

Robust Methods for Fusing Heterogeneous Spatiotemporal Data

By Charlotte L. Ellison, Zachary J. Roth, and Crystal Chen

INTRODUCTION: Modern technology has offered the capability to document and explore characteristics of human patterns of life. One such case is the abundance of time-indexed, ordered locations known as *spatiotemporal (ST) trajectories*. This increase in geospatial data, as well as other relevant contextual information, poses both opportunities and challenges to the Army geospatial community. This data has the potential to allow for gaining novel insight into human behavior, or opposite, previously developed methods may not be adequate for addressing the volume, sparsity, or complexity inherent in these data sets. New frameworks and methods for understanding the structure within the data will be key to providing additional insights into human behavior, which in turn, can expand the Army geo-intelligence-human-intelligence (GEOINT-HUMINT) spectrum of operations.

The analysis of an ST trajectory in isolation (i.e., in conjunction with no additional data) is able to identify only those characteristics that are represented in the trajectory information such as locations of interest, sequences of locations, and movement characteristics. Current methods of trajectory analysis can identify historical trends in locations or movements, but these methods lack the ability to detect interactions between human movement and the surrounding environment. In order to move beyond this limitation, data from multiple ST modalities, global positioning system (GPS) signals, social media interactions, recordings of events, and more, must be analyzed in unison. Each such ST modality captures distinct features that have the potential to reveal key insights during an analysis.

This work investigates the challenge of merging spatial trajectories with contextual features in a generalizable framework in order to understand the relationships between human movement and the surrounding environment. To do this, disparate ST datasets are merged together using a tensor-based representation that lends itself to higher dimensional data modeling and information extraction. This framework of data representation stores additional contextual information along with the spatial trajectories in question to allow for the detection of patterns that do not exist in the spatial trajectories alone.

In brief, this paper (1) proposes a method of modeling spatial trajectories, social interactions, events, and more via binary tensors, and (2) applies a tensor decomposition to detect key features and discover hidden correlations within the data. This new framework will allow for greater understanding of human movement by revealing latent features in the data.

BACKGROUND AND RELATED WORK: There are several sources of data that contain information related to human movement. The most obviously relevant example is the movement of an object through a geographic space captured as a time-indexed, chronologically ordered set of location points called a *spatial trajectory*.



Additional examples include data about weather patterns, events (e.g., conferences and traffic incidents), and social interactions (e.g., posts on social media). These data sets can often provide unifying information about the movement observed in individual spatial trajectories. Though highly disparate, these auxiliary sources and ST trajectories all share spatial and temporal dimensions. As a result, the different data modalities can be merged along those common ST dimensions. The method of merging produces an object referred to as a tensor, and the method of analysis involves tensor decomposition.

In the following section, tensors and related terminology are formally defined, a particular tensor decomposition referred to as CP decomposition is introduced, and some related and motivating applications of tensors and their decompositions are reviewed. The remainder of this section is based primarily on the excellent review paper by Kolda and Bader (2006).

Tensor Fundamentals. For the purposes of this technical note (TN), a tensor is a multidimensional array. Mathematically, a tensor \mathcal{X} is represented by an element of the vector space $F^{I_1 \times I_2 \times \dots \times I_N}$ for field F and positive integers N and I_1, \dots, I_N . In this case, the *order* of \mathcal{X} is N ; and \mathcal{X} is called an *order- N* tensor. For example, a vector is an order-1 tensor and a matrix is an order-2 tensor. Each array dimension¹ of the tensor is referred to as a tensor *mode*. A tensor is *rank-one* if it can be expressed as the outer product of vectors, with each vector corresponding to a single mode of the tensor; that is, a tensor \mathcal{X} is rank-one if there are vectors $v_1, \dots, v_N \in I_N$ such that $\mathcal{X} = v_1 \circ \dots \circ v_N$, where the operator denotes the outer product. For example, if \mathcal{X} is a rank-one, order-three tensor, then there are vectors $\mathbf{u} \in I_1$, $\mathbf{v} \in I_2$, and $\mathbf{w} \in I_3$ such that $\mathcal{X} = \mathbf{u} \circ \mathbf{v} \circ \mathbf{w}$. Trivially, any tensor can be decomposed as the sum of rank-one tensors; however, a single tensor may have several decompositions into sums of rank-one tensors. The *rank* of a tensor \mathcal{X} , written as $\text{rank}(\mathcal{X})$, is the minimum number of rank-one tensors required in any such decomposition of \mathcal{X} . The decomposition of an order-3 tensor into a sum of rank-one tensors is depicted in Figure 1. Such rank-one decompositions are of interest because each rank-one tensor from the decomposition contains structured information about the original tensor and, as such, is more useful for analysis and interpretation.

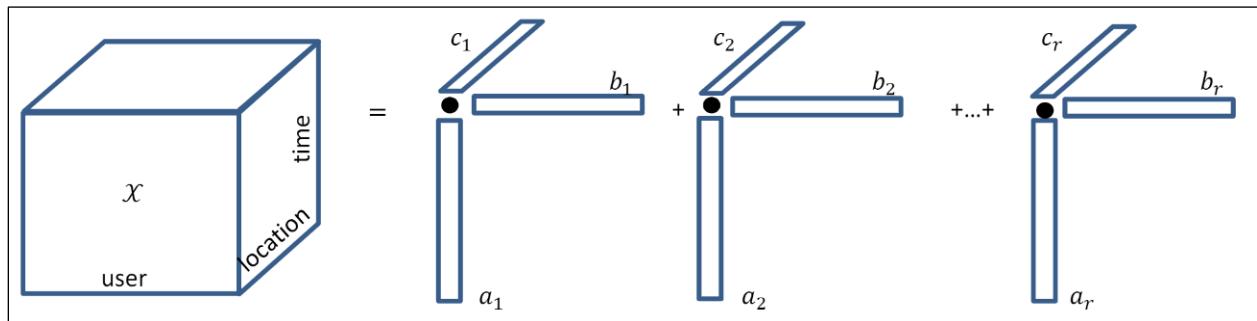


Figure 1. Illustration of rank-one decomposition of an order-three tensor.

The problem of determining the rank of a tensor is very computationally intensive (Kolda and Bader 2006), however, algorithms exist to compute decompositions that are approximately equal to the input tensor. One such algorithm is *CANDECOMP/PARAFAC (CP) decomposition*, which

¹ The term *array dimension* is used to distinguish from the dimension of the vector space.

approximates a tensor \mathcal{X} by a rank- R tensor for some pre-determined rank. This algorithm produces a list of *factor matrices* $X_i \in F^{I_i \times R}$ with $i \in \{1, \dots, n\}$ (i.e., one factor matrix for each modality). Given any matrix A , let $\text{col}_i(A)$ denote column i of A . Then the approximation $\widehat{\mathcal{X}}$ of \mathcal{X} produced by the decomposition is given by

$$\widehat{\mathcal{X}} = \sum_{j=1}^R \text{col}_j(X_1) \circ \cdots \circ \text{col}_j(X_n) = X_1 \circ \cdots \circ X_N \quad (1)$$

where $\text{col}_j(X_1) \circ \cdots \circ \text{col}_j(X_n)$ is a rank-one component for each j and \circ is an outer product. The trade-off between rank and approximation error means CP decomposition can be used to get a less accurate, but easier to understand decomposition, or a higher rank more faithful reconstruction.

Related Works. Existing work on tensor analysis predominately comes from the signal processing, acoustics, and chemometrics domains, where tensors offer the ability to extract significant features from high-dimensional, noisy data (Kolda and Bader 2006). Tensors naturally support the integration of multiple dimensions and, thus, are well-suited to manage complex data while still maintaining interpretability. In the geospatial domain, some work has begun to apply tensors to spatial datasets for the purpose of data imputation (Li et al. 2014; Li et al 2012). A literature review reveals there has not been research into applying tensors for representing moving object (MO) data together with contextual information to discover potential relationships between external features and MO behaviors (Kolda and Bader 2006).

A large amount of existing trajectory work attempts to predict future paths, make location/route recommendations, detect anomalous routes, or identify commonly visited areas of interest. Current spatial trajectory research mainly focuses on movement pattern extraction rather than pattern understanding. For instance, Li et al. (2012) developed an approach to discover periodic behaviors in animal movements using Fourier Transform and autocorrelation; however, their approach does not handle the application of more variable movements that occur in sporadic temporal windows, which limits the types of patterns they can discover. Bachmann et al. (2013) use density or frequency analysis for the purpose of trajectory prediction, while others employ historical trend information for anomaly detection (Fanaswala and Krishnamurthy. 2013). Many of these existing works focus on intra-trajectory analysis, studying the features of single trajectories with a smaller cohort of research conducted on inter-trajectory relationships or simultaneous feature discovery. Vespier et al. (2013) studied the simultaneous impact of various attributes by performing multi-scale analysis on moving-object and time-series data. Li et al. (2013) studies MO data where objects follow one another by considering sequence alignment and similarity. What is lacking in these studies is the consideration of the impact of multiple external features on human mobility, limiting their ability to accurately explain patterns and relationships between MO and their environment.

Other works on spatial trajectories have begun to consider the addition of external data attributes to enhance MO recognition. Most of these works merge the datasets in an ad-hoc manner that are not generalizable for merging with other ST data (Wang et al. 2016; Wang et al. 2014; Zhang et al. 2015; Zheng et al. 2012). For instance, Wang et al. (2016) developed an approach to predict crime rates based on point-of-interest (POI) attributes and “taxi flow.” Due to the specificity of the approach, the model is not easily generalizable to incorporate a wider range of urban data features. Spatial trajectories and their combination with other data classes can be achieved by modeling them as graphs, matrices, and tensors to simplify, manage, and extract patterns. Others have

performed map-matching to connect trajectories to an underlying road network, which can then be modeled as a directed weighted graph (Pan et al. 2013; Zheng et al. 2012; Zheng and Xie 2011). Analyzing graph segment traversals or assessing graph connectivity allows for the determination of route predictions, recommendations, and anomalies. Matrix-based trajectory representation has been primarily used to recommend routes or discover anomalies via Collaborative Filtering and other factorization approaches (Zheng et al. 2010; Chawla et al. 2012). In both cases of graphs and matrices, associating multiple data features led to a cumbersome model that hinders interpretability and effectiveness.

Unlike graphs and matrices, tensors naturally support the integration of multiple dimensions. However, the few works that do utilize tensors for trajectories do so exclusively in order to perform data imputation (Li et al. 2016; Wang et al. 2014). These works consider contextual information involving road network statistics and POIs but do not handle the sudden occurrence of external events, especially in the absence of historical data. Zheng et al. (2014) developed work most closely related to the proposed research, the authors merge user complaint data, social media information, POIs, and road network data to improve their understanding of noise pollution in New York City. However, their approach does not effectively make use of the additional ST information that is embedded with trajectory data such as motifs. Furthermore, the authors do not exploit the tensor decomposition results to understand or discover correlations between external features and MO behaviors. Rather, the tensors and additional contextual information are utilized to interpolate missing data. Currently, no existing works have utilized tensors to discover correlations between multimodal ST datasets and trajectories to understand the impact of features on human movement and vice-versa.

The tensor framework for data integration that are introduced in the following section allows for joint analysis of a variety of ST datasets to improve the understanding of trajectory movements and the external attributes that impact them.

METHODS: The approach described in this TN uses tensors to model the trajectories and contextual data in question. These various pieces of data are often referred to as data modalities.

For the purposes of this document, a *data modality* is any finite set.¹ Before getting to formal notation, consider the modalities $M_1 = \{\text{Fred, Barney}\} = \{F, B\}$, $M_2 = \{\text{Thursday, Friday, Saturday}\} = \{t, f, s\}$, and $M_3 = \{\text{quarry, lodge}\} = \{q, l\}$. Then the data set $\mathcal{D} = \{(F, t, q), (F, s, l), (B, f, q), (B, s, l)\}$ show that Fred was at the quarry on Thursday, that Barney was at the quarry on Friday, and that both Fred and Barney were at the lodge on Saturday.

To set some notation, let M_1, \dots, M_N be a list of distinct data modalities with $M_k = \{m_1^{(k)}, \dots, m_{I_k}^{(k)}\}$ for each $k = 1, \dots, N$. Note that the cardinality of M_k is I_k for each k . With this convention, a (*binary*) *data set* \mathcal{D} is simply a subset of the Cartesian product $M_1 \times \dots \times M_N$; and $(d_1, \dots, d_N) \in \mathcal{D}$ indicates co-occurrence of (d_1, \dots, d_N) . Given a data set $\mathcal{D} \subseteq M_1 \times \dots \times M_N$, the *induced tensor* is defined $\mathcal{X}_{\mathcal{D}} = [x_{(i_1, \dots, i_N)}] \in \{0, 1\}^{I_1 \times \dots \times I_N}$ by $x_{(i_1, \dots, i_N)} = 1$ if $(m_{i_1}^{(k)}, \dots, m_{i_k}^{(k)}) \in$

¹ Many modalities may be infinite in nature, such as is the case with time and location. However, to fit into this framework, those modalities must be discretized.

\mathcal{D} and $x_{(i_1, \dots, i_N)} = 0$ otherwise.¹ In particular, note that modality M_k of a data set \mathcal{D} corresponds to mode k of the induced tensor $\mathcal{X}_{\mathcal{D}}$.

Once the induced tensor of a data set is created, CP decomposition is applied to discover relationships among the features in the datasets. Some additional details about the data and about CP decomposition are given in Section ‘CP Decomposition’ and Section ‘Experimental and Computational Details’, respectively.

CP Decomposition. Once the spatial trajectories and contextual features are represented as a single tensor \mathcal{X} , CP decomposition can be applied to create an approximation $\widehat{\mathcal{X}}$ of \mathcal{X} that captures latent features in the data. Specifically, CP decomposition solves the following minimization problem:

$$\min_{\widehat{\mathcal{X}}} \|\mathcal{X} - \widehat{\mathcal{X}}\|_F^2, \quad (2)$$

where $\widehat{\mathcal{X}}$ is defined from the factor matrices as shown in Equation 1 and where $\|\cdot\|_F^2$ is the Frobenius norm of a tensor (i.e., the square root of the sum of the squares of all the elements in the tensor). A popular method for solving CP decomposition is Alternating Least Squares (ALS), which requires an initial (usually random) guess for the factor matrices X_i and a user-defined number R of component rank-one tensors. ALS iteratively solves simpler sub-problems of Equation 2 by fixing all but one factor matrix in order to constrain the search space. That is, for each factor matrix X_i in turn, ALS iteratively calculates an improved estimate of \mathcal{X} .

The *reconstruction error* of a non-zero tensor \mathcal{X} by another tensor $\widehat{\mathcal{X}}$ is defined to be Frobenius norm of their difference normalized by the magnitude of \mathcal{X} :

$$\text{Frob}_{\mathcal{X}}(\widehat{\mathcal{X}}) = \frac{\|\mathcal{X} - \widehat{\mathcal{X}}\|_F^2}{\|\mathcal{X}\|_F^2}. \quad (3)$$

For a fixed \mathcal{X} , minimizing $\text{Frob}_{\mathcal{X}}(\widehat{\mathcal{X}})$ is equivalent to the optimization problem shown in Equation 2; however, the reconstruction error provides a measurement relative to the data being approximated.

The tensor used in this is binary (see Section ‘Experimental and Computational Details’). As such, a variation of CP decomposition that restricts the entries of the resultant factor matrices to be non-negative was chosen. Specifically, the non-negative CP decomposition used here optimizes the objective function from Equation 2 subject to the restriction $X_i \in \mathbb{R}_+^{I_i \times R}$, where \mathbb{R}_+ is the set of non-negative real numbers Xu and Yin (2013). Imposing this restriction was meant to produce results that were more-easily interpretable. The non-negative summands produced by this method ensure that each element contributes positively (if at all) and does not allow for portions of multiple

¹ These definitions can be extended to non-binary data sets over a field F . In this case, a data set is a map $f: M_1 \times \dots \times M_N \rightarrow F$; and the induced tensor is defined by $x_{(i_1, \dots, i_N)} = f(m_{i_1}^{(1)}, \dots, m_{i_N}^{(N)})$.

rank-one tensors to cancel each other out, as could happen with the unrestricted version of CP decomposition.

When decomposed, each factor matrix captures the defining features of each mode. Any information not captured in the factor matrices is considered insignificant for explaining the moving object observations and is discarded from further investigation. The number of component rank-one tensors used to reconstruct the original tensor is a user-defined parameter, dependent on the level of resolution required for the application. A higher number of component rank-one tensors produces an improved tensor reconstruction extraction of more detailed patterns, while a reduced number of components allows for quicker computation and extraction of more obvious and frequently occurring patterns.

EXPERIMENTAL AND COMPUTATIONAL DETAILS: For this work, a synthetic data set was created to investigate the effectiveness of CP decomposition at uncovering patterns embedded in data involving spatial trajectories and other contextual information. This data set contains trajectory information for five users and, additionally, information about the activities performed by each user and the social interactions between these users. As such, the tensor induced by this data set contains five modes corresponding to information about the user, location, time, activity, and social interactions. This data serves to illustrate the capabilities of the proposed methodologies. As is common with synthetic data, the primary purpose of this data is not to perfectly represent reality, but rather provide a simple data set for experimentation with the proposed methodologies.

In the synthetic data, each user moves independently of all other users through a 10-by-10 grid, and each user is active for some portion of a fixed time window (composed of 120 time steps). They move more-or-less randomly with the following modification: Portions of each user's trajectory may repeat during the given time period. Indeed, this is the case for all but user 5. The basic structure of these trajectories is summarized here

- User 1 is active during time steps 1–69. The trajectory during time steps 1–23 is repeated during time steps 24–46 and 47–69.
- User 2 is active during time steps 1–120. The trajectory during time steps 1–30 is repeated during time steps 31–60, 61–90, and 91–120.
- User 3 is active during time steps 1–57. The trajectory during time steps 28–37 is repeated during time steps 38–47 and 48–57.
- User 4 is active during time steps 1–65. The trajectory during time steps 26–35 is repeated during time steps 36–45, 46–55, and 56–65.
- User 5 is active during time steps 1–34. No portion of the trajectory is repeated.

Both location and time information are unitless in this data set.

This leaves only the following two modalities to be described: activity, and social interactions. Each user performs exactly one activity for the entire time that the user is active. Users 1, 3, and 5 perform activity 1; users 2 and 4 perform activity 2. For example, user 1 performs activity 1 during

time steps 1–69 and no activity outside of those time steps. Similarly, each user has social interactions with a fixed set of other individuals for the entire duration of the user's trajectory regardless of the duration of the other individuals' trajectories. Specifically, users 1, 3, and 5 interact only with each other; and users 2 and 4 interact only with each other. The use of this synthetic data set allowed for a comparison of the results of CP decomposition to the patterns that were explicitly embedded in the data. The following section describes several patterns observed in the output.

RESULTS AND DISCUSSION: The results of CP decomposition on the data set can be examined from several different perspectives. Looking at the reconstruction error as a function of the number of rank-one tensors used shows that increasing the number of rank-one tensors improves the approximation. In order to understand what is inside the approximations themselves, it is necessary to consider what a rank-one tensor can and cannot represent. With this in mind, it is then possible to find and interpret the underlying patterns preserved in the factor matrices. Throughout this section, \mathcal{X} denotes the tensor induced by the synthetic data described in the previous section.

Reconstruction Error. Since CP decomposition is designed to approximate a tensor by another tensor of a pre-determined rank, the analysis begins by quantifying the ability of CP to approximate the synthetic data. Figure 2 shows the average reconstruction error of approximations of \mathcal{X} produced by CP decomposition. Note that the rank of any tensor is at most the number of nonzero entries in that tensor, which is 505 for the induced tensor \mathcal{X} in this case study. Thus, in theory, a CP decomposition of rank 505 should perfectly reconstruct the data. This provides the reasoning for the maximum rank used for CP in Figure 2. As can be seen in that figure, the reconstruction error approaches, but never quite reaches, zero, as the rank approaches that upper bound. This is not surprising since the algorithm used by CP decomposition provides no guarantee that a global minimum will be attained and can be affected by the initial conditions. Additionally, a tighter upper bound on the tensor rank can be computed by counting the number of unique locations visited by each user and summing them together. Doing this, it can be found that the rank of the tensor is at most 159 (with users 1–5 contributing at most 23, 30, 37, 35, and 34, respectively, to the total rank). Indeed, Figure 2 shows that the lowest reconstruction error reaches nearly zero between rank 159 and rank 505.

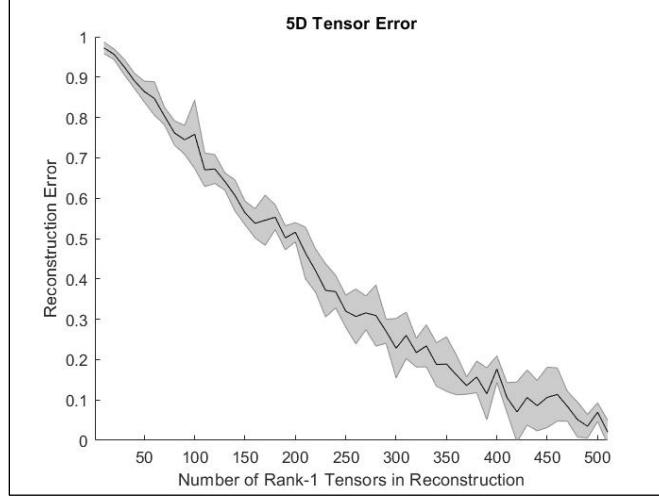


Figure 2. Reconstruction error. For each rank in $\{10, 20, \dots, 510\}$, the reconstruction error of each of ten CP decompositions of the induced tensor \mathcal{X} is computed. Each CP decomposition was initialized with random input of the size determined by the rank. The central black line shows the mean of the reconstruction error produced by CP decomposition on the ten initializations for a given rank. The surrounding ribbon indicates one standard deviation from the mean.

Figure 3 further illustrates the reconstruction error by showing some samples of how portions of the data are reconstructed for various ranks. In this figure, the original tensor along with approximations of rank 50 and 200 are considered. Although aspects of the data are represented in each reconstruction, this figure demonstrates how certain features may be lost when reconstruction error is high or, equivalently, when the number of rank-one tensors used in the decomposition is low. For example, the trajectory of user 3 within the first 27 time points is more accurately captured by the rank-200 reconstruction than the rank-50 reconstruction. In effect, because no repetitions occur for those locations, that portion of the trajectory seems to be regarded as noise to the rank-50 decomposition. A similar situation occurs with the entire set of data for user 5.

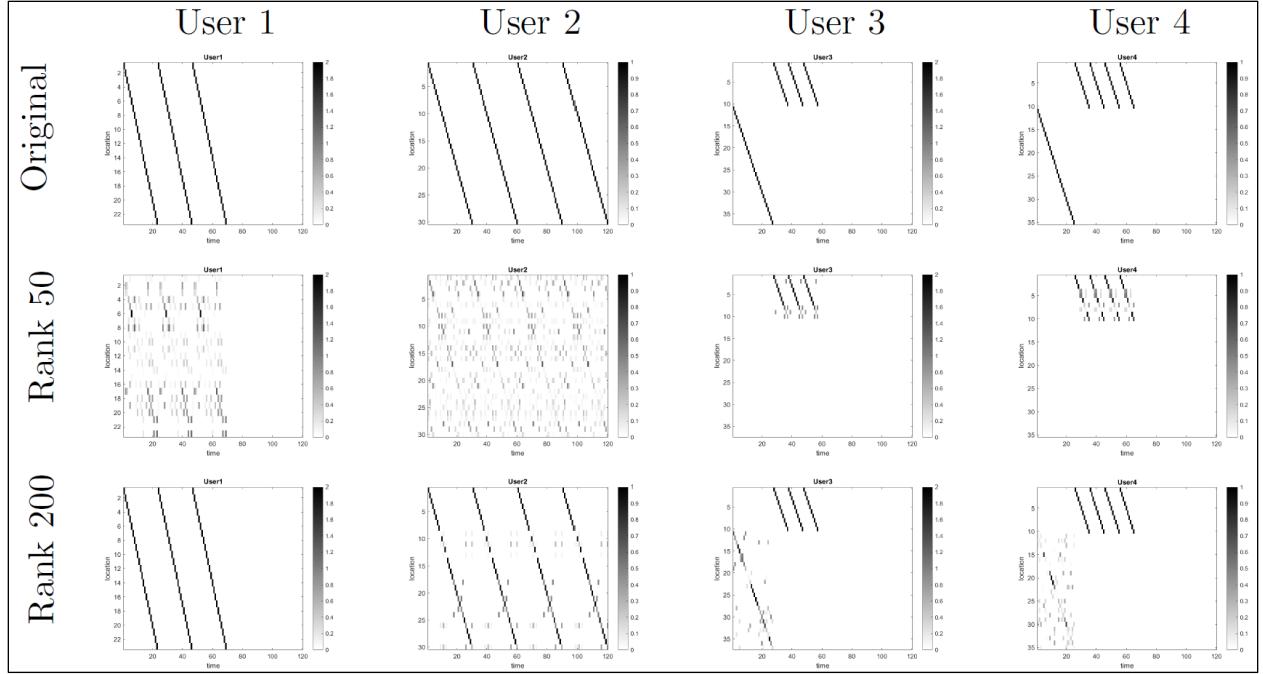


Figure 3. Trajectories and reconstructions (alternate format). This shows visually how accurately trajectories are reconstructed in the synthetic data set. Each column corresponds to an individual user. Each row corresponds to a different tensor: the original tensor (top row), a rank-50 approximation (center row), and a rank-200 approximation (bottom row). The reconstruction error of the approximation with 50 components is 0.850, and the reconstruction error of the approximation with 200 components is 0.429. Each matrix created by summing the sub-tensor corresponding to the user over the activity and connected-user dimensions, leaving only the time and location dimensions as indicated. Overall, it can be easily seen that the reconstruction from the rank-200 tensor is much more accurate than that for the rank-50 tensor. Notice that repeated trajectories seem, on the whole, to be represented strongly in the reconstructions. In contrast to this, any portion of a trajectory that is not repeated seems to be considered as noise in the rank-50 reconstruction (and is still noisy in the rank-200 reconstruction).

The matrices in the center and bottom rows of Figure 3 also demonstrate an important characteristic of the approximations produced by CP decomposition: The reconstructed tensors contain real values despite the binary nature of the original tensor. This makes sense from the perspective of minimizing the Frobenius norm. For example, for the matrix $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$, the smallest reconstruction error obtainable by any binary, rank-one matrix is 1 (obtained by the all-ones matrix), whereas the real-valued, rank-one matrix $\frac{3}{4} \cdot \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, obtains a smaller reconstruction error of $\frac{3}{4}$. However, this characteristic makes interpretation of the reconstructed tensors less straightforward. In particular, although the top row of Figure 3 shows a trajectory for each user (i.e., at most one location per time point), the reconstructed tensors (center and bottom rows) approximate trajectories since several points in time (i.e., matrix columns) have multiple nonzero location values.

Rank-one blocks. To make sense of the results of CP decomposition, one must have an understanding of the sorts of patterns that can be detected by CP decomposition. Fundamentally, the task that CP decomposition is trying to perform is approximation by a sum of rank-one tensors.

This constrains the idea of *pattern* from the perspective of CP decomposition: A pattern is a rank-one tensor. To gain insight into what such a pattern conveys about the underlying data, a binary (0 or 1), rank-one tensor is considered. Such a tensor contains a unique maximal sub-tensor in which all entries are 1 and outside of which all entries are 0. Two examples of such rank-one matrices (i.e., rank-one, order-2 tensors), unrelated to the data described in the rest of the paper, are

$$\begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{matrix} \text{ and } \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}, \end{matrix}$$

which can be decomposed, respectively, as

$$\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \circ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \circ \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

Note that these two matrices are equivalent up to permutation of rows and columns. If rows represent users and if columns represent locations, then the tensor on the left indicates that users 1 and 2 are identical within locations 1 through 3.

The generalization of binary, rank-one tensors to real-valued, rank-one tensors loosens the idea of being “identical” from the perspective of the decomposition; but not all of the notion of similarity is lost. Another example of a rank-one tensor for illustrative purposes, again with rows representing users and columns representing locations, containing real numbers is

$$\begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 0.1 & 0.4 & 1.2 & 0 & 0 \\ 0.5 & 2.0 & 6.0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 5 \\ 0 \end{pmatrix} \circ \begin{pmatrix} 0.1 \\ 0.4 \\ 1.2 \\ 0 \\ 0 \end{pmatrix}. \end{matrix}$$

In this example, each location visited by user 1 (represented by row 1) seems to be visited exactly five times as often by user 2. Similarly, locations 1, 2, and 3 are visited by users 1 and 2 in similar ways. This might indicate for instance, that location 2 is visited four times as often as location 1 by both users 1 and 2. Regardless of the perspective, the rank-one tensor in question indicates that users 1 and 2 visit the same locations *with the same relative frequencies*, but that user 2 visits those locations more often. In this way, even when considering real-valued CP decompositions, rank-one tensors seem to indicate a sense of being identical up to some scaling. For CP decomposition, this leads to the perspective that a pattern in \mathcal{X} is a collection $\{C_i \subseteq M_i | i = 1, \dots, N\}$ of subsets of

modalities such that all elements of each C_i are highly similar to each other when considered over the restricted range $\prod_{j \neq i} C_j$ of other modalities.

Scope of Rank-One Approximation. While CP decomposition is useful in uncovering the patterns described here, there are several other types of patterns for which CP decomposition is less successful. Broadly speaking, patterns that appear in the tensor as small blocks of the form described in the previous section will not affect the cost function that CP decomposition is attempting to minimize as much as those patterns that exist as large blocks will affect the cost function. As such, the sorts of patterns that manifest as small blocks will often not be well-represented by a CP decomposition of low rank. Hence, it behooves an individual to understand whether a particular pattern of interest will manifest itself as a large coherent block in an induced tensor.

For example, infrequently visited locations may result in relatively small blocks (as observed in Figure 3). In some cases, this may be a desirable feature, eliminating entries in the reconstructed tensor that may well be considered to be noise. In this way, the low-rank reconstruction performed by CP decomposition can be considered to perform a type of noise reduction. On the other hand, such infrequently visited locations in another data set may be exactly the features of interest. In this case, CP decomposition can provide a reference from which such deviations can be more easily identified.

Another pattern that will not form large blocks in the induced tensor is that of one person who is following another. The delay between the users could produce an offset in the spatial mode of the tensor, which would not produce the larger blocks for which CP decomposition searches. Such an offset could also occur in the activity mode of the induced tensor if two users are performing different activities during a common trajectory. Again, this would result in a representation ill-suited for CP decomposition. In this way, the addition of any modality for which at most one element is possible at a given point in time will result in a sparsification of the data into a higher-dimensional tensor. In other words, even though a researcher may be primarily focused on trajectories, each modality in a tensor serves to segment the trajectories in unintuitive ways. In particular, this means that the addition of more modalities must be done with care for a researcher to leverage the tensor representation and CP decomposition most effectively.

Regardless, an individual must understand the data and the way that its characteristics are represented in the induced tensor to appropriately use CP decomposition and interpret its output. For example, with the construction used for the synthetic data set, a user who performs many activities may be over-represented in a CP decomposition relative to a user who only performs one activity; in this case, the user performing only one activity may be thought of as noise. Decomposition of such a scenario could be misleading if the tensor representation is not properly understood.

Factor Matrices and Correlations. The output of CP decomposition can also be considered as a list of factor matrices whose columns can be used to reconstruct the rank-one tensors used in the approximation. Example factor matrices are shown in Figure 4. While these factor matrices contain all of the information contained in the corresponding rank-one tensors, they significantly simplify the representation and provide immediate verification that the corresponding tensors are rank-one. Hence, the factor matrices allow for another layer in which to consider and search for patterns.

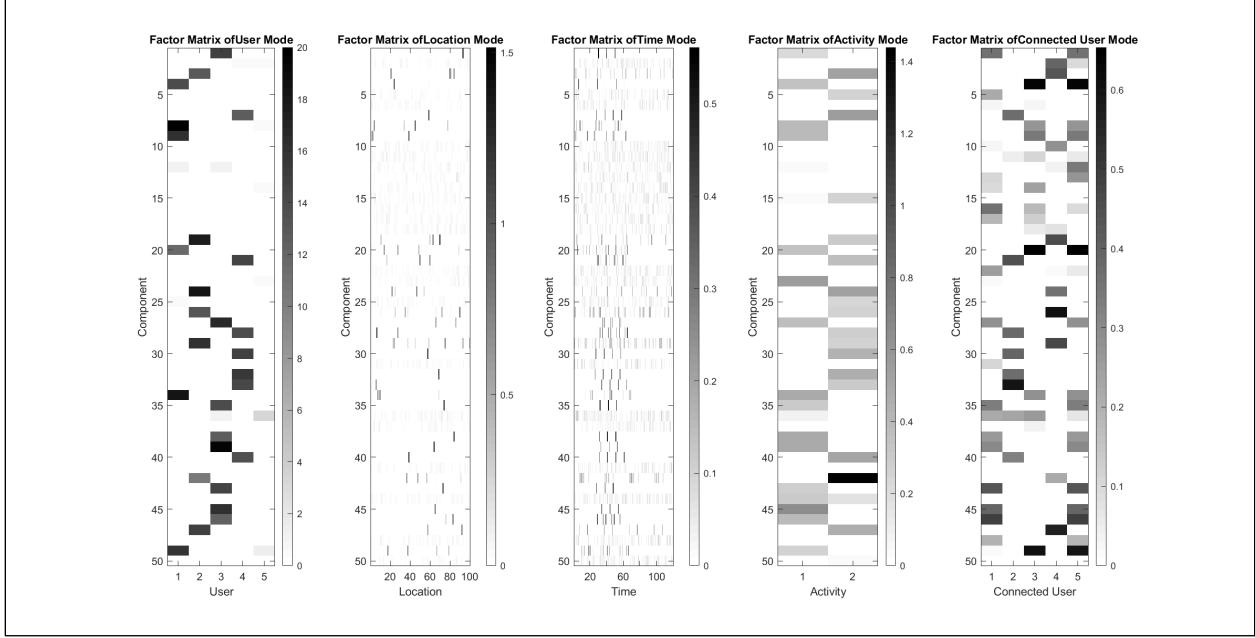


Figure 4. Factor matrices produced by CP decomposition. The decomposition was performed on the original tensor with rank $R = 50$. Note that the transpose of the factor matrices are shown here for ease of visualization, leading to the horizontal axis being the elements of each mode and the vertical axis being the component number. The i th rank-one component tensor can be computed by taking the outer product of the i th row of each (transposed) factor matrix shown here. The reconstructed approximation of the original tensor calculated by CP decomposition is the summation of the 50 rank-one component tensors.

In particular, recall that each factor matrix corresponds to a single, distinct mode/modality in the induced tensor. This means the modality name can be used to refer to each corresponding factor matrix. Further, each factor matrix contains all of the information captured about the corresponding modality. So, for instance, all information captured by the decomposition about a given location is represented in the corresponding row of the location factor matrix (column in transposed factor matrix visible in Figure 4). This observation allows for direct comparison between arbitrary elements of any single modality by computing the correlation between the corresponding columns of the factor matrix. For example, in Figure 5 there are red diagonal lines between times 26 and 56, suggesting points of high correlation. These diagonals indicate that times i and $i \pm 10, 20, 30$ for $i \in [26, 56]$ are highly correlated. This matches the periodicity in the trajectories of users 3 and 4, where both of the users are repeating the same movements (as well as activity and connected user) between approximately those times. Fainter diagonal lines can be seen in the bottom right and top right areas of the correlation matrix. These entries correspond to the periodicity in the trajectory of user 2. Looking at the correlation between the entities in a given mode's factor matrix can illustrate which entities are similar in the original data.

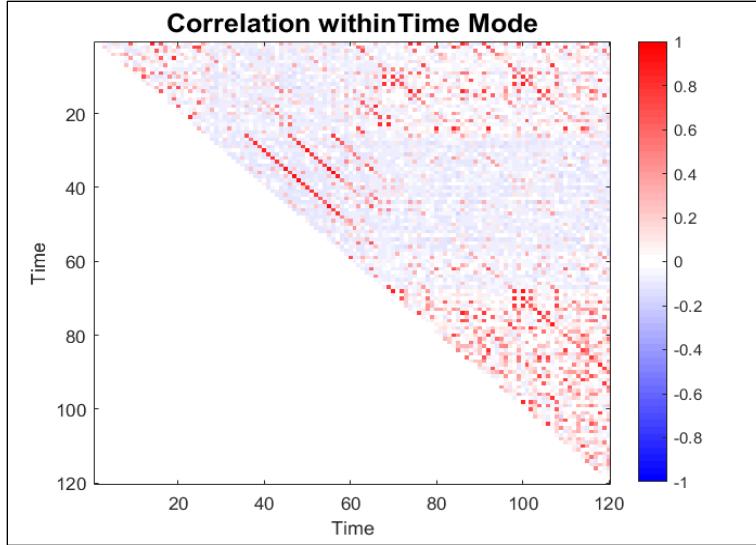


Figure 5. Correlations within single factor matrix. The Pearson correlation coefficient for all pairs of select component vectors (i.e., columns of a factor matrix) was computed. These pairs of select correlation vectors can be seen along the axes of the shown matrix. This figure shows an example of within-mode patterns that are captured by a CP decomposition of rank $R = 50$. All component vectors are shown in the factor matrices of Figure 4. A diverging color scale is used and ranges from negative one (blue) to zero (white) to one (red).

Given that each factor matrix has exactly one column for each rank-one tensor in the decomposition, the rows of distinct factor matrices can also be directly compared to each other, allowing for an inspection of the relationship between, say, user 1 and location 2. This method provides a high-level perspective in which the large trends are exposed and which may produce useful insights, although additional investigation will be required to further determine the nature of the similarity.

Examples of this technique are shown in Figure 6. The images in this figure show the correlation between the users and elements of other modalities. User 5 exhibits weaker correlations compared to the other users due to the lack of high component values in the factor matrices (column five of the leftmost image in Figure 4) which in turn is a manifestation of the lack of patterned behavior in the original data involving user 5. As for users 1–4, Figure 6a shows that there is positive correlation between activity 1 and users 1 and 3; further, there is positive correlation between activity 2 and users 2 and 4. These features are a clear reflection of the original data in which users 1 and 3 (and 5) always perform activity 1 and in which users 2 and 4 always perform activity 2. In a similar manner, Figure 6b shows a positive correlation between users 2 and 4 and among users 1, 3, and 5, which accurately reflects the structure of the original data. Figure 6c shows the correlations between users and times. The positive values corresponding to users 1–4 seem to indicate the time periods during which a user was occupying a location that the user occupied multiple times. While areas of positive correlation are in concurrence with the ground truth of the original embedded patterns, it is unclear why areas where no correlation was expected actually display negative values in Figure 6.

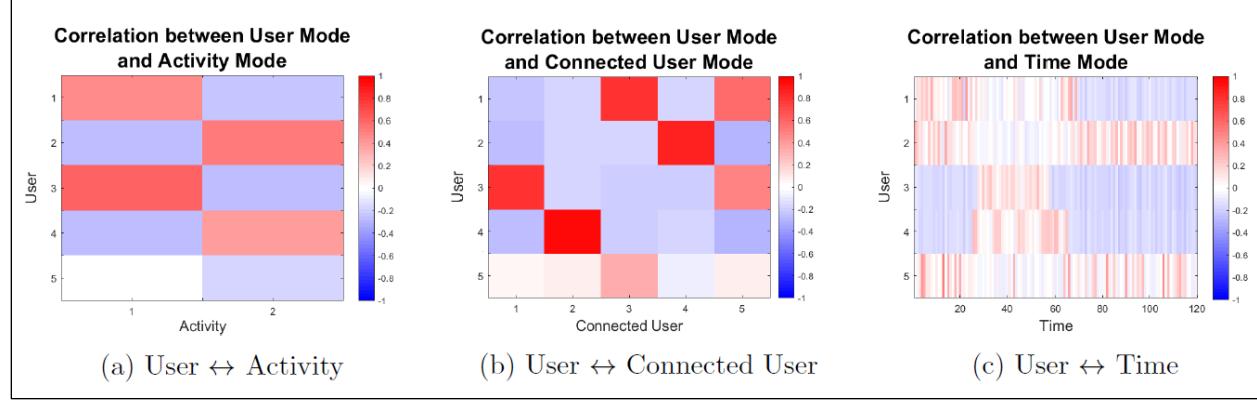


Figure 6. Correlations between component vectors. This figure shows examples of between-mode patterns that are captured by a CP decomposition of rank $R = 50$. The Pearson correlation coefficient for all pairs of select component vectors (i.e., columns of factor matrices) was computed. These pairs of selection correlation vectors can be seen along the axes of the shown matrices (and in their respective captions). For example, panel (a) displays all possible correlation values between user component vectors and activity component vectors. All component vectors are shown in the factor matrices of Figure 4. A diverging color scale is used and ranges from negative one (blue) to zero (white) to one (red).

Overall, the factor matrices provide a simplified view of the most prominent patterns within the original data. Instead of contending with a high-dimensional, hard-to-visualize tensor, the factor matrices provide a simpler view for each mode or pair of modes. An individual can control the level of fidelity to the data and conversely the noise reduction. In addition, since each entity in each mode is now characterized by a vector consisting of the values that entity is contributing to each rank-one tensor, correlations between entities within modes as well as entities across modes can be established.

CONCLUSION AND FUTURE WORK: The use of tensor representations for the analysis of multi-modal, spatiotemporal data shows promise. By fusing different data sets together, from spatial trajectories to social interactions, events, and more, the tensor-based framework allows for the analysis of the data as a whole, not just its constituent parts, thus providing a fuller understanding.

This TN illustrated that CP decomposition has the ability to extract strong features from induced tensors, and demonstrated multiple techniques for analyzing the factor matrices and rank-one components produced by CP decomposition. This led to the identification of trends within a fixed modality such as the time domain as shown in Figure 5 and between pairs of modalities such as between users and activity as shown in Figure 6. Additionally, the rate of convergence of the tensor approximations produced by CP decomposition by decomposing the synthetic data set over a wide range of tensor ranks was inspected and can be seen in Figure 2.

Observed limitations of CP decomposition, as described in ‘Scope of Rank-One Approximation,’ point toward the use of additional decomposition techniques that are tailored for the detection of other sorts of patterns in the data. In particular, the notion at close locations are likely similar points toward including spatiotemporal coherence explicitly in these tensor-decomposition techniques. A first step toward this end will be to design objective functions such as Equation (3) that treat spatial

and temporal modes in accordance with the expected coherence. Beyond this, there may also be a need to develop specialized methods for the optimization of these new objective functions.

As data related to moving objects continues to increase in quantity and variety, the ability to synthesize disparate data sets will become ever-increasingly crucial to ones understanding of how individuals navigate and interact with their environment. This proliferation in data-source collection technologies continues to present a challenge to the Army geospatial community. A vital part of addressing the gap in the current pattern-of-life domain is the existence and availability of frameworks and methods to integrate and analyze disparate data sources in a unified and uniform manner, as is described in this document.

In summary, the ability to synthesize both moving-object trajectories and contextual data with the goal of gaining insights into the latent interactions among various modalities will provide further insights into human behavior and patterns of life, thereby enhancing the Army GEOINT-HUMINT domain. To this end, a new approach for fusing multi-modal, spatiotemporal data was developed and their interactions were analyzed. As the research continues, the techniques herein described will be expanded to allow (1) for detection of a greater variety of patterns, (2) for integration of additional data types, and (3) for increased accuracy, robustness, and efficiency of pattern detection.

REFERENCES

- Bachmann, A., C. Borgelt and G. Gidófalvi. 2013. Incremental frequent route based trajectory prediction. *In Proceedings of the Sixth ACM SIGSPATIAL International Workshop on Computational Transportation Science*.
- Chawla, S., Y. Zheng, and J. Hu. 2012. Inferring the root cause in road traffic anomalies. *In Data Mining (ICDM) 2012 IEEE 12th International Conference*.
- Fanaswala, M. and V. Krishnamurthy. 2013. Detection of anomalous trajectory patterns in target tracking via stochastic context-free grammars and reciprocal process models. *In IEEE Journal of Selected Topics in Signal Processing* 7:76–90.
- Kolda, T. G., and B. W. Bader. 2009. Tensor decompositions and applications. *Society for Industrial and Applied Mathematics (SIAM)* 51(3):455–500. <https://doi.org/10.1137/07070111X>.
- Li, X., .M. Li, Y.-J. Gong, X.-L. Zhang, and J. Yin. 2016. T-desp: Destination prediction based on big trajectory data. *In IEEE Transactions on Intelligent Transportation Systems* 17(8):2344–2354.
- Li, Z., J. Han, B. Ding and R. Kays. 2012. Mining periodic behaviors of object movements for animal and biological sustainability studies. *Data Mining and Knowledge Discovery* 24(2): 355–386.
- Li, Z., F. Wu, and M. C. Crofoot. 2013. Mining following relationships in movement data. *In Data Mining (ICDM), 2013 IEEE 13th International Conference*.
- Pan, P., Y. Zheng, D. Wilkie, and C. Shahabi. 2013. Crowd sensing of traffic anomalies based on human mobility and social media. *In Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*.
- Wang, H., D. Kifer, C. Graif, and Z. Li. 2016. Crime rate inference with big data. *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Wang, Y., Y. Zheng, and Y. Xue. 2014. Travel time estimation of a path using sparse trajectories. *In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Vespier, U., S. Nijssen, and A. Knobbe. 2013. Mining characteristic multi-scale motifs in sensor-based time series. *In Proceedings of the 22nd ACM International Conference on Information and Knowledge Management*.

- Xu, Y., and W. Yin. 2013. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *Society for Industrial and Applied Mathematics (SIAM) Journal on Imaging Sciences* 6:1758–1789.
- Zhang, F., N. J. Yuan, D. Wilkie, Y. Zheng, and X. Xie. 2015. Sensing the pulse of urban refueling behavior: A perspective from taxi mobility. *ACM Transactions on Intelligent Systems and Technology (TIST)* 6:37.
- Zheng, K., Y. Zheng, X. Xie, and X. Zhou. 2012. Reducing uncertainty of low-sampling-rate trajectories. In *Data Engineering (ICDE), 2012 IEEE 28th International Conference*.
- Zheng, V. W., Y. Zheng, X. Xie and Q. Yang. 2010. Collaborative location and activity recommendations with GPS history data. In *Proceedings of the 19th International Conference on World Wide Web*.
- Zheng, V. W., Y. Zheng, X. Xie, and Q. Yang. 2012. Towards mobile intelligence: Learning from GPS history data for collaborative recommendation. *Artificial Intelligence* 84:17–37.
- Zheng, Y., and X. Xie. 2011. Learning travel recommendations from user-generated GPS traces. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2:2.
- Zheng, Y., T. Liu, Y. Wang, Y. Zhu, Y. Liu, and E. Chang. 2014. Diagnosing New York city's noises with ubiquitous data. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*.

NOTE: The contents of this technical note are not to be used for advertising, publication, or promotional purposes. Citation of trade names does not constitute an official endorsement or approval of the use of such products.