

Bipart: Learning Block Structure for Activity Detection

Yang Mu, Henry Z. Lo, Wei Ding, Kevin Amaral, and Scott E. Crouter

Abstract

Physical activity consists complex behavior, typically structured in bouts which can consist of one continuous movement (e.g. exercise) or many sporadic movements (e.g. household chores). Each bout can be represented as a block of feature vectors corresponding to the same activity type. This paper introduces a general distance metric technique to use this block representation to first predict activity type, and then uses the predicted activity to estimate energy expenditure within a novel framework. This distance metric, dubbed Bipart, learns block-level information from both training and test sets, combining both to form a projection space which materializes block-level constraints. Thus, Bipart provides a space which can improve the bout classification performance of all classifiers. We also propose an energy expenditure estimation framework which leverages activity classification in order to improve estimates. Comprehensive experiments on waist-mounted accelerometer data, comparing Bipart against many similar methods as well as other classifiers, demonstrate the superior activity recognition of Bipart, especially in low-information experimental settings.

Index Terms

Accelerometers, semisupervised learning, distance learning.

1 INTRODUCTION

In time series classification tasks, samples adjacent in time often have block structure, in which adjacent samples correspond to the same class. Given the potential benefit of knowing same-class samples, it would be folly not to use this information. This paper proposes a method to learn this block information from both training and test sets, and shows that such information improves classification performance empirically.

Our method is tested on waist-mounted accelerometer data, with the aim of determining activity type. In this dataset, each participant performed activities in blocks. The feature vectors extracted from each minute spent in a single block correspond to a single activity label [18], [21]. Both training and test data contain information about which feature vectors belong to which blocks. Though this study only uses waist data, analysis methods could be applied to data sets collected from other body locations in a similar manner [9].

To use this block structure, a classifier may label each vector individually, and vote on one class label within the block structure. However, this only takes into account block structure during classification, not during the learning phase. The proposed Bipart distance metric instead learns from the class labels, when given, and the block structure. This would hypothetically utilize information not otherwise used, and thus improve classification performance.

Ideally, feature vectors belonging to the same block should be well-clustered in feature space. The proposed distance metric method makes this clustering more apparent by creating a projection matrix which moves

same-block instances closer together. The method, which is dubbed Bipart, learns block structure from both the training and the test sets, and then combines the two parts to form a space which clusters same-class and same-block instances together. The samples embedded in this resulting Bipart space contains the same-block information, thus allowing any classifier to take advantage of this information by using the embedded samples.

Fig. 1 shows sample data projected onto 2-dimensional space; first in its original form, then with one Bipart projection (learned from the training set), and then with both Bipart matrices combined. Clustering between items in the same class improves as Bipart projections are applied to the dataset. This makes classification on the Bipart space easier compared to the original feature space, as any classifier operating in Bipart space will implicitly consider block-level information.

In this study, accelerometer data is classified into activities in order to effectively predict energy expenditure (measured as metabolic equivalents, or METs). Current models for translating accelerometer data (e.g. counts; “area under the curve” aggregated over a specific time interval, such as 1 sec) to a physical activity outcome (e.g. energy expenditure, time spent in moderate activity) mainly use single or multiple regression models, which don’t utilize the full capability of the data collected [5], [13]. To date, most accelerometer algorithms focus on energy expenditure without having a context for the activity taking place, which limits the accuracy of these models.

This work proposes that by first predicting the activity type, one should be able to better estimate energy ex-

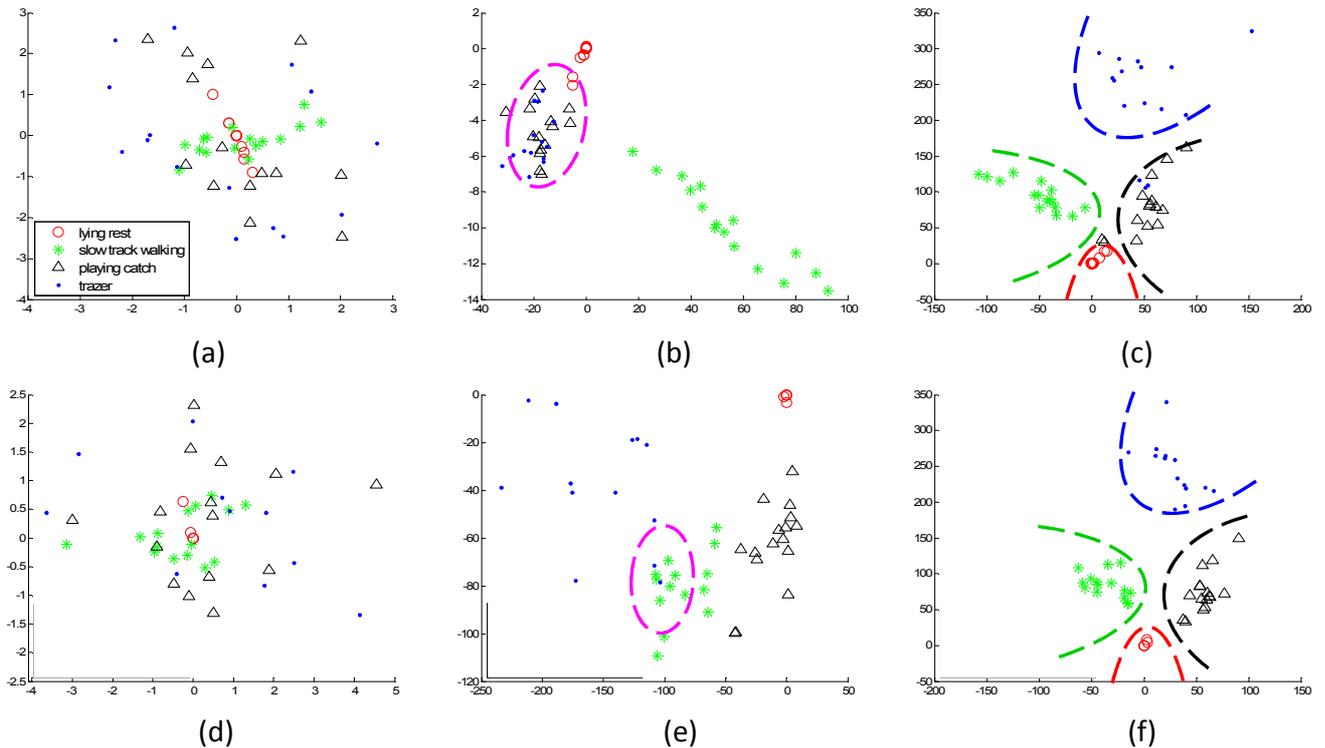


Fig. 1. Visualizations of real data. Top row shows training data, bottom row shows test data. Subfigures (a) and (d) shows the 2-dimensional PCA projection of the data. (b) and (e) shows the effect of projecting the data onto one distance metric (learned from the training set). (c) and (f) show the data projected onto Bipart space, combining the effects of both the training and test distance metric. The different activities begin to separate under one distance metric, and separates even further under Bipart space. Pink circles show problem areas, which disappear with more Bipart information. Other regions surrounded by dotted lines show the improved separation of data.

penditure, as this provides more information than count values. The proposed model, shown in Fig. 2, first uses classification to determine the activity of accelerometer bouts, then uses the activity class to select an appropriate regression model. One regression model is trained for each class; thus the classification piece is crucial for obtaining the correct energy expenditure. Using machine learning in the proposed model allows it to be more flexible and robust than the specific accelerometer, single-regression models which have predominated in the physical activity measurement field.

In summary, the contributions of this paper are as follows:

- 1) A framework which uses activity classification and multiple regression models to first predict activity type and then predicts energy expenditure (in the form of metabolic equivalents, or METs) using accelerometer data.
- 2) Formulation of the many-to-one classification problem, in which some data points are known to share the same class, and which generalizes to many other problems.
- 3) The Bipart method, a distance metric learning method which utilizes block structure in both the training and test sets, in addition to labeled training set data.

- 4) Extensive experiments which demonstrate the use of Bipart compared to other classifiers for the given problem.

2 RELATED WORK

The case study described in this paper relates three disparate fields of study: activity prediction and estimating energy expenditure using accelerometer data, multi-instance single label classification problems, and distance metric learning.

2.1 Energy Expenditure Estimation and Activity Prediction

The primary goal of this study is to use Bipart to classify activities or groups of activities using accelerometer data; the secondary goal is to develop methods to estimate energy expenditure from activities and activity groups.

A linear relationship between accelerometer counts and energy expenditure has been shown during locomotion [8]. Since this work, linear regression methods (which are specific to the activities developed on, and accelerometer model), have been the primary way to convert accelerometer data to a physical activity outcome. Recently, there has been a movement away from single regression models on limited activities [5], [13]. For

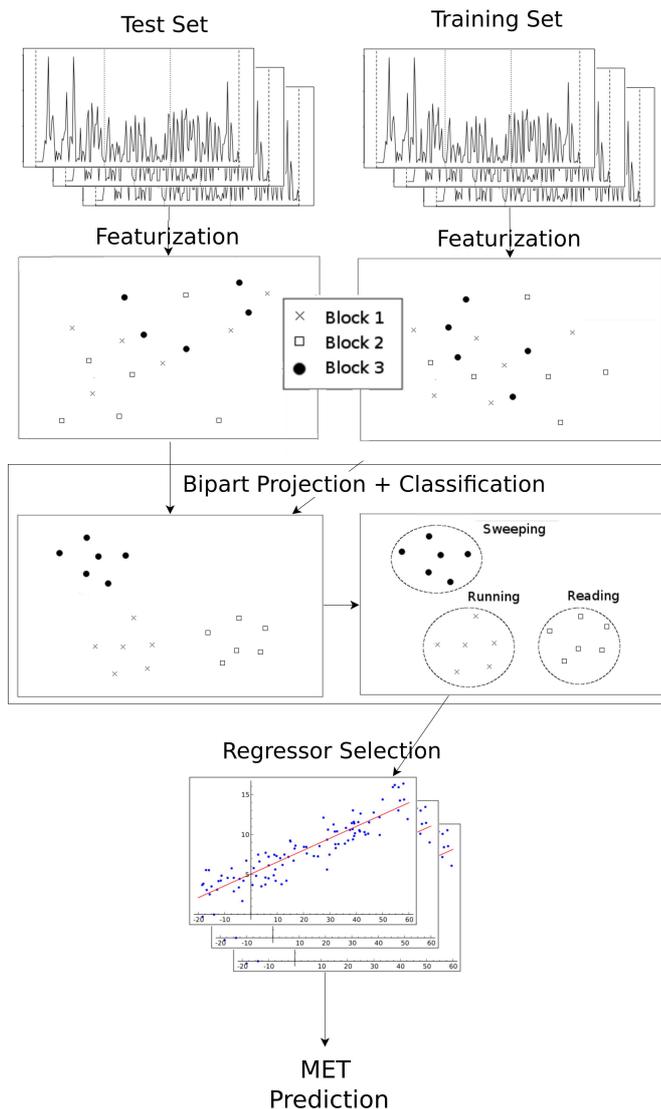


Fig. 2. Overview of energy expenditure (MET) prediction framework. Each bout of activity is formed into several feature vectors. One Bipart projection matrix is learned from the training set and one from the test set. The data samples are projected into the combination of the two. An activity is determined through classification, and this activity is used to select a regression model with which to predict energy expenditure.

example, Crouter and colleagues have developed a two-regression model that differentiates between walking and running activities and intermittent lifestyle activities based on the variability in the accelerometer counts [3], [4]. Compared to other models available, the 2-regression model reduces both the mean group error and individual error for estimating energy expenditure and time spent in intensity categories (e.g. moderate activity) [18], [22].

With advancements in technology and reduced cost of the devices, rapid advancements are taking place in how accelerometer data is used for physical activity assessment. Activity classification has begun to gain

momentum as a feasible way to get activity type and then estimate energy expenditure, especially with machine learning techniques. Among these, feedforward backpropagation neural networks are the most popular and tend to be very successful [7], [18], [22]. Naive Bayes and other classifiers have also been applied to the problem of task classification [12], [20].

2.2 Multiple Instance Single Label Problem

The Bipart activity classification method uses block structure information from both the training and test sets. In particular, it applies to datasets in which data instances are grouped, and within these groups, the data instances are known to share the same class label. In addition to class labels, Bipart learns about this group membership information from both the training and test set data in order to transform the original feature space.

This block structure is superficially similar to the structure delineated in the multiple instance single label problem (MISL) literature. MISL classification also generalizes normal classification by assigning labels to bags of feature vectors, rather than to feature vectors themselves. However, there are a few key differences:

- 1) In the original formulation of MISL, only binary classification is allowed [6].
- 2) MISL requires that any instance in a “bag” found to be positive makes the entire bag positive.

Thus, many approaches derived for MISL do not apply to our many-to-one classification problem [30]. One exception is citation kNN (CkNN), a lazy learning method which extends the kNN method for multiple-instance classification [23].

2.3 Distance Metric Learning

Bipart incorporates block-level information by materializing block-level relationships as closer distances in a projection matrix. The decision to use distance metric learning allows for other classifiers to use the block level information, thereby allowing for more flexibility.

Distance metric learning approaches transform data into a representation which reflects relationships between data points. This typically means moving members of the same class (similar) together, and separating samples of different classes (dissimilar). Many existing approaches generalize Mahalanobis distance metrics [10], [19], [24], [26], [27]. It is worth pointing out that all the generalized Mahalanobis distances are equivalent to Euclidean distance under a projected space [17].

Xing’s algorithm is typical of many global distance metric methods; it uses convex optimization to satisfy both similarity and dissimilarity constraints simultaneously [26]. These constraints are built globally over the entire dataset. Some methods, such as large margin nearest neighbor (LMNN) [24] and local Fisher discriminant analysis (LFDA) [19], utilize only neighborhood constraints, rather than all constraints in the dataset, to learn

the distance. It has been shown that global constraint methods have difficulty with multimodal distributions, which local constraint methods do not suffer from [15].

Like LMNN and LFDA, Bipart uses local constraints to learn its distance metric. We decided on the local approach, which has been shown to have superior discrimination and more robustness to multimodal distributions.

Bipart uses a distance metric constructed from the training set and modified by the bag (not class) information in the test set. This differs from the above approaches, which only obtain discriminative information from the training set. The closest to Bipart are the semi-supervised approaches, e.g, semi-supervised discriminant analysis (SDA) [1], which considers both labeled and unlabeled samples. However, the dataset in this paper includes bag-level information, which cannot be utilized by SDA and related methods. Bipart takes advantage of this information to similarity and dissimilarity constraints, and thus is novel in this regard.

3 OVERVIEW

The framework used in this paper first categorizes a group of accelerometer signals into an activity, uses the activity to select a regression model, then applies the regression on the original data to estimate energy expenditure.

Specifically, training the framework consists of two steps:

- 1) Using activities as class labels, and using a bout of activity as a bag to be classified, learn a Bipart matrix with the appropriate constraints.
- 2) Using class labels to divide the dataset, train one regression model for each activity.

After training, the framework is ready to be applied to test data. It processes such data as follows:

- 1) Using bag information in the test set, learn a second Bipart matrix.
- 2) Combine the two Bipart matrices to form a unified distance matrix, and project each data-point into the space defined by the matrix.
- 3) Classify the projected data-point using kNN.
- 4) Use the predicted activity to select the regression model for that activity, and use the model to estimate energy expenditure from the original accelerometer data.

This block structure differentiates the classification problem in this paper with typical classification. In the latter, one example is associated with one label. That is, given a data set $(\mathbf{X}, Y) = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_i, y_i), \dots, (\mathbf{x}_n, y_n)\}$, where $\mathbf{x}_i \in \mathbb{R}^d$, and y_i is the label of \mathbf{x}_i , the goal is to generate a model to classify unknown examples.

For this many-to-one classification problem, instead of classifying unknown examples, the goal is to classify the unknown block, which is defined as $\mathbf{B}_i = \{\mathbf{x}_1^{\mathbf{B}_i}, \dots, \mathbf{x}_{k_i}^{\mathbf{B}_i}\}$, where k_i is the number of examples in block B_i . The corresponding label $y^{\mathbf{B}_i}$ of \mathbf{B}_i is defined

\mathbf{X}	Dataset of n samples with d dimensions
\mathbf{x}_i	Data sample i
Y	Labels for each element in \mathbf{X}
y_i	Label for data sample i
\mathbf{B}_i	i -th block of samples, $\{\mathbf{x}_1^{\mathbf{B}_i}, \dots, \mathbf{x}_{k_i}^{\mathbf{B}_i}\}$
$y^{\mathbf{B}_i}$	Label for block i
k_i	Number of elements in block B_i
$d_{\mathbf{A}}$	Distance metric defined by matrix \mathbf{A}
\mathbf{A}	The unified objective distance metric
\mathbf{W}_1	Distance metric learned from training
\mathbf{W}_2	Distance metric learned from testing
\mathbf{B}_i^d	Block nearest \mathbf{B}_i with different class label
\mathbf{B}_i^s	Block nearest \mathbf{B}_i with same class label
$\mathbf{x}_q^{\mathbf{B}_i^d}$	q^{th} sample from block \mathbf{B}_i^d
$\mathbf{x}_q^{\mathbf{B}_i^s}$	q^{th} sample from block \mathbf{B}_i^s
β	Balancing parameter for training and test metrics
\mathbf{S}_i	Selection matrix
n_1	number of samples in the test set

TABLE 1
Notation used.

as $y^{\mathbf{B}_i} = y_1^{\mathbf{B}_i}, \dots, y_{k_i}^{\mathbf{B}_i}$. All vectors in the same block have the same label.

4 THE BIPART METHOD

The Bipart metric is forged from two distance metrics: one learned from the labels and block structure of the training set, and the other is learned just from the block structure of the test set. The two are combined into a single metric, which is used to project the data onto for classification.

4.1 Distance Metric Learning

Between any two samples of time series data, \mathbf{x}_i and \mathbf{x}_j , is the distance $d_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j)$, defined by the metric $d_{\mathbf{A}}$. Many distance metric learning methods generalize the Mahalanobis distance, which are of the form:

$$d_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{A} (\mathbf{x}_i - \mathbf{x}_j)}, \quad (1)$$

where \mathbf{A} is positive semi-definite. Note that when \mathbf{A} is the identity matrix, this simplifies to Euclidean distance. Technically, this allows pseudometrics, i.e. $d_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j) = 0$ does not imply $\mathbf{x}_i = \mathbf{x}_j$. Using the Cholesky decomposition, \mathbf{A} can be replaced with $\mathbf{W}^T \mathbf{W}$ in Equation (1), giving:

$$\begin{aligned} d_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j) &= \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{W} \mathbf{W}^T (\mathbf{x}_i - \mathbf{x}_j)} \\ &= \|\mathbf{W}^T (\mathbf{x}_i - \mathbf{x}_j)\|. \end{aligned} \quad (2)$$

Approaches differ in how to learn \mathbf{A} in Equation (1) [24], [26], [27], but all ensure that similar examples have a small distance under the learned metric.

Our proposed Bipart distance metric is similar to the second form (Equation 2). It uses two distance metrics by replacing \mathbf{W} with $\mathbf{W}_1\mathbf{W}_2$:

$$d_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{W}_2^T \mathbf{W}_1^T (\mathbf{x}_i - \mathbf{x}_j)\|, \quad (3)$$

where \mathbf{W}_1 and \mathbf{W}_2 correspond to the distance metrics learned from test and training data respectively.

Equation (3) is equivalent to projecting all samples onto the space defined by projection matrix \mathbf{W}_1 , then to \mathbf{W}_2 . The projection matrix \mathbf{W}_2 defines the space in which all the data ends up, and so should be learned from more reliable data. Thus, \mathbf{W}_2 is learned from the training set, as it includes activity label information, and \mathbf{W}_1 is learned from the test set.

4.2 Bipart Distance Metric Objective

Learning the projection matrices \mathbf{W}_1 and \mathbf{W}_2 in Equation (3) requires finding a metric space that keeps all the examples in the same classes and blocks close, and those from different classes and blocks separated. The local patch alignment framework [28] and similarity and dissimilarity constraints [26] formulates two objectives: first, to minimize the distance between any two samples in the same labeled blocks, and second, to maximize the distance between any two samples in two different labeled blocks.

Previous studies [24], [26], [27] have shown that building constraints from only neighborhood information is superior to the global constraints approach in dealing with multi-modal distributions. Taking this into account, Bipart forms its objectives using local constraints. For any example \mathbf{x}_i in block \mathbf{B}_i^s , similarity constraints are formed from other elements in the same block. Dissimilarity constraints are formed only from elements in the nearest (according to the minimal Hausdorff distance) block in a different class, denoted as \mathbf{B}_i^d .

The training procedure for the Bipart distance metrics contains two phases: first, learn \mathbf{W}_1 from the test set, and second, learn \mathbf{W}_2 from the training set. As there is no mathematical difference between learning from \mathbf{W}_1 and \mathbf{W}_2 , we use the test data distance metric \mathbf{W}_1 to illustrate the training procedure.

Let \mathbf{x}_i be a sample, the k_i^s vectors in the nearest block with the same class $\mathbf{x}_p^{\mathbf{B}_i^s} \in \mathbf{B}_i^s$, and k_i^d vectors in the nearest different-class block $\mathbf{x}_q^{\mathbf{B}_i^d} \in \mathbf{B}_i^d$. The following objective function minimizes the similarity constraints:

$$\arg \min_{\mathbf{A}_1} \sum_{i=1}^{n_1} \sum_{p=1}^{k_i^s} d_{\mathbf{A}_1}^2(\mathbf{x}_i, \mathbf{x}_p^{\mathbf{B}_i^s}), \quad (4)$$

The following objective function maximizes the dissimilarity constraints:

$$\arg \max_{\mathbf{A}_1} \sum_{i=1}^{n_1} \sum_{q=1}^{k_i^d} d_{\mathbf{A}_1}^2(\mathbf{x}_i, \mathbf{x}_q^{\mathbf{B}_i^d}) \quad (5)$$

where $A_1 = \mathbf{W}_1^T \mathbf{W}_1$ is the distance metric learned from the test set.

Equations (4) and (5) can be combined into one objective function, utilizing the scaling parameter β :

$$\arg \min_{\mathbf{A}_1} \sum_{i=1}^{n_1} \left(\sum_{p=1}^{k_i^s} d_{\mathbf{A}_1}^2(\mathbf{x}_i, \mathbf{x}_p^{\mathbf{B}_i^s}) - \beta \sum_{q=1}^{k_i^d} d_{\mathbf{A}_1}^2(\mathbf{x}_i, \mathbf{x}_q^{\mathbf{B}_i^d}) \right) \quad (6)$$

The distance metric \mathbf{A}_1 , as well as \mathbf{W}_1 , can be solved from the objective function in Equation (6). Similarly, \mathbf{W}_2 can be solved using the training set under the distance metric \mathbf{A}_1 . With \mathbf{W}_1 and \mathbf{W}_2 , we can obtain the final distance metric \mathbf{A} using Equation (3). Under this distance metric, the abundant discriminative information of training set as well as the test set is well preserved.

4.3 Bipart Distance Metric Solution

In this section, the closed form solution for \mathbf{W}_1 in Equation (6) is derived.

Let the test sample be \mathbf{x}_i , its same-class blocks be \mathbf{B}_i^s , and nearest different-class block \mathbf{B}_i^d be combined into a matrix \mathbf{X}_i , where:

$$\begin{aligned} \mathbf{X}_i &= [\mathbf{x}_i, \mathbf{B}_i^s, \mathbf{B}_i^d] \\ &= [\mathbf{x}_i, \mathbf{x}_1^{\mathbf{B}_i^s}, \dots, \mathbf{x}_{k_i^s}^{\mathbf{B}_i^s}, \mathbf{x}_1^{\mathbf{B}_i^d}, \dots, \mathbf{x}_{k_i^d}^{\mathbf{B}_i^d}]. \end{aligned} \quad (7)$$

Let the coefficients w_i be defined as follows:

$$w_i = \left[\overbrace{1, \dots, 1}^{k_i^s}, \overbrace{-\beta, \dots, -\beta}^{k_i^d} \right]. \quad (8)$$

Using Equations (7) and (8), Equation (6) can be reduced to:

$$\begin{aligned} &\arg \min_{\mathbf{A}_1} \sum_{i=1}^{n_1} \left(\sum_{j=1}^{k_i^s+k_i^d} d_{\mathbf{A}_1}^2(\mathbf{X}_i\{1\}, \mathbf{X}_i\{j+1\})(w_i)_j \right) \\ &= \arg \min_{\mathbf{W}_1} \sum_{i=1}^{n_1} \left(\sum_{j=1}^{k_i^s+k_i^d} \|\mathbf{W}_1(\mathbf{X}_i\{1\} - \mathbf{X}_i\{j+1\})\|_2^2 (w_i)_j \right) \\ &= \arg \min_{\mathbf{W}_1} \sum_{i=1}^{n_1} \text{tr}(\mathbf{W}_1^T \mathbf{X}_i \mathbf{L}_i \mathbf{X}_i^T \mathbf{W}_1), \end{aligned} \quad (9)$$

where $\mathbf{X}_i\{j\}$ is the j^{th} column of matrix \mathbf{X}_i , $(w_i)_j$ is the j^{th} element of (w_i) , and $\mathbf{L}_i \in \mathbb{R}^{(k_i^s+k_i^d+1) \times (k_i^s+k_i^d+1)}$ is given by

$$\mathbf{L}_i = \begin{bmatrix} \sum_{j=1}^{k_i^s+k_i^d} (w_i)_j & -w_i^T \\ -w_i & \text{diag}(w_i) \end{bmatrix}. \quad (10)$$

Since \mathbf{X}_i is selected from the entire test data set \mathbf{X} , \mathbf{X}_i can be written as:

$$\mathbf{X}_i = \mathbf{X} \mathbf{S}_i, \quad (11)$$

where $\mathbf{S}_i \in \mathbb{R}^{n_1 \times (k_i^s+k_i^d+1)}$ is a selection matrix, with elements defined as follows:

$$(\mathbf{S}_i)_{pq} = \begin{cases} 1 & \text{if } p = D_i\{q\} \\ 0 & \text{else} \end{cases}, \quad (12)$$

where $D_i = [i, i_1^{\mathbf{B}^s}, \dots, i_{k_i^s}^{\mathbf{B}^s}, i_1^{\mathbf{B}^d}, \dots, i_{k_i^d}^{\mathbf{B}^d}]$ is the index set for \mathbf{X}_i . With all this, Equation (9) can be rewritten as:

$$\begin{aligned} & \arg \min_{\mathbf{W}_1} \sum_{i=1}^{n_1} \text{tr}(\mathbf{W}_1^T \mathbf{X}_i \mathbf{L}_i \mathbf{X}_i^T \mathbf{W}_1) \\ &= \arg \min_{\mathbf{A}_1} \text{tr} \left(\mathbf{W}_1^T \mathbf{X} \sum_{i=1}^{n_1} (\mathbf{S}_i \mathbf{L}_i \mathbf{S}_i) \mathbf{X}^T \mathbf{W}_1 \right) \\ &= \arg \min_{\mathbf{A}_1} \text{tr}(\mathbf{W}_1^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{W}_1), \end{aligned} \quad (13)$$

where $\mathbf{L} = \sum_{i=1}^{n_1} \mathbf{S}_i \mathbf{L}_i \mathbf{S}_i^T \in \mathbb{R}^{n_1 \times n_1}$ is the alignment matrix [29] [28].

To make the projection matrix \mathbf{W}_1 linear and orthogonal, we impose the constraint condition $\mathbf{W}_1^T \mathbf{W}_1 = \mathbf{I}_d$, where \mathbf{I}_d is a $d \times d$ identity matrix. The objective function in Equation (13) then becomes:

$$\arg \min_{\mathbf{A}_1} \text{tr}(\mathbf{W}_1^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{W}_1) \text{ s.t. } \mathbf{W}_1^T \mathbf{W}_1 = \mathbf{I}_d. \quad (14)$$

Solutions of Equation (14) can be obtained by using standard eigen-decomposition:

$$\mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{u} = \lambda \mathbf{u}. \quad (15)$$

Let the column vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d$ be the solution of Equation (15), ordered according to the eigenvalues $\lambda_1 < \lambda_2 < \dots < \lambda_d$. The optimal projection matrix \mathbf{W}_1 is then given by: $\mathbf{W}_1 = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{d_1}]$, where $d_1 < d$. Once \mathbf{W}_1 is calculated, the distance metric of the first part \mathbf{A}_1 can be obtained by Equation (2), which is not required to be calculated explicitly.

Similarly, \mathbf{W}_2 can be obtained by using projected training data by \mathbf{W}_1 . Finally, we have the final projection matrix $\mathbf{W} = \mathbf{W}_1 \mathbf{W}_2$ and the corresponding Bipart distance metric \mathbf{A} . \mathbf{W}_1 reduces the dimension from d to d_1 , and \mathbf{W}_2 further reduces the dimension to d_1 from d_2 . d_2 is the dimension of the final low-dimensional discriminative Bipart distance metric space.

5 MET PREDICTION

5.1 Classification

Projecting the dataset onto the Bipart metric preserves block structure information by ingrainning it into the resulting dataset. In this study, classification is done on the resulting dataset using a k-nearest neighbor approach. Each example in a block is classified individually, and the resulting classes are voted on to assign the label of the entire block.

Though this is a relatively unsophisticated classifier, it is hypothesized to perform better than methods which do not consider block-level information. Block level information could be exploited by voting within the block; however, this only takes advantage of block structure

during the testing phase, not training. Thus, Bipart kNN is expected to outperform classification with voting as well.

5.2 Multi-Linear Regression

The label outputted by the kNN classifier is used to select an appropriate multi-linear regression model [25]. The models are pre-trained. For the activity classification paradigm, one model was trained for each activity, and for categorized classification, one model was trained for each category.

The linear regression method requires finding a linear model β which, when applied to \mathbf{X} , results in the predicted METs \mathbf{y} with minimal error ε .

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \quad (16)$$

where \mathbf{X} is an $n \times (d+1)$ matrix representing n samples. The values in the extra dimension are always 1; this is for learning the constant bias β_0 .

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_d x_{i,d}$$

The model is learned from the training set \mathbf{X}_{train} by minimizing ε_{train} and solving for β in the following equation:

$$\mathbf{y}_{train} = \mathbf{X}_{train} \beta + \varepsilon_{train}$$

The model can then be applied to predict the MET values \mathbf{y}_{test} for the testing set \mathbf{X}_{test} .

$$\mathbf{y}_{test} = \mathbf{X}_{test} \beta + \varepsilon_{test}$$

6 EXPERIMENTS

6.1 Data Description and Feature Representation

This was part of a larger study, and the data and the participant characteristics and methods have been published elsewhere [3]. Data from indirect calorimetry and waist-mounted accelerometers attached to 112 children were used in this study. Each child performed lying rest (30 minutes) and six of the other 18 physical activities (7 minutes each). The physical activities, and the corresponding categories, are:

- Sedentary activities: lying rest, reading, watching TV, searching the internet
- Household chores: sweeping, vacuuming
- Locomotion: slow track walking, brisk track walking, walking with a 10-lb backpack, track running
- Interactive Video Games: Nintendo Wii, Light Space, Wall Light Space, Dance Dance Revolution, Trazer
- Exercise and Sports: playing catch, soccer around cones, sport wall, workout video

During all activity measurements energy expenditure was measured using indirect calorimetry (Cosmed K4b2) so that the predicted energy expenditure estimates could be compared to a gold standard. Accelerometer measurements were simultaneously collected using an ActiGraph GT3X tri-axial accelerometer, worn on the right hip. Accelerometer measurements in the x, y, and z directions

were aggregated to produce one count for every dimension and every second. From this aggregate, a feature vector block of 60 instances for every minute of activity was constructed. This feature block was associated with one class label. The types of features used are the same as in other energy expenditure estimation studies using neural networks [18], [22], except that all three axes of data are used, whereas the authors of those papers only used x-axis data.

The constructed feature vectors consist of the following:

- Block ID.
- 10th, 25th, 50th, 75th, and 90th percentile values for 60 one-second counts.
- Lag-1 to lag-9 autocorrelations, to represent temporal relations.

6.2 Experimental Design

Two general types of experiments were performed: activity classification and estimation of energy expenditure (i.e. METs). Within these two experiments, three types of training validation were performed:

- Leave-one-person-out (LOPO), as in [18]. All participants but one were used for training, and the held out participant's activities were used for validation. This is the most realistic experimental setting.
- Random splitting (RS). The percentage of subjects used in training varied incrementally from 10% to 90%, and the rest were used for testing. This setting tests the performance of various classifiers under different training conditions (insufficient/sufficient training data).
- 10-fold cross validation (CV). This setting is widely used in many data mining problems to combat overfitting.

Two different types of datasets were used. As shown in Table 2 and Fig. 3, the first dataset contains all 19 class labels, and the second dataset categorizes the 19 activities into five category labels.

6.3 Activity Classification

The following classifiers were tested:

- State-of-the-art classifiers, which have been used in previous work on mining accelerometer data [7], [12], [18], [20], [22].
 - Feedforward Backpropagation Artificial Neural Network (ANN)
 - k Nearest Neighbor (kNN)
 - Support Vector Machine, using the one-vs-all method to handle multiple classes [16], and the following kernels:
 - 1) Linear kernel (SVM-linear)
 - 2) Radial basis function kernel (SVM-RBF)
 - Naive Bayes
- Citation-kNN (CkNN) [23], a multi-instance classifier. CkNN is suitable for the proposed problem,

while other multi-instance multi-label approaches [14], [30], [31] are not, as they are trained based on the diversity of the blocks.

- The proposed method, Bipart, using a 3NN classifier.
- The following distance metric learning methods, with a 3NN classifier.
 - No distance metric (Euclidean)
 - Xing's method (Xing)
 - Local Fisher's discriminant analysis (LFDA)
 - Semi-supervised discriminant analysis (SDA)

Classification was performed in two different ways:

- In the first, each feature vector was classified, as in typical classification problem (no-voting).
- In the second, majority voting between labels in a block were used to determine the block label (voting). For CkNN and Bipart, there is no difference between voting and no-voting.

For a visual summary of the different experimental variations, see Fig. 4.

Classification is evaluated using accuracy, which is the ratio of correct classifications over the total number of test samples [11].

6.4 Classifier Parameters

The feedforward backpropagation neural network had one hidden layer and 25 hidden neurons, as in [18], [22].

The CkNN classifier was used with $k = 2$ and $c = 4$, optimal values in [23].

The kNN classifier had $k = 3$.

Naive Bayes was used with default settings. Linear kernel SVM was applied with optimal settings on validation sets.

The Bipart distance metric had two parameters: the scaling parameter β , as shown in Equation (6), and dimension d_2 , as discussed for Equation (15). β is selected in a range of $(2^{-3} - 2^3)$. d_2 is automatically decided when 90% energy is achieved according to the eigenvalues.

6.5 Regression

Regression models for each activity (or category) were trained to predict METs from the feature representation shown in Section 6.1. In the classification-regression framework, all classifiers share the same linear regression models. That is, if two classifiers result in the same activity classification, then the same regression model is selected, and the resulting predicted MET will be the same.

The classification step determines the class label, and this labeled activity (or category) is used to select the regression trained specifically for this activity (or category). The selected regression model is used on the original data in order to predict a MET value.

Regression results are reported using the RMSE (root-mean-square error). For a visualization of the scale of RMSE, see Fig. 3.

Category	Activity	MET range (min. - max.)	
		Activity	Category
Sedentary	Lying Rest	1.0000 - 1.0000	0.6448 - 2.4799
	Reading	0.7702 - 2.4799	
	Watching TV	0.6523 - 2.1141	
	Searching Internet	0.6448 - 1.6608	
Chores	Sweeping	1.2728 - 5.8562	1.2728 - 5.8562
	Vacuuming	1.7355 - 4.8597	
Locomotion	Slow Track Walking	2.0546 - 7.2180	2.0546 - 11.2163
	Brisk Track Walking	2.3348 - 8.8780	
	Walking with 10 lb Backpack	2.3274 - 6.4440	
	Track Running	4.6846 - 11.2163	
Interactive Video Games	Nintendo Wii	1.1206 - 5.7367	1.1206 - 9.1458
	Light Space	2.4098 - 9.1458	
	Wall Light Space	2.5449 - 8.6164	
	Dance Dance Revolution	1.7943 - 6.1126	
	Trazer	1.8256 - 8.7463	
Exercise and Sports	Playing Catch	1.6448 - 5.8235	1.4361 - 10.9344
	Soccer Around Cones	2.0343 - 10.5173	
	Sport Wall	3.0160 - 10.9344	
	Workout Video	1.4361 - 4.5338	

TABLE 2

Physical activities, categories of physical activities, and the corresponding range of measured METs for those activities and categories.

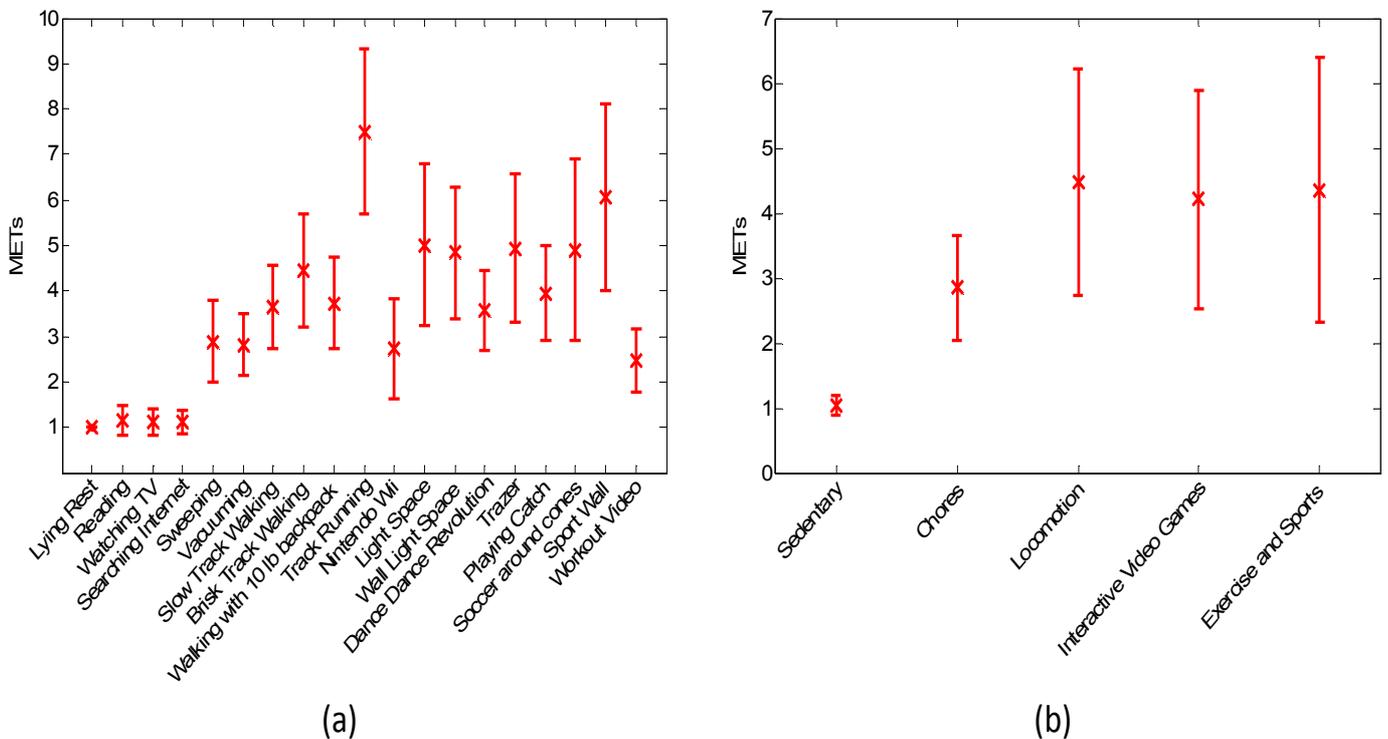


Fig. 3. Distribution of measured energy expenditure for the different physical activities and categories. Energy expenditure is described by measured METs. "x" marks represent the mean values, and bars correspond to standard deviations. (a) Activities. The x-axis shows the 19 physical activities. (b) Categories. The x-axis shows the five categories of activities.

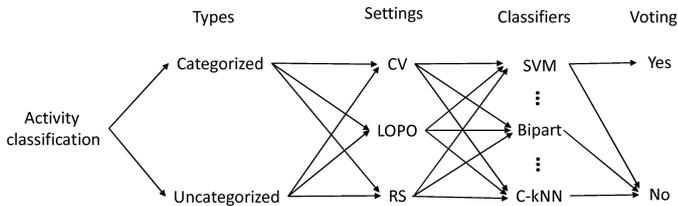


Fig. 4. Structure of classification experiments. Experiments are divided by class label, evaluation methodology, algorithm type, and whether or not voting is done. At each level, all experimental conditions to the right are applied; e.g., CV, LOPO, and RS are done for both categorized and uncategorized data. The sole exception is that voting was not done for the CkNN classifier.

7 RESULTS

7.1 Classification

Results for activity classification experiments are shown in Table 3 for classifiers with no voting, and Table 4 with voting. Category classification results are shown in Table 5 before voting, and Table 6 after voting.

The difficulty of multi-class classification is directly determined by the number of classes - for example, in our 19 activity classification scenario, random guess would only yield an accuracy of $\frac{1}{19} = 5.26\%$; comparatively, naive Bayes (the worst) achieves 16.35% and Bipart 53.00% accuracy on pre-voting cross-validation, as seen in Table 3. Categorizing activities will benefit accuracy both by reducing the number of classes to 5, and by grouping similar activities together in a meaningful way; in categorized classification, random guess will yield $\frac{1}{5} = 20.00\%$ accuracy. Comparatively, CkNN (the worst) achieves an accuracy of 57.66% and Bipart yields 81.64% on pre-voting cross-validation, as seen in Table 5. More detailed explanations are given below.

In experiments without voting (Tables 3 and 5), Bipart comes out as a clear winner. Thus, Bipart with kNN outperforms any unaided classifier or distance metric method. In voting experiments, LFDA and neural networks approached Bipart's performance under certain experimental conditions.

Bipart also performs best in LOPO experiments, which are the most realistic. It also consistently outperforms in situations with low training data, as is evident in the random split conditions, especially in "rs_9", in which only 10% of subjects were used for training. This is presumably due to Bipart using block-level information in the testing phase.

All classifiers performed much better in predicting categories than predicting activity types. As shown in [2]–[4], categorization improves classification performance. The confusion matrix in Table 7 show that sedentary activities often get confused as one another, as do locomotor activities. Table 8 shows that there is much less confusion after categorization.

The predictability of these activities and categories

were not obviously related to size of the range of measured METs as shown in Table 2. However, they were somewhat related to the "regularity" of the activities in terms of the accelerometer measurements; sedentary activities all involve little movement, but there is little in common in the many different actions performed in exercise and sports.

Table 9 shows that sedentary activities are difficult to differentiate. Sweeping and vacuuming are difficult for all classifiers, though categorization improves performance, as seen in 10. Locomotion activities are the easiest to distinguish, aside from sedentary. Exercise and sports and interactive video games have varying difficulty.

7.2 Regression

Regression results of each classifier over all experimental conditions are presented in Table 11. Prediction accuracy is directly related with classification accuracy under the proposed framework; thus, Bipart performs best in most settings, ceding a few to neural networks. LOPO results can be seen in Fig. 5. Though neural networks followed closely, Bipart achieved the lowest RMSE.

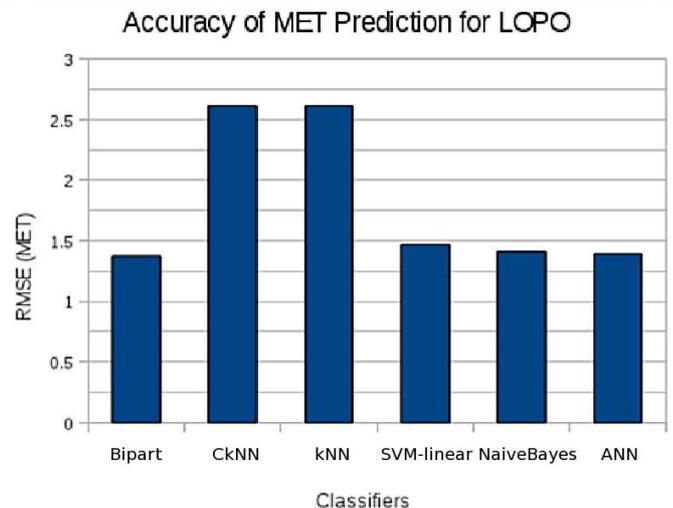


Fig. 5. Root mean square error (RMSE) for each classifier for estimation of METs across all activities in the LOPO experiment. Bipart achieves the lowest RMSE at 1.37, followed by ANN at 1.39, and Naive Bayes at 1.41.

MET prediction results using activity category are shown in Table 12. Bipart performs better than other approaches in interactive video games or exercise and sports, and is comparable on sedentary activities. In locomotion activities, neural networks outperforms by Bipart by a small margin.

RMSEs achieved by Bipart are relatively low, considering the range of MET values per category as shown in Fig. 3. The wide range of MET values for each activity group puts a limit on the accuracy of regression results.

	Other Classifiers					kNN				
	SVM-linear	SVM-RBF	NaiveBayes	ANN	CkNN	Euclidean	Xing	LFDA	SDA	Bipart
cv	47.80	49.99	16.35	51.25	39.06	46.96	43.85	44.48	42.68	53.00
lopo	48.93	51.08	17.44	52.21	39.96	48.02	44.53	48.73	43.23	59.81
rs_1	46.53	48.70	15.28	48.65	38.36	46.58	42.72	42.87	42.03	52.29
rs_2	48.63	51.15	16.24	51.70	39.13	48.74	45.03	45.25	44.30	54.03
rs_3	47.17	50.02	15.95	49.95	38.36	46.73	42.67	42.81	41.96	51.54
rs_4	46.92	50.34	18.51	50.38	38.58	47.39	43.17	43.07	42.54	51.94
rs_5	45.54	49.21	37.53	48.85	38.57	46.39	42.37	42.16	42.13	51.12
rs_6	45.39	49.37	46.30	48.48	38.58	46.18	42.62	41.97	41.94	51.22
rs_7	45.13	49.31	45.06	47.86	38.67	46.06	41.94	41.33	41.69	51.49
rs_8	43.52	47.56	37.95	46.18	37.89	45.00	41.05	39.85	40.61	50.09
rs_9	41.41	45.96	43.69	44.59	38.37	43.78	40.29	39.13	39.53	48.43

TABLE 3

Accuracy (%) under various experimental paradigms for classifying individual activities, without voting. Best results highlighted.

	Other Classifiers					kNN				
	SVM-linear	SVM-RBF	NaiveBayes	ANN	CkNN	Euclidean	Xing	LFDA	SDA	Bipart
cv	50.38	52.67	16.46	54.21	39.06	50.49	51.44	52.12	48.86	53.00
lopo	51.23	53.82	18.00	55.97	39.96	52.32	53.29	57.94	51.29	59.81
rs_1	48.65	50.96	15.69	50.95	38.36	50.16	50.02	52.13	47.82	52.29
rs_2	52.05	54.52	17.05	55.45	39.13	52.62	53.34	53.25	51.48	54.03
rs_3	49.59	52.54	15.39	53.21	38.36	50.58	51.15	50.95	50.02	51.54
rs_4	48.41	53.14	18.78	54.17	38.58	51.40	52.03	51.26	50.23	51.94
rs_5	46.26	51.88	38.31	51.74	38.57	49.82	49.71	49.69	49.83	51.12
rs_6	46.33	51.86	47.44	51.06	38.58	49.83	50.17	50.01	49.38	51.22
rs_7	45.74	51.53	46.20	50.49	38.67	49.34	49.54	48.41	48.45	51.49
rs_8	43.98	49.63	39.28	47.95	37.89	48.09	48.09	47.08	47.49	50.09
rs_9	41.51	47.29	45.04	45.53	38.37	46.41	46.80	45.15	45.85	48.43

TABLE 4

Accuracy (%) for several classifiers under various experimental paradigms for classifying individual activities, after voting. Best results highlighted.

	Other Classifiers					kNN				
	SVM-linear	SVM-RBF	NaiveBayes	ANN	CkNN	Euclidean	Xing	LFDA	SDA	Bipart
cv	74.02	76.30	61.14	78.59	57.66	74.93	72.33	74.41	71.24	81.64
lopo	74.73	77.26	59.96	79.63	58.21	75.32	72.85	77.28	72.06	87.05
rs_1	74.89	77.66	61.73	79.00	58.39	75.09	72.37	74.27	71.70	80.92
rs_2	74.03	77.17	62.68	79.14	57.54	75.61	73.09	73.83	71.91	82.41
rs_3	73.47	75.53	64.38	77.34	57.61	74.80	71.45	72.25	70.67	80.90
rs_4	73.51	75.90	63.03	76.80	57.89	74.89	71.83	72.26	70.97	80.30
rs_5	72.45	74.36	63.12	75.86	57.18	73.73	70.66	70.99	69.92	79.18
rs_6	72.41	74.82	64.75	75.09	57.33	74.01	70.91	70.32	70.12	79.16
rs_7	72.23	74.82	64.91	74.65	57.54	73.88	70.57	70.15	70.08	79.79
rs_8	71.11	73.97	69.02	73.83	56.74	72.94	69.70	69.42	69.17	78.00
rs_9	68.96	72.44	69.69	72.21	56.75	71.48	68.35	67.36	68.41	77.02

TABLE 5

Accuracy (%) for several classifiers under experimental paradigms for classifying activity categories, without voting. Best results highlighted.

7.3 Discussion and Future Directions

Despite the difficulty of classifying activities, as shown in the confusion matrices, Bipart with kNN outperforms other classifiers overall.

The performance of kNN in Euclidean space suggests that using 3NN as a classifier in Bipart space may limit the accuracy of the Bipart method. Results may improve if more sophisticated classifiers, such as SVMs or neural networks, were used in Bipart space. As Bipart allows for any classification method to be adapted for the block

classification problem, future work may involve other classifiers. Comparing performance in both Euclidean and Bipart space will help demonstrate the utility of bag-level information.

Categorization improves performance, but as seen in the confusion matrices, there is still some overlap between categories, particularly between exercise and games. Choosing a different categorization, for example based on MET level, may improve results [21]. Noting that these categorizations are arbitrary, perhaps

	Other Classifiers					kNN				
	SVM-linear	SVM-RBF	NaiveBayes	ANN	CkNN	Euclidean	Xing	LFDA	SDA	Bipart
cv	76.03	78.12	63.55	81.91	57.66	78.49	78.49	80.02	79.28	81.64
lopo	77.02	79.88	63.13	83.22	58.21	79.78	79.36	84.84	79.66	87.05
rs_1	76.41	80.14	63.47	81.87	58.39	78.14	78.49	80.50	78.54	80.92
rs_2	76.28	79.25	65.21	82.91	57.54	80.43	80.53	81.00	80.01	82.41
rs_3	75.61	77.95	66.21	80.38	57.61	79.56	80.00	79.50	77.71	80.90
rs_4	75.39	77.76	65.47	79.99	57.89	79.49	79.02	80.09	78.63	80.30
rs_5	73.86	76.49	65.21	78.31	57.18	77.83	77.42	78.13	76.71	79.18
rs_6	73.86	76.67	66.66	77.34	57.33	78.46	78.09	78.20	77.40	79.16
rs_7	73.50	77.09	66.71	76.57	57.54	78.14	77.24	77.07	76.81	79.79
rs_8	72.06	75.65	70.32	75.52	56.74	77.00	76.94	76.49	76.53	78.00
rs_9	70.03	73.86	70.69	73.21	56.75	75.34	74.79	74.29	75.10	77.02

TABLE 6

Accuracy (%) under various experimental paradigms for classifying activity categories, after voting. Best results highlighted.

	Lying Rest	Reading	Watching TV	Searching Internet	Sweeping	Vacuuming	Slow Track Walking	Brisk Track Walking	Walking w/ 10 lb Backpack	Track Running	Nintendo Wii	Light Space	Wall Light Space	Dance Dance Revolution	Trazer	Playing Catch	Soccer around Cones	Sport Wall	Workout Video	
Lying Rest	1922	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Reading	225	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Watching TV	184	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Searching Internet	189	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sweeping	74	0	7	0	76	17	0	0	0	0	20	12	5	0	0	7	0	0	7	
Vacuuming	81	0	0	0	25	50	0	0	0	0	20	0	0	0	0	0	7	0	5	
Slow Track Walking	28	0	0	0	10	0	153	24	10	0	0	0	0	0	0	0	6	0	0	
Brisk Track Walking	34	0	0	0	0	0	41	174	17	5	3	0	0	0	0	0	0	0	0	
Walking with 10 lb Backpack	0	0	0	0	0	0	55	55	69	0	0	0	0	0	0	0	0	0	5	
Track Running	0	0	0	0	0	0	16	31	5	52	0	0	0	0	0	0	4	0	0	
Nintendo Wii	140	5	0	0	7	7	0	0	0	0	45	0	12	0	5	0	0	0	10	
Light Space	32	0	0	0	0	0	0	0	0	0	10	142	11	10	5	5	12	7	0	
Wall Light Space	25	0	0	0	5	0	5	0	0	0	5	45	87	0	0	0	0	10	0	
Dance Dance Revolution	80	0	0	0	16	0	5	0	0	0	15	10	5	20	5	15	10	0	5	
Trazer	5	0	0	0	5	0	0	0	0	0	5	10	0	0	161	0	0	0	0	
Playing Catch	33	0	0	0	15	5	0	0	0	0	5	20	20	5	17	45	0	15	0	
Soccer around cones	25	0	0	0	20	15	10	5	5	0	5	15	5	0	0	5	56	15	0	
Sport Wall	19	0	0	0	0	0	0	0	0	0	0	10	0	5	5	0	3	141	5	
Workout Video	97	0	0	0	0	10	0	0	0	0	12	10	0	0	5	0	0	10	45	

TABLE 7

Confusion matrix for activity data. Only Bipart and LOPO experiments considered. Rows represent true class, and columns represent predicted class.

	Sedentary	Chores	Locomotion	Interactive Video Games	Exercise and Sports
Sedentary	2520	0	0	0	0
Chores	114	228	0	57	14
Locomotion	50	0	742	5	0
Interactive Video Games	172	31	0	789	27
Exercise and Sports	116	68	15	97	442

TABLE 8

Confusion matrix for categorized data. Only LOPO experiments considered. Rows represent true class, and columns represent predicted class.

deriving natural categories, through clustering or other techniques, may improve classification.

The linear regression model used in this study may limit MET prediction performance, as even with perfect classification, error still exists.

Though multiple linear regression models allow more accuracy than one, they are still unable to cope with nonlinear relationships, and counts and METs may not be linear. Though it is outside the scope of the current project, future work may allow for non-linear regression models to be used for energy estimation. This may include kernel support vector regression, neural networks, and regression methods used in Bipart space.

8 CONCLUSION

This study proposes a novel distance metric learning method which utilizes block-level constraints. The Bipart method exploits block structure, which is assumed to be known for both the training and the test set. Two distance metrics, learned from both test and training sets, are combined into the Bipart metric, and a kNN classifier is used. Experiments show that Bipart performs favorably compared to other classifiers and distance metrics, especially in LOPO and low-information conditions. These results demonstrate the utility of the Bipart method on datasets which contain feature vectors known to belong to the same class.

9 ACKNOWLEDGMENTS

This study was partially supported by a grant from the National Institutes of Health (R21HL093407) to develop novel approaches to monitor physical activity in children.

REFERENCES

- [1] Deng Cai, Xiaofei He, and Jiawei Han. Semi-supervised discriminant analysis. *ICCV*, pages 1–7, 2007.
- [2] Scott E. Crouter, Kurt G. Clowers, and David R. Bassett Jr. A novel method for using accelerometer data to predict energy expenditure. *J. of Applied Physiology*, 100(4):1324–1331, April 2006.
- [3] Scott E. Crouter, Magdalene Horton, and David R. Bassett Jr. Use of a two-regression model for estimating energy expenditure in children. *Med Sci Sports Exerc*, 44(6):1177–85, 2012.
- [4] Scott E. Crouter, Erin Kuffel, Jere D. Haas, Edward A. Frongillo, and David R. Bassett Jr. Refined 2-regression model for the actigraph accelerometer. *Med Sci Sports Exerc*, 42(5):1029–37, 2010.
- [5] Suzanne M. de Graauw, Janke F. de Groot, Marco van Brussel, Marjolein F. Streur, and Tim Takken. Review of prediction models to estimate activity-related energy expenditure in children and adolescents. *International Journal of Pediatrics*, 111, 2010.
- [6] Thomas G. Dietterich, Richard H. Lathrop, and Tomás Lozano-Pérez. Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89(1-2):31–71, January 1997.
- [7] Patty S. Freedson, Kate Lyden, Sarah Kozey-Keadle, and John Staudenmayer. Evaluation of artificial neural network algorithms for predicting mets and activity type from accelerometer data: Validation on an independent sample. *J. of Applied Physiology*, 111, 2011.
- [8] Patty S. Freedson, Edward Melanson, and John Sirard. Calibration of the computer science and applications, inc. accelerometer. *Med Sci Sports Exerc*, 5, 1998.
- [9] Yuichi Fujiki, Panagiotis Tsiamyrtzis, and Ioannis Pavlidis. Making sense of accelerometer measurements in pervasive physical activity applications. In *CHI '09*, New York, NY, USA, 2009. ACM.
- [10] Jacob Goldberger, Sam Roweis, Geoff Hinton, and Ruslan Salakhutdinov. Neighbourhood components analysis. *NIPS*, 2005.
- [11] Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques 3rd Edition*. Morgan Kaufmann Publishers Inc., 2011.
- [12] Jonathan Lester, Tanzeem Choudhury, and Gaetano Borriello. A practical approach to recognizing physical activities. In *Proceedings of Pervasive 2006*. Springer, 2006.
- [13] Kate Lyden, Sarah L. Kozey, John W. Staudenmayer, and Patty S. Freedson. A comprehensive evaluation of commonly used accelerometer energy expenditure and met prediction equations. *European Journal of Applied Physiology*, 111, 2011.
- [14] Oded Maron and Aparna Lakshmi Ratan. Multiple-instance learning for natural scene classification. In *ICML*, pages 341–349, San Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc.
- [15] Yang Mu, Wei Ding, and Dacheng Tao. Local discriminative distance metrics ensemble learning. *Pattern Recognition*, 46(8):2337–2349, 2013.
- [16] Ryan Rifkin and Aldebaro Klautau. In defense of one-vs-all classification. *Journal of Machine Learning Research*, 5:101–141, 2004.
- [17] Sam T. Roweis and Laurence K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000.
- [18] John Staudenmayer, David Pober, Scott Crouter, David Bassett, and Patty Freedson. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *J. of Applied Physiology*, 107, 2009.
- [19] Masashi Sugiyama. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *JMLR*, 8:1027–1061, 2007.
- [20] Emmanuel Munguia Tapia, Stephen Intille, William Haskell, Kent Larson, Julie Wright, Abby King, and Robert Friedman. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In *ISWC, ISWC '07*, Washington, DC, USA, 2007. IEEE Computer Society.
- [21] Stewart G. Trost, Paul D. Loprinzi, Rebecca Moore, and Karin A. Pfeiffer. Comparison of accelerometer cut points for predicting activity intensity in youth. *Med Sci Sports Exerc*, 43(7):1360–1368, Jul 2011.
- [22] Stewart G. Trost, Weng-Keen Wong, Karen A Pfeiffer, and Yonglei Zheng. Artificial neural networks to predict activity type and energy expenditure in youth. *Med Sci Sports Exerc*, 2012 Apr 19 2012.
- [23] Jun Wang and Jean-Daniel Zucker. Solving the multiple-instance problem: A lazy learning approach. In *ICML*, pages 1119–1125. Morgan Kaufmann, 2000.

	SVM-linear	SVM-RBF	NaiveBayes	ANN	CkNN	kNN	Xing	LFDA	SDA	Bipart
Lying Rest	100.00	100.00	2.19	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Reading	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Watching TV	0.00	0.00	89.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Searching Internet	0.00	0.00	2.65	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sweeping	0.00	27.56	0.00	17.78	0.00	29.78	23.11	38.22	23.11	33.78
Vacuuming	0.00	2.66	21.28	5.32	0.00	5.32	18.62	18.62	21.28	26.60
Slow Track Walking	50.22	45.89	8.23	53.25	45.89	64.94	53.68	55.84	47.62	66.23
Brisk Track Walking	61.68	59.12	13.14	56.20	31.75	35.40	37.23	47.08	30.29	63.50
Walking with 10 lb backpack	0.00	4.89	74.46	7.61	0.00	11.96	20.65	45.11	19.57	37.50
Track Running	37.04	48.15	79.63	51.85	31.48	50.93	52.78	54.63	34.26	48.15
Nintendo Wii	0.00	0.00	0.00	0.00	0.00	0.00	3.03	3.03	6.06	19.48
Light Space	61.54	57.26	16.67	47.44	0.00	53.85	39.32	52.14	49.57	60.68
Wall Light Space	0.00	0.00	35.71	16.49	0.00	21.98	35.71	32.97	30.77	47.80
Dance Dance Revolution	8.06	5.38	8.06	5.38	0.00	2.69	9.14	18.82	0.00	10.75
Trazer	59.68	84.41	62.90	100.00	5.38	79.03	76.34	79.57	71.51	86.56
Playing Catch	42.78	45.56	51.11	48.33	0.00	9.44	19.44	45.56	11.11	25.00
Soccer around cones	35.91	44.75	41.99	69.61	0.00	24.86	38.67	50.28	27.62	30.94
Sport Wall	61.70	62.77	13.30	73.94	0.00	59.04	60.64	60.11	38.30	75.00
Workout Video	0.00	5.29	0.00	5.29	0.00	7.94	5.29	15.87	19.58	23.81

TABLE 9

Classification performance (in terms of accuracy) of each activity. Only LOPO experiments were considered. Best results highlighted.

	SVM-linear	SVM-RBF	NaiveBayes	ANN	CkNN	kNN	Xing	LFDA	SDA	Bipart
Sedentary	100.00	100.00	71.27	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Chores	0.00	31.96	93.46	51.09	0.00	35.84	29.30	50.85	37.05	55.21
Locomotion	92.35	91.72	85.07	93.73	76.54	93.73	93.10	94.48	93.85	93.10
Interactive Video Games	73.31	72.82	36.80	72.03	0.98	62.32	61.83	71.93	58.98	77.43
Exercise and Sports	23.44	27.51	29.00	42.14	2.03	38.48	39.70	52.85	41.87	59.90

TABLE 10

Classification performance (in terms of accuracy) of each activity category. Only Bipart and LOPO experiments were considered. Best results highlighted.

	CV	LOPO	RS1	RS2	RS3	RS4	RS5	RS6	RS7	RS8	RS9
SVM-linear	1.50	1.47	1.41	1.52	1.42	1.53	1.51	1.53	1.63	1.74	2.19
SVM-RBF	1.47	1.46	1.48	1.43	1.41	1.46	1.44	1.47	1.63	1.75	2.28
NaiveBayes	1.46	1.41	1.43	1.53	1.47	1.57	1.60	1.61	1.79	1.84	2.17
ANN	1.40	1.39	1.39	1.44	1.36	1.47	1.45	1.46	1.55	1.66	2.16
CkNN	2.55	2.62	2.40	2.40	2.22	2.35	2.18	2.28	2.14	2.21	2.39
kNN	2.55	2.62	2.40	2.40	2.22	2.35	2.18	2.28	2.14	2.21	2.39
Xing	1.42	1.43	1.41	1.41	1.39	1.46	1.44	1.50	1.66	1.80	2.27
LFDA	2.22	2.04	2.02	2.56	2.35	2.57	2.84	2.84	2.88	2.96	2.92
SDA	1.41	1.41	1.45	1.42	1.38	1.50	1.47	1.50	1.65	1.82	2.30
Bipart	1.42	1.37	1.37	1.43	1.37	1.45	1.46	1.48	1.58	1.68	2.12

TABLE 11

MET expenditure error for each experimental condition and each classifier. Performance is given by root mean square error (RMSE). Best results highlighted.

	SVM-linear	SVM-RBF	NaiveBayes	ANN	CkNN	kNN	Xing	LFDA	SDA	Bipart
Sedentary	0.42	1.41	0.68	0.42	0.42	0.42	1.41	2.04	1.41	0.42
Chores	2.14	1.46	1.40	1.64	2.25	2.25	1.46	2.04	1.46	1.61
Locomotion	1.64	1.43	1.64	1.63	1.99	1.99	1.43	2.04	1.43	1.64
Sports and Games	1.65	1.35	1.60	1.62	3.54	3.54	1.35	2.04	1.35	1.57
Exercise and Sports	1.62	2.04	1.67	1.63	3.65	3.65	2.04	2.04	2.04	1.59

TABLE 12

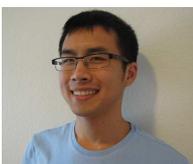
Root mean square error (RMSE) showing of each classifier for estimating METs for each activity category. LOPO results only.

- [24] Kilian Weinberger and Lawrence Saul. Distance metric learning for large margin nearest neighbor classification. *J. of Machine Learning Research*, 10:207–244, June 2009.
- [25] Sanford Weisberg. *Applied Linear Regression (3rd Edition)*, volume 528. John Wiley and Sons, 2005.
- [26] Eric Xing, Andrew Ng, Michael Jordan, and Stuart Russell. Distance metric learning, with application to clustering with side-information. In *NIPS*, pages 505–512. MIT Press, 2002.
- [27] Liu Yang, Rong Jin, Rahul Sukthankar, and Yi Liu. An efficient algorithm for local distance metric learning. In *AAAI*, pages 543–548, 2006.
- [28] Tianhao Zhang, Xuelong Li Dacheng Tao, and Jie Yang. Patch alignment for dimensionality reduction. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1299–1313, September 2009.
- [29] Zhenyue Zhang and Hongyuan Zha. Principal manifolds and nonlinear dimension reduction via local tangent space alignment. *SIAM J. of Scientific Computing*, 26:313–338, 2002.
- [30] Zhi-Hua Zhou and Min-Ling Zhang. Ensembles of multi-instance learners. In *ICML*, pages 492–502. Springer, 2003.
- [31] Zhi-Hua Zhou and Min-Ling Zhang. Multi-Instance Multi-Label learning with application to scene classification. In Bernhard Schölkopf, John C. Platt, and Thomas Hoffman, editors, *NIPS*, pages 1609–1616. MIT Press, 2006.



Yang Mu Yang Mu received his B.S. and M.S. degree from Jilin University and University of Massachusetts Boston in 2008 and 2012 respectively. He is currently pursuing his Ph.D. degree of Computer Science at the University of Massachusetts Boston in Knowledge Discovery Lab working on a general framework for efficient and effective data analysis on large-scale data. Prior to his PhD studies, he worked at Microsoft ATC as an intern for mobile education and in Nanyang Technological University as a research

assistant working on large-scale image retrieval. His research interests include online learning, distance learning and feature selection. His papers have been published on top venues such as Pattern Recognition, IEEE T-SMC part B, ACM SIGKDD and IEEE ICDM.

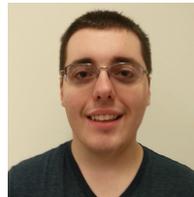


Henry Lo Henry Lo is a PhD student in Data Mining at the University of Massachusetts Boston, where he is a member of the Knowledge Discovery Lab. Prior to his work there, he received his Bachelor of Science degree in both computer science and psychology from the same institution, and worked in Tongji University in Shanghai as an EAPSI China fellow. Henry has done consulting work as a web developer and data scientist for various startups and small companies. His research interests are in data

mining, with a focus on itemset mining, temporal and spatial data, and tensor analysis.



Wei Ding Wei Ding has been an Assistant Professor of Computer Science in the University of Massachusetts Boston since 2008. Her main research interests include Data Mining, Machine Learning, Artificial Intelligence, Computational Semantics, and with applications to health science, astronomy, geosciences, and environmental sciences. She has published more than 90 referred research papers, one book, and has one patent. She is an Associate Editor of Knowledge and Information Systems (KAIS), an editorial board member of the Journal of System Education (JISE), the Journal of Big Data, and the Social Network Analysis and Mining Journal. She is an IEEE senior member and an ACM member. Her research projects are currently sponsored by NASA and DOE.



Kevin Amaral Kevin Amaral is a Research Assistant at the University of Massachusetts Boston, where he has been a member of the Knowledge Discovery Lab since the Summer of 2012, when he worked as an REU intern for Professor Wei Ding. In the Summer of 2013, he worked as an REU intern at the Artificial Intelligence Lab at the University of Houston-Downtown under the guidance of Professor Ping Chen. Kevin has pursued many teaching and mentorship opportunities as an undergrad, as a guest lecturer and facilitated study group leader. He is anticipating his Bachelor of Science degree in computer science at the end of December 2013 at the University of Massachusetts Boston. His research interests include data mining, classification, time series data, and artificial intelligence.



Scott Crouter Scott Crouter is an Assistant Professor in the Department of Kinesiology, Recreation and Sport Studies at The University of Tennessee Knoxville. His main research area includes measuring physical activity and energy expenditure in adults and children using devices such as accelerometers, pedometers and heart rate monitors. Much of his work has been focused on improving how accelerometers are used to estimate energy expenditure in free-living individuals and has developed novel techniques for estimating energy expenditure with accelerometers.