# Human-Activity Monitoring and Recognition Using Motion Sensors

Faisal Sikder and Dilip Sarkar

## I. Linear Regression Model

Let $X = [X_1, X_2, .....X_m]$ be a matrix of $n$ rows and $m$ columns, where each $X_j = [x_{1j}, x_{2j}, \cdots, x_{nj}]^T$, for $1 \le j \le m$, is a column of $n$ elements. Let $y = [y_1, y_2, ......, y_n]^T$ be a column vector of $n$ elements.

Let a linear model $f(\cdot)$ relate element $y_i$ of $y$ with $i$th row, $[x_{i1}, x_{i2}, \cdots, x_{im}]$, of $X$. Mathematically,

$$y_i = f(x_{i1}, x_{i2}, \cdots, x_{im}) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im} \quad (1)$$

If we denote $\beta = [\beta_0, \beta_1, ......\beta_m]^T$ as a column vector, Equation 1 can be rewritten as,

$$y_i = [1, x_{i1}, x_{i2}, \cdots, x_{im}]\beta.$$

Inserting a '1' before the first element of each row of $X$, we get a matrix of $n$ rows and $(m+1)$ columns, $\boldsymbol{X}$ = [1,$X$]. Now using matrix notation we can write,

$$y = f(\boldsymbol{X}) = \boldsymbol{X}\beta \quad (2)$$

Suppose $y$ and $X$ are obtained from observing a system, and we want to estimate a model $\hat{f}(\cdot)$ from the data. This is equivalent to estimating values of $\beta$ from $y$ and $\boldsymbol{X}$. Let $\hat{\beta}$ denote an estimate of $\beta$, that is, $\hat{\beta} = [\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_m]$. Now the approximation of $y$ from $\boldsymbol{X}$ using $\hat{f}(\cdot)$ (that is, $\hat{\beta}$) can be written as,

$$\hat{y} = \hat{f}(\boldsymbol{X}) = \boldsymbol{X}\hat{\beta} \quad (3)$$

The selection of the parameters $\hat{\beta}$ for approximation of the function $\hat{f}(\cdot)$ depends on the minimization of errors,

$$error_i = y_i - [1, x_{i1}, x_{i2}, \cdots, x_{im}]\hat{\beta}. \quad (4)$$

By far the most popular method for selection of $\hat{\beta}$ is that minimizes the *residual sum of squares* [2].

$$RSS(\hat{\beta}) = \sum_{i=1}^{n} \left( y_i - [1, x_{i1}, x_{i2}, \cdots, x_{im}]\hat{\beta} \right)^2 \quad (5)$$

Let us denote mean of $RSS(\hat{\beta})$ as $\bar{e}$,

$$\bar{e} = RSS(\hat{\beta})/n \quad (6)$$

Being a quadratic function of the parameters $\hat{\beta}$, its minimum always exits. The minimum is unique if $\boldsymbol{X^T X}$ is nonsingular, and the corresponding approximation of $\beta$ is given by

$$\hat{\beta} = (\boldsymbol{X^T X})^{-1} \boldsymbol{X}^T y. \quad (7)$$

Faisal Sikder and Dilip Sarkar are with the Department of Computer Science, University of Miami. E-Mails: f.sikder@umiami.edu and sarkar@miami.edu

IPDPS 2014: Tuesday 4th August, 2015

Above result can be found in any standard book on matrix computation, but completeness, next we state it as a theorem.

**Theorem 1.** (Least square error model [3] ) *The linear model in Equation 2 that minimizes sum of squared error (given by Equation 5) always has a solution. The solution is unique if and only if $\boldsymbol{X}^T\boldsymbol{X}$ is nonsingular.*

□

From this estimated model $\hat{f}(.)$ we can get an estimate of $y_k$ from a new observation $x_k$. Expressing in vector-matrix notation,

$$\hat{y} = \boldsymbol{X}\hat{\beta}. \quad (8)$$

It is also important to recall that vector $\hat{\beta}$ points in the steepest uphill direction in the input subspace [2]. Therefore, two characterizing features of the data are $\hat{\beta}$ and $RSS(\hat{\beta})$. In this study we will use both of these features.

Before we discuss their applications for identification of daily activities, we briefly introduce entropy, KL-distance, and establish some of its relation to *log-sum measures* of two sequences of positive numbers.

## II. Kullback-Leibler divergence and Log-Sum inequality

### A. Entropy and KL-divergence

Let $p = \{p_1, p_2, \cdots, p_m\}$ and $q = \{q_1, q_2, \cdots, q_m\}$ be two probability mass distributions. Note that $p$ and $q$ satisfy all axioms of probability mass distributions, including$\sum_{i=1}^{m} p_i = \sum_{i=1}^{m} q_i = 1$.

**Definition 1.** *(*Entropy*) The* entropy $H(p)$ *of a probability mass distribution* $p$ *is defined by*

$$H(p) = -\sum_{i=1}^{m} p_i \log p_i \quad (9)$$

□

**Definition 2.** *(*KL-distance *[1]) The* Kullback-Leibler distance *or* KL-distance *or* relative entropy *between two probability mass distributions* $p$ *and* $q$ *is defined as,*

$$D(p||q) = \sum_{i=1}^{m} p_i \log \frac{p_i}{q_i} \quad (10)$$

□

**Theorem 2.** *(*KL-distance inequality [1]*) For two probability mass distributions* $p$ *and* $q$

$$D(p||q) \ge 0. \quad (11)$$

*Equality holds if and only if $p_i = q_i$ for all $1 \leq i \leq m$.* □

While above theorem is used in many applications to measure entropy distance between probability mass distributions $p$ and $q$, it should be noted that the measure is not symmetric, that is, $D(p||q) \neq D(q||p)$ unless $p$ and $q$ are identical. For application that benefits from a symmetric measure, sum of $D(p||q)$ and $D(q||p)$ is used, and the corresponding inequality is

$$D(p||q) + D(q||p) \geq 0 \tag{12}$$

Next we introduce a symmetric log-sum distance measure for two sequences of positive numbers of identical length.

### B. Log-Sum Distance

**Definition 3.** (Log-Sums distance) *For two sequences of positive numbers $U = <u_1, u_2, \cdots, u_m>$ and $V = <v_1, v_2, \cdots, v_m>$ log-sum distance of $U$ and $V$ is defined as,*

$$LD(U||V) = \sum_{i=1}^{m} u_i \log \frac{u_i}{v_i} + \sum_{i=1}^{m} v_i \frac{v_i}{u_i} \tag{13}$$

Next we show that $LD(U||V)$ is nonnegation.

**Theorem 3.** (Log-sums inequality) *For two sequences of positive numbers $U = <u_1, u_2, \cdots, u_m>$ and $V = <v_1, v_2, \cdots, v_m>$,*

$$LD(U||V) \geq 0 \tag{14}$$

*with equality if and only if $u_i = v_i$ for all $1 \leq i \leq m$.*

*Proof.* Assume without loss of generality that $u_i > 0$ and $v_i > 0$. We start with the definition of $LD(U||V)$ in Equation 13,

$$\begin{aligned} LD(U||V) &= \sum_{i=1}^{m} u_i \log \frac{u_i}{v_i} + \sum_{i=1}^{m} v_i \log \frac{v_i}{u_i} \\ &= \sum_{i=1}^{m} (u_i - v_i) \log \frac{u_i}{v_i} \end{aligned} \tag{15}$$

We show that value of each term of the sum is $(i)$ zero when $u_i = v_i$ and $(ii)$ greater than zero when $u_i \neq v_i$. Thus, the sum cannot be negative and, moreover, it is greater than zero if there exist at least one pair of $u_i$ and $v_i$ such that $u_i \neq v_i$.

For each term of the sum in Equation 15, we have to consider two cases: $u_i = v_i$, and $u_i \neq v_i$.
Case 1 ($u_i = v_i$): In this case, since $(u_i - v_i) = 0$ and $\log(u_i/u_i) = 0$, we have $(u_i - v_i)\log\frac{u_i}{v_i} = 0$.
Case 2 ($u_i \neq v_i$): In this case we have to consider two situations, $u_i < v_i$, and $u_i > v_i$. If $u_i < v_i$, both $(u_i - v_i)$ and $\log(u_i/v_i)$ are negative numbers, and hence, their product is greater than zero. On the other hand, if $u_i > v_i$, both $(u_i - v_i)$ and $\log(u_i/v)$ are greater than zero, and hence, their product is also greater than zero.

Thus, the sum in the right-side of Equation 15 is zero, if $u_i = v_i$, for all $1 \leq i \leq m$. On the other hand, if for one or more terms of the right-hand side of Equation 15 is greater than zero, the right-hand side of the equation is greater than zero. This completes the if part of the proof.

For the only if part of the proof it is easy to show that if sum is zero, then each individual term must be zero. Because if that is not the case then one or more terms of the sum must be negative; in that case $(u_i - v_i)$ and $\log(u_i/v_i)$ must have opposite signs, which is impossible. Similar arguments hold for the case when the sum is greater than zero. □

Next we establish a relation between $LD(U||V)$ and KL-distance.

**Theorem 4.** (Log-Sums distance and KL-distance relation) *For two sequences of positive numbers $U$ and $V$,*

$$LD(U||V) = (u - v)\log\frac{u}{v} + uD(p||q) + vD(q||p) \tag{16}$$

*where $u = \sum_{i=1}^{m} u_i$, $v = \sum_{i=1}^{m} v_i$, $p_i = u_i/u$, and $q_i = v_i/v$ for all $1 \leq i \leq m$.*

Note that $p = \{p_1, p_2, \cdots, p_m\}$ and $\sum_{i=1}^{m} p_i = 1$. Also, $q = \{q_1, q_2, \cdots, q_m\}$ and $\sum_{i=1}^{m} q_i = 1$. While $p$ and $q$ are not probability mass distributions, they can be used for measuring $D(p||q)$ and $D(q||p)$ because sum of the terms in $p$ (and $q$) is 1.

*Proof.* Since $u_i = up_i$, we have

$$\begin{aligned} \sum_{i=1}^{m} u_i \log \frac{u_i}{v_i} &= \sum_{i=1}^{m} up_i \log \frac{up_i}{vq_i} \\ &= u \sum_{i+1}^{m} p_i \log \frac{u}{v} + u \sum_{i=1}^{m} p_i \log \frac{p_i}{q_i} \\ &= u \log \frac{u}{v} + uD(p||q) \end{aligned} \tag{17}$$

Similarly, it can be shown that,

$$\sum_{i=1}^{m} v_i \log \frac{v_i}{u_i} = v \log \frac{v}{u} + vD(q||p) \tag{18}$$

For completing the proof, first we use the definition of $LD(U||V)$ and then Equations 17 and 18.

$$\begin{aligned} LD(U||V) &= \sum_{i=1}^{m} u_i \log \frac{u_i}{v_i} + \sum_{i=1}^{m} v_i \log \frac{v_i}{u_i} \\ &= u \log \frac{u}{v} + uD(p||q) + v \log \frac{v}{u} + vD(q||p) \\ &= (u - v) \log \frac{u}{v} + uD(p||q) + vD(q||p) \end{aligned}$$

This completes the proof. □

**Corollary 1.** (Log-sum and KL-distance inequality) *For $U, V, u, v, p$, and $q$ as defined earlier,*

$$LD(U||V) \geq uD(p||q) + vD(q||p). \tag{19}$$

*Proof.* Since for any two positive numbers $u$ and $v$, $(u - v)\log(u/v) \geq 0$ (see proof of Theorem 3), the proof of the inequality follows immediately from Theorem 4. □

In the next section we present methods that utilize the results in the Sections I and II-B. For identification of human activities.

## III. Activity Detection Algorithms

Let $A = \{u, d, w\}$ be the set of activities *walking upstairs*, *walking downstairs*, and *walking forward on a level floor*, respectively. Let the set $\{1, 2, \cdots, m\}$ be denoted by $M$. During each activity $k$, wireless sensing units with *inertial measurement units* (IMUs) are connected to different locations on the body of a person for collecting motion data. Each sensor collects $m$ time series for a period of interest.

Let $X^u$, $X^d$ and $X^w$ be the observed time series data for activities $u$, $d$, and $w$ respectively. For $k \in A$, $m$-time series are denoted as $X^k = [X^k_1, X^k_2, \cdots, X^k_m]$. When $X^k_l$, for $k \in M$, is removed from $X^k$ and a unit column vector $\mathbf{1}$ of appropriate size is appended as the first column, let the new matrix be denoted as $\boldsymbol{X^k_{l'}}$. That is, $\boldsymbol{X^k_{l'}} = [\mathbf{1}, X^k_1, \cdots, X^k_{l-1}, X^k_{l+1}, \cdots, X^k_m]$.

In this study, we use the method described in Section I for estimating $m$ linear models $\hat{f}^k_l(.)$, for each $k \in A$ and each $l \in M$, to relate $X^k_l$ with $\boldsymbol{X^k_{l'}}$. Thus,

$$\hat{X}^k_l = \hat{f}^k_l(\boldsymbol{X^k_{l'}}) = \boldsymbol{X^k_{l'}} \hat{\beta}^k_l, \tag{20}$$

where

$$\hat{\beta}^k_l = ((\boldsymbol{X^k_{l'}})^T \boldsymbol{X^k_{l'}})^{-1} (\boldsymbol{X^k_{l'}})^T \boldsymbol{X^k_l}.$$

Recall that $\hat{\beta}^k_l$ points in the steepest uphill direction and minimizes residual error $RSS(\hat{\beta}^k_l)$. For each activity $k \in A$ we have $m$ steepest uphill direction vectors $\hat{\beta}^k_l = [\hat{\beta}^k_{l0}, \hat{\beta}^k_{l1}, \cdots, \hat{\beta}^k_{lm}]$, for $l \in M$. Also, corresponding to each $\hat{\beta}^k_l$ we have a mean of $RSS(\hat{\beta}^k_l)$ value $\bar{e}^k_l$. We will use these values to study several activity monitoring and recognition algorithms.

### A. KL-distance Algorithms

In this section we present two versions of KL-distance algorithms; One for asymmetric distance and the other for symmetric measures.

*1) KL-Distance algorithms:* The algorithms in this subsection consider $D(p||q)$, $D(q||)$, and $D(p||q)+D(q||)$ for activity identification.

*2) Symmetric and Weighted KL-distance Algorithm:* This version of the algorithm considers $uD(p||q) + vD(q||)$ for activity identification.

### B. Log-Sum Algorithm

This version of the algorithm considers $DL(U||V)$ for activity identification.

### C. Angle Similarity Algorithm

## References

[1] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, 2nd edition, 2006.

[2] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, 2nd edition, 2008.

[3] G.W. Stewart. *Introduction to Matrix Computations*. Academic Press, 1st edition, 1973.