. The solution of this problem $\boldsymbol{\chi}^*$ provides the optimal sensor selection (selected sensors) and the minimum-cardinality of $\mathcal{C}$ where $k \triangleq \operatorname{card}(\mathcal{C}) = f(\boldsymbol{\chi}^*)$, provided $w_j = 1, \forall\, j$.

To solve problem (5) the binary covering constants c_{ij} must be specified. This matrix can be constructed if the "distance" δ is appropriately defined. In the classic facility location problem, δ has the meaning of spatial distance. Here the spatial distance cannot be employed as a metric as explained in Section I. A measure of correlation based on the correlation or mutual information (see Section II) between random variables can be used as a "distance" metric to provide sufficient coverage and information accuracy.

Consider a transport network with n mid-block link sensors reporting flow and occupancy observations. Suppose that the time-occupancy data in each sensor is described by a discrete random variable $X_i \in \mathcal{X}$ (the time index is omitted for clarity), $i = 1, 2, \ldots, n$ with $\mathcal{X} = \{0, 1, 2, \ldots, 100\}$ the finite set of occupancy observations (0-100%). The main idea here is to look for the correlation of all pairs (X_i, X_j) for all $i, j = 1, 2, \ldots, n$, see (3), (4). High correlated random variables (with $\varrho \approx 1$) provide on average the same information (their expected value is the entropy (1)), and thus their measurements contribute in the same way in the construction of the traffic model. Therefore, it would be desirable the IP problem (5) to exclude a number of those detectors providing similar coverage. On the other hand, low correlated or independent random variables (with $\varrho \approx 0$) provide more information and their measurements are important for the construction of the traffic model. It is thus desirable, the IP problem (5) to include those detectors in the final solution. With these observations in mind a natural selection of "distance" is $\delta = \varrho$. For given δ, the binary covering matrix $\mathbf{C} = \left[c_{ij}\right]$ is given then by

$$
c_{ij} = \begin{cases} 1, & \text{if } d_{ij} \ge \delta, \\ 0, & \text{otherwise,} \end{cases} \qquad i, j = 1, 2, \ldots, n, \tag{6}
$$

with $d_{ij} = \varrho(X_i, Y_j)$.

Clearly if $\delta = 1$ has chosen then $\mathbf{C} = \mathbf{I}_n$ and the IP optimal solution is $f(\boldsymbol{\chi}^*) = n$, i.e., all sensors are selected $\chi_j = 1, \forall\, j$ (full-measurement model). If $\delta < 1$ then the optimization will exclude a number of sensors and a sparse-measurement model can be constructed from the selected sensors. Obviously, the accuracy of the traffic model approximation will be reduced (e.g., its shape will change) as δ decreased, but the optimal value of the objective function in (5) will be improved. This procedure reduces significantly the number of sensor locations that need to be checked, thus speeding up computation of the IP. Second, it provides an upper-bound on the value of the optimal selection (for $\delta = 1$), which can be used to bound the cost of other exact or heuristic approaches. To check the accuracy of the traffic model approximation for a particular δ (i.e., for a particular solution of (5)) we calculate its sparse-measurement pmf. We then compare (see KL-div) the sparse-measurement candidate pmf with an empirical (ground truth) reference pmf reflecting the full-measurement model.

C. Kullback-Leibler Divergence and Model Selection

To achieve this goal, we use the information-theoretic KL-div (2), to measure the dissimilarity or "distance" between pmf corresponding to different models obtained from the solution of the IP problem (5) for different δ. Entropy (1) reflects the average information included in our data set and gives a theoretical lower bound on the number of bits needed to encode a random variable X or its pmf $p_X(x)$. The KL-div reflects the average loss of using another code (or model $q(x)$) to encode a random variable X or its pmf $p(x)$. Certainly, we are interested in pmf or models that preserve the most information from our original data (i.e., from the empirical pmf). The relative entropy can also be interpreted as the information gain achieved about X if p can be used instead of q. Under certain regulatory conditions, KL-div is a monotonically decreasing function with information gain, while it is zero if and only if two distributions are equal (see Lemma 1). Therefore KL-div minimisation leads to information loss minimisation. On the other hand, the optimal value of the objective function in (5) is improved with information loss as δ is decreased while KL-div is increased. Note that KL-div is unbounded from above (see Section II). Concluding, efficient model selection (and sparse-measurement MFD approximation) is a trade-off between the KL-div and the cost of the cost function in (5).

IV. EMPIRICAL DATA & RESULTS

A. Data Description and Setup

Flow-occupancy experimental data (1.5-min samples) from 58 inductive-loop mid-block detectors and spanning one week in June 2006 [18], were available for the testing of the proposed data inference and model selection scheme. This data set was obtained in a field evaluation of the TUC/HYBRID signal control strategy with the semi-real-time strategy TASS developed by Siemens in the central business district (CBD) of Chania, Greece [18]. The CBD includes about 24 closely spaced signalised intersections and 71 links with varying lengths. It has been showed in [19] that the CBD exhibits a network-wide MFD that is reproduced under different traffic conditions (different days) but its shape and critical parameters (e.g., critical occupancy) depend on the applied signal control and the distribution of congestion in the network. Figs 2(a) and 2(c) depict the full-measurement MFD (circulating flow versus occupancy obtained from 58 sensors, indicated with black dots) for two representative weekdays (Monday and Friday). Each measurement point on the MFD corresponds to 1.5 min.

The full-measurement MFD of each day is approximated by a number of models obtained from the solution of the IP problem (5) for different $\delta \in \{0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 1\}$. These δ values were selected to investigate the effectiveness and accuracy of approximation whenever highly correlated data (in descending order) are excluded from the IP formulation (as reflected in $\mathbf{C}$ matrix). For each IP solution (δ value) the minimum-cardinality subfamily $\mathcal{C} \subseteq \mathcal{S}$ is obtained and then used to calculate the corresponding pmf q^{δ}.

TABLE I

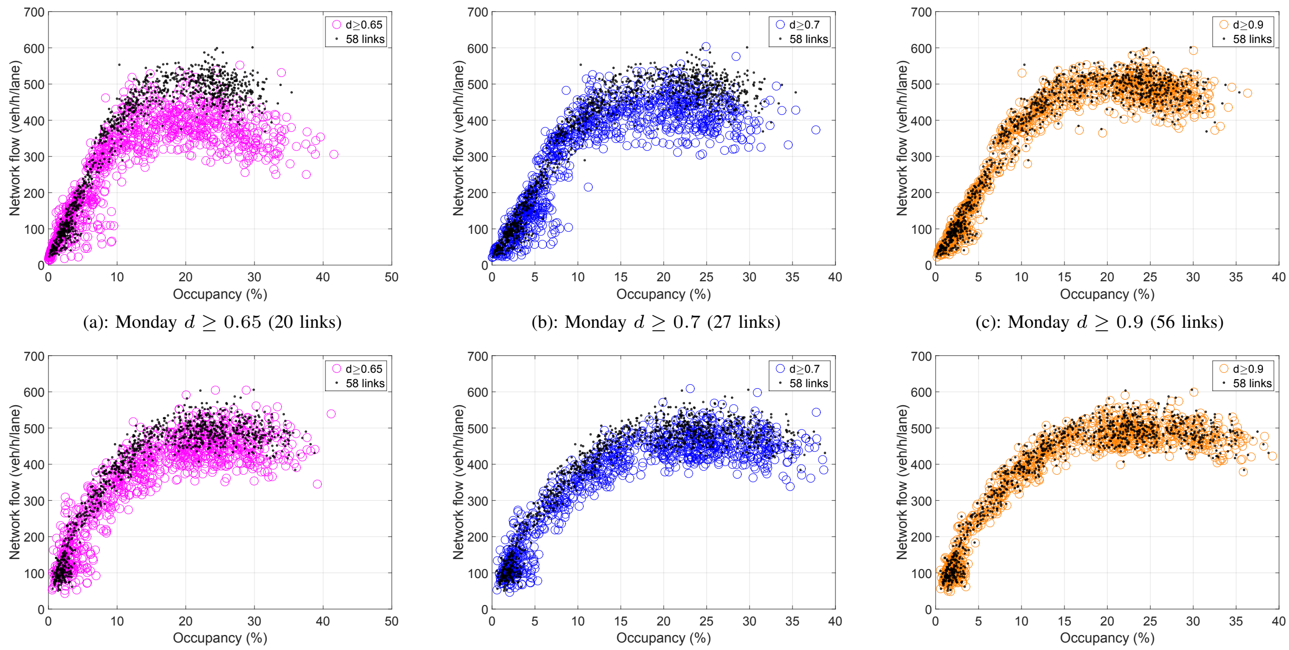
SOLUTION OF THE SET COVERING PROBLEM, ENTROPY AND KL-DIV FOR MONDAY & FRIDAY & DIFFERENT $d \geq \delta$ VALUES.

Day		"distance" ($d \geq \delta$)							
		$d \geq 0.6$	$d \geq 0.65$	$d \geq 0.7$	$d \geq 0.75$	$d \geq 0.8$	$d \geq 0.85$	$d \geq 0.9$	$d \geq 1.0$
Monday June 5, 2006	Total # of selected links	13	20	27	39	48	52	56	58
	$\mathcal{C}_{\text{IP}}$	{4, 8, 14, 17, 23, 24, 25, 30, 33, 37, 49, 52, 57}	{ 2, 3, 4, 7, 8, 13, 14, 17, 19, 23, 24, 30, 32, 33, 37, 38, 45, 52, 56, 57 }	{ 2, 3, 4, 7, 8, 11, 12, 13, 14, 15, 17, 19, 23, 24, 30, 32, 33, 37, 38, 39, 44, 48, 50, 52, 56, 57, 58 }	{ 1, 2, 3, 4, 7, 8, 10, 11, 12, 13, 14, 15, 17, 19, 21, 23, 24, 25, 27, 29, 30, 31, 32, 33, 35, 37, 38, 40, 44, 47, 48, 50, 52, 53, 54, 55, 56, 57, 58 }	{ 1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15 ,16, 18, 19, 21, 23 , 24, 26, 27, 28, 30, 31, 32, 33, 35, 36, 37, 38, 39, 40, 41, 43, 44, 45, 46, 47, 48, 49, 51, 52, 53, 54, 55, 56, 57, 58}	{ 1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 17, 18, 19, 21, 22, 23, 24, 25, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 51, 52, 53, 54, 55, 56, 57, 58 }	{1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58 }	{1, 2, …, 58}
	KL-div (bits)	0.038	0.038	0.022	0.019	0.0042	0.0015	0.00025	0
	Entropy (bits)	5.01	4.91	4.93	5.02	5.06	5.10	5.16	5.16
Friday June 9, 2006	Total # of selected links	20	27	36	43	50	55	56	58
	$\mathcal{C}_{\text{IP}}$	{2, 4, 8, 11, 14, 16, 17, 22, 23, 24, 25, 30, 32, 37, 38, 41, 44, 52, 57, 58}	{2, 3, 4, 7, 8, 10, 11, 12, 13, 14, 19, 21, 23, 24, 29, 30, 32, 33, 37, 38, 42, 44, 49, 52, 56, 57, 58}	{ 1, 2, 3, 4, 7, 8, 9, 11, 12, 13, 14, 15, 18, 19, 20, 22, 23, 24, 25, 29, 30, 32, 33, 37, 38, 40, 41, 42, 44, 45, 51, 52, 53, 56, 57, 58 }	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 17, 22, 23, 24, 26, 27, 29, 30, 32, 33, 35, 36, 37, 38, 40, 41, 42, 43, 44, 45, 46, 47, 48, 51, 52, 53, 54, 56, 57, 58, }	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 17, 18, 21, 22, 23, 24, 25, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 51, 52, 53, 54, 55, 56, 57, 58}	{ 1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58 }	{1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58}	{1, 2, …, 58}
	KL-div (bits)	0.022	0.012	0.0078	0.0059	0.003	0.0006	0.00027	0
	Entropy (bits)	5.49	5.29	5.35	5.25	5.25	5.3	5.34	5.34

Consider a random vector $\mathbf{X}$ of random variables $X_1, X_2, \ldots, X_k$ (time index is omitted) defined in a finite set $\mathcal{X} \in \{0, 1, 2, \ldots, 100\}$ (occupancy 0-100%), with $k \triangleq \text{card}(\mathcal{C}) = f(\boldsymbol{\chi}^*)$. The value $p_{\mathbf{X}}(x) = \mathbf{P}(X_k = x, \text{for all } k)$ is the probability that the random vector $\mathbf{X}$ takes the value x. Now for a given time-occupancy data set the pmf $p(x)$ is:

$$\mathbf{P}(X_k = x, \text{for all } k) = \frac{\phi_x + 1}{N + \sum_{i=0}^{N-1} \phi_i}, \tag{7}$$

with $N \triangleq \text{card}(\mathcal{X})$ and ϕ_x the frequency of observation x in the dataset. Eq. (7) is a smoothing technique to obtain a satisfactory probabilistic model in case of data sparsity (e.g., if many events x are unobserved $\phi_x = 0$). Note that



(a): Monday $d \geq 0.65$ (20 links)
(b): Monday $d \geq 0.7$ (27 links)
(c): Monday $d \geq 0.9$ (56 links)
(d): Friday $d \geq 0.65$ (27 links)
(e): Friday $d \geq 0.7$ (36 links)
(f): Friday $d \geq 0.9$ (56 links)

Fig. 2. Full-measurement (black dots) and sparse-measurement MFD approximations (pink and blue circles) for $\delta \in \{0.65, 0.7, 0.9\}$.

entropy and KL-div require $p(x), q(x) > 0$ and pmf that sum to 1. Given the pmf $q^{\delta}(x)$ for each IP solution (for each $\delta \in \{0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 1\}$), entropy and KL-div $\Delta(p \,\|\, q^{\delta})$ are then calculated to investigate the accuracy of different models. Here $p(x)$ is the empirical (ground truth) reference pmf reflecting the full MFD.

B. *Optimal Sensor Selection and Model Selection*

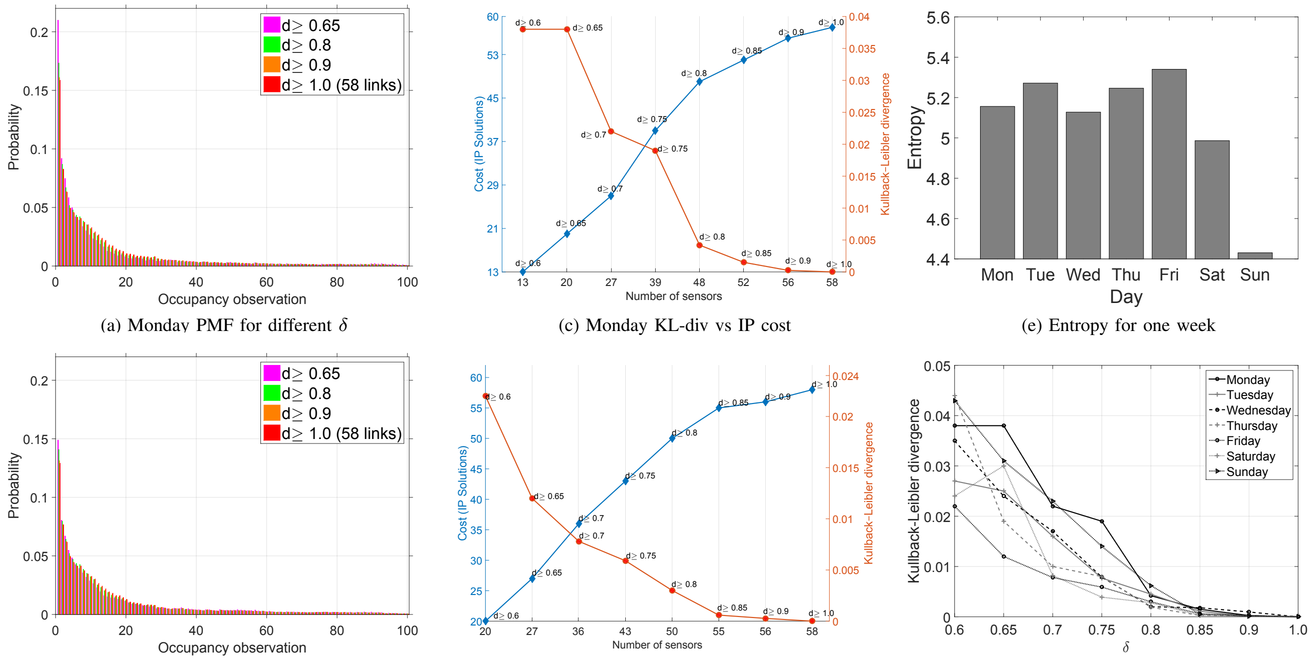
Table I displays the obtained results from the solution of the IP problem (5) for Monday and Tuesday and different $\delta \in \{0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 1\}$. As can be seen, the optimal value of the IP solution as reflected from the cardinality of $\mathcal{C}$ is improved (i.e., less sensors selected) as δ decreases. On the other hand, KL-div is increased as δ decreases reflecting the information loss under data sparsity. An important observation is that the optimal value of the IP solution is different from day to day for the same δ. For instance, $d \geq 0.7$ for Monday results 27 sensors (around 47%) while the same δ for Friday results 36 sensors (around 62%). Note that each IP solution in Table I except the optimal cost of sensor allocation (cardinality of $\mathcal{C}$) provides the subset of links (numbers in brackets) selected for sensor selection.

Figs 3(a) and 3(b) display the obtained pmf (models) for Monday and Friday and different $\delta \in \{0.65, 0.8, 0.9, 1\}$. The pmf for $\delta = 1$ (58 links) corresponds to the full-measurement MFD shown in Figs 2(a) and 2(d). All models indicate a geometric distribution with $\mathbf{P}(X = x) = (1-p)^x p$, $x = 0, 1, 2, \ldots$. Figs 3(c) and 3(d) depict the optimal cost of the IP problem versus KL-div for different δ on Monday and Friday. The primary vertical axis on the left is used for the optimal cost of the IP problem, whereas the secondary vertical axis on the right side is for the KL-div. These plots indicate the trade-off between sensor allocation costs and information loss or information gain. Figs 3(c) and 3(d) suggest that IP solutions for any $\delta \geq 0.75$ and $\delta \geq 0.7$ are acceptable on Monday and Friday, respectively. Models corresponding to $\delta = 0.75$ (39 sensors, Monday) and $\delta = 0.7$ (36 sensors, Friday) can be selected for the construction of the sparse-measurement MFD, since they provide the best trade-off between sensor allocation costs and info loss.

Figs 3(e) and 3(f) present collective results for all data sets spanning one week. Fig. 3(e) gives the entropy for each day if 58 sensors selected ($\delta = 1$), the theoretical lower bound on the number of bits needed to encode the pmf of each day. As can be seen, Monday and Wednesday (market is closed in the evening) indicate more or less the same lower bound. The same observation holds for Tuesday and Thursday (market is open in the evening). These two days are busier compared to Monday and Tuesday and thus more bits of information is needed to encode their models. Friday is usually congested and reveals high entropy.

Fig. 3(f) depicts the KL-div (dissimilarity between the pmf of the full MFD for $\delta = 1$ and rival models corresponding to different solutions of the IP program, i.e., different $\delta < 1$) trajectories in function of δ for seven days. This graph confirms that KL-div is monotonically decreasing (by definition) with respect to δ, it is nonnegative and zero if and only if two pmf are equal (case of $\delta = 1$). For $\delta = 0.6$ on Saturday KL-div is decreased (see $\delta = 0.65$) due to the discrete nature of the pmf and data sparsity. These trajectories also reflect the information loss induced by selecting δ smaller to 1.

Fig. 2 depicts the sparse-measurement MFDs (circle markers in colour) obtained for selected models (δ values, see pmf in Figs 3(a)–(b)) when contrasted with the empirical full-measurement (reference or ground truth) MFD (circle markers in black). As can be seen in Figs 2(a)–2(c), the model with $\delta = 0.9$ provides excellent approximation of the full-measurement MFD while approximation deteriorates



(a) Monday PMF for different δ (c) Monday KL-div vs IP cost (e) Entropy for one week

(b) Friday PMF for different δ (d) Friday KL-div vs IP cost (f) KL-div for one week

Fig. 3. (a)-(b): PMFs for different δ; (c)-(d): KL-div vs IP cost; (e) Entropy (in bits) for one week; (f) KL-div in function of δ for one week.

for models with $\delta < 0.75$ (cf. Fig. 3(c)). Similarly, models with $\delta \in \{0.7, 0.9\}$ provide excellent approximation of the full-measurement MFD of Friday (cf. Figs 2(d)–2(f) with Fig. 3(d)). Most of the models preserve the shape and main features of the full-measurement MFD. These results underline the satisfactory performance of the proposed information-theoretic framework in model selection and sparse-measurement MFD approximation.

V. CONCLUSIONS

This paper developed an information-theoretic framework for the efficient model selection and approximation of sparse-measurement MFDs. A measure of correlation between random variables is introduced as a "distance" metric to provide sufficient coverage and information accuracy. The KL-divergence was used to measure the dissimilarity between probability mass functions corresponding to different models obtained from the solution of a set covering problem. The proposed framework was evaluated with empirical loop-detector data of one week from a CBD with around sixty sensors. Results demonstrated that the obtained sparse-measurement rival diagrams were able to preserve the shape and main features of the full-measurement diagram. Therefore approximated MFDs, which are in principle less costly in terms of infrastructure requirements, can be used for the efficient monitoring and control of congested urban areas.

REFERENCES

[1] G. Laporte, S. Nickel, and F. Saldanha da Gama, Eds., *Location Science*. Cham, Switzerland: Springer, 2015.

[2] C. Diakaki, M. Papageorgiou, and K. Aboudolas, "A multivariable regulator approach to traffic-responsive network-wide signal control," *Control Engineering Practice*, vol. 10, pp. 183–195, 2002.

[3] A. Mazloumian, N. Geroliminis, and D. Helbing, "The spatial variability of vehicle densities as determinant of urban network capacity," *Philos. Trans. Royal Soc. A*, vol. 368, no. 1928, pp. 4627–4647, 2010.

[4] N. Geroliminis and J. Sun, "Properties of a well-defined macroscopic fundamental diagram for urban traffic," *Transp. Res Part B*, vol. 45, no. 3, pp. 605–617, 2011.

[5] C. Papadimitriou, "Optimal sensor placement methodology for parametric identification of structural systems," *Journal of Sound and Vibration*, vol. 278, no. 4, pp. 923–947, 2004.

[6] F. Zhao and L. Guibas, *Wireless Sensor Networks: An Information Processing Approach*. San Mateo, CA: Morgan Kaufmann, 2004.

[7] V. Gupta, T. H. Chung, B. Hassibi, and R. M. Murray, "On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage," *Automatica*, vol. 42, no. 2, pp. 251–260, 2006.

[8] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. Cambridge, MA: MIT Press, 2009.

[9] A. Krause, S. Ajit, and C. Guestrin, "Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies," *J. Mach. Learn. Res*, vol. 9, pp. 235–284, 2008.

[10] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, no. 27, pp. 379–423, 1948.

[11] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[12] H. Mahmassani, J. Williams, and R. Herman, "Performance of urban traffic networks," in *10th Int. Symp on Transport & Traffic Theory*, N. Gartner and N. Wilson, Eds., Amsterdam, The Netherlands, 1987.

[13] C. F. Daganzo, "Urban gridlock: Macroscopic modeling and mitigation approaches," *Transport Res Part B*, vol. 41, no. 1, pp. 49–62, 2007.

[14] N. Geroliminis and C. F. Daganzo, "Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings," *Transportation Research Part B*, vol. 42, no. 9, pp. 759–770, 2008.

[15] M. Keyvan-Ekbatani, M. Papageorgiou, and I. Papamichail, "Urban congestion gating control based on reduced operational network fundamental diagrams," *Transport Res Part C*, vol. 33, pp. 74–87, 2013.

[16] K. Aboudolas and N. Geroliminis, "Perimeter and boundary flow control in multi-reservoir heterogeneous networks," *Transportation Research Part B*, vol. 55, pp. 265–281, 2013.

[17] J. Ortigosa, M. Menendez, and H. Tapia, "Study on the number and location of measurement points for an MFD perimeter control scheme," *EURO J. Transport & Log*, vol. 3, no. 3, pp. 245–266, 2014.

[18] A. Kouvelas, K. Aboudolas, M. Papageorgiou, and E. Kosmatopoulos, "A hybrid strategy for real-time traffic signal control of urban road networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 3, pp. 884–894, 2011.

[19] K. Ampountolas and A. Kouvelas, "Real-time estimation of critical vehicle accumulation for maximum network throughput," in *2015 American Control Conference*, 2015, pp. 2057–2062.