Conditional mixture model for modeling attributed dyadic data

Loc Nguyen
Independent scholar, Vietnam
Email: ng_phloc@yahoo.com
Homepage: www.locnguyen.net

Abstract

Dyadic data contains co-occurrences of objects, which is often modeled by finite mixture model which in turn is learned by expectation maximization (EM) algorithm. Objects in traditional dyadic data are identified by names, causing the drawback which is that it is impossible to extract implicit valuable knowledge under objects. In this research, I propose the so-called attributed dyadic data (ADD) in which each object has an informative attribute and each co-occurrence of two objects is associated with a value. ADD is flexible and covers most of structures / forms of dyadic data. Conditional mixture model (CMM), which is a variant of finite mixture model, is applied into learning ADD. Moreover, a significant feature of CMM is that any co-occurrence of two objects is based on some conditional variable. As a result, CMM can predict or estimate co-occurrence values based on regression model, which extends applications of ADD and CMM.

Keywords: dyadic data, co-occurrence data, attributed dyadic data (ADD), mixture model, conditional mixture model (CMM), regression model.

1. Introduction to dyadic data and mixture model

Suppose data has two parts such as hidden part \( X \) and observed part \( Y \) and we only know \( Y \). A relationship between random variable \( X \) and random variable \( Y \) is specified by the joint probabilistic density function (PDF) denoted \( f(X, Y \mid \Theta) \) where \( \Theta \) is parameter. Given sample \( \{Y_1, Y_2, \ldots, Y_N\} \) whose all \( Y_i \) (s) are mutually independent and identically distributed (iid), it is required to estimate \( \Theta \) based on such sample whereas \( X \) is unknown. Expectation maximization (EM) algorithm is applied to solve this problem when only \( Y_i \) (s) are observed. EM has many iterations and each iteration has two steps such as expectation step (E-step) and maximization step (M-step). At some \( t^{th} \) iteration, given current parameter \( \Theta^{(t)} \), the two steps are described as follows:

E-step:

The expectation \( Q(\Theta \mid \Theta^{(t)}) \) is determined based on current parameter \( \Theta^{(t)} \), according to equation 1.1 (Nguyen, Tutorial on EM tutorial, 2020, p. 50).

\[
Q(\Theta \mid \Theta^{(t)}) = \sum_{i=1}^{N} \int f(X \mid Y_i, \Theta^{(t)}) \log(f(X, Y_i \mid \Theta)) dX
\]

M-step:

The next parameter \( \Theta^{(t+1)} \) is a maximizer of \( Q(\Theta \mid \Theta^{(t)}) \) with subject to \( \Theta \). Note that \( \Theta^{(t+1)} \) will become current parameter at the next iteration (the \( (t+1)^{th} \) iteration).

EM algorithm will converge after some iterations, at that time we have the estimate \( \Theta^{(t)} = \Theta^{(t+1)} = \Theta^{*} \). Note, the estimate \( \Theta^{*} \) is result of EM.

Given two finite sets \( X = \{x_1, x_2, \ldots, x_N\} \) and \( Y = \{y_1, y_2, \ldots, y_M\} \) with note that \( x_i \) (s) and \( y_j \) (s) represent \( X \)-objects and \( Y \)-objects, respectively; exactly, they are names of objects. An observational pair \( (x_i, y_j) \in X \times Y \) is called a co-occurrence of \( x_i \) and \( y_j \). Dyadic data or co-occurrence data \( S \) contains these co-occurrences with note that a co-occurrence \( (x_i, y_j) \) can exist more than one time. So, each co-occurrence \( (x_i, y_j) \) is indexed by an index \( r \). As a result, each co-occurrence is denoted by the triple \( (x_i, y_j, r) \) and we have (Hofmann & Puzicha, 1998, p. 1):
\[ S = \{ (x_i, y_j, r) : 1 \leq r \leq |S| \} \]  

(1.2)

Where,

\[ x_i \in \mathcal{X} = \{ x_1, x_2, ..., x_{|x|} \} \]

\[ y_j \in \mathcal{Y} = \{ y_1, y_2, ..., y_{|y|} \} \]

Of course, the size of \( S \) is \( |S| \). As a convention, \( x(r) \) and \( y(r) \) indicate that \( X \)-object and \( Y \)-object at the \( r^{th} \) co-occurrence are \( x_i \) and \( y_j \), respectively. Thus, the triplet \( (x_i, y_j, r) \) can be denoted as \( (x(r), y(r), r) \).

Suppose each co-occurrence \( (x_i, y_j) \) belongs to a latent variable \( C \) and \( C \) has \( K \) values \( c_k \) (s). These values \( c_k \) (s) are called classes or aspects and thus, mixture model for dyadic data is also called aspect model or latent class model which aims to discover the latent variable \( C \). Without loss of generality, let \( c_k = k \) where \( k = 1, 2, ..., K \). The random variable \( C \) has discrete distribution such that every value has an associated probability \( \alpha_k \). Of course, there are \( K \) probabilities \( \alpha_k \) (s). There are three kinds of dyadic mixture model (Hofmann & Puzicha, Statistical Models for Co-occurrence Data, 1998, p. 2) for dyadic data such as symmetric mixture model (SMM), asymmetric mixture model (AMM), and product-space mixture model (PMM). In this section, EM algorithm is applied to build up these mixture models.

SMM is defined as follows (Hofmann & Puzicha, Statistical Models for Co-occurrence Data, 1998, p. 2):

\[ P(x_i, y_j | \Theta) = \sum_{k=1}^{K} \alpha_k P(x_i, y_j | k) = \sum_{k=1}^{K} \alpha_k P_{il|k} q_{j|k} \]  

(1.3)

By applying EM, given dyadic sample \( S \), at some \( t^{th} \) iteration, given current parameter \( \Theta(t) = (\alpha_k(t), p_{il|k}(t), q_{j|k}(t))^T \), the next parameter \( \Theta(t+1) = (\alpha_k(t+1), p_{il|k(t+1)}, q_{j|k(t+1)})^T \) of SMM is calculated at M-step as follows:

\[ \alpha_k(t+1) = \frac{\sum_{i=1}^{X} \sum_{j=1}^{Y} n(x_i, y_j) P(k|x_i, y_j, \Theta(t))}{\sum_{j=1}^{Y} \sum_{i=1}^{X} n(x_i, y_j) P(k|x_i, y_j, \Theta(t))} \]  

(1.4)

\[ p_{il|k}(t+1) = \frac{\sum_{j=1}^{Y} n(x_i, y_j) P(k|x_i, y_j, \Theta(t))}{\sum_{i=1}^{X} \sum_{j=1}^{Y} n(x_i, y_j) P(k|x_i, y_j, \Theta(t))} \]  

(1.5)

\[ q_{j|k}(t+1) = \frac{\sum_{i=1}^{X} n(x_i, y_j) P(k|x_i, y_j, \Theta(t))}{\sum_{i=1}^{X} \sum_{j=1}^{Y} n(x_i, y_j) P(k|x_i, y_j, \Theta(t))} \]  

(1.6)

Where \( n(x_i, y_j) \) denotes the number of co-occurrences \( (x_i, y_j) \). The conditional probability \( P(k | x_i, y_j, \Theta(t)) \) of SMM is calculated at E-step according to Bayes’ rule as follows:

\[ P(k|x_i, y_j, \Theta(t)) = \frac{\alpha_k(t) p_{il|k}(t)}{\sum_{l=1}^{K} \alpha_l(t) p_{il|l}(t)} \]  

(1.7)

The mixture model of dyadic data is called asymmetric mixture model (AMM) if \( \alpha_k \) (s) are only independent from \( x_i \) or from \( y_j \). Without loss of generality, given \( \alpha_k \) (s) are only independent from \( y_j \) (of course, it is dependent on \( x_i \)), AMM is defined as follows (Hofmann & Puzicha, Statistical Models for Co-occurrence Data, 1998, p. 3):

\[ P(x_i, y_j | \Theta) = p_i q_{j|i} = p_i \sum_{k=1}^{K} \alpha_{k|i} q_{j|k} \]  

(1.8)
By applying EM, given dyadic sample $\mathcal{S}$, at some $n$th iteration, given current parameter $\Theta^{(n)} = (a_{ij}^{(n)}, p_{i}^{(n)}, q_{jk}^{(n)})^T$, the next parameter $\Theta^{(n+1)} = (a_{ij}^{(n+1)}, p_{i}^{(n+1)}, q_{jk}^{(n+1)})^T$ of AMM is calculated at M-step as follows:

$$
a_{k|i}^{(n+1)} = \frac{\sum_{j=1}^{\lvert \mathcal{Y} \rvert} n(x_i, y_j)P(k|x_i, y_j, \Theta^{(n)})}{\sum_{j=1}^{\lvert \mathcal{Y} \rvert} n(x_i, y_j)}
$$

(1.9)

$$
p_{i}^{(n+1)} = \frac{\sum_{j=1}^{\lvert \mathcal{Y} \rvert} n(x_i, y_j)}{\sum_{i=1}^{\lvert \mathcal{X} \rvert} \sum_{j=1}^{\lvert \mathcal{Y} \rvert} n(x_i, y_j)}
$$

(1.10)

$$
d_{j|k}^{(n+1)} = \frac{\sum_{i=1}^{\lvert \mathcal{X} \rvert} n(x_i, y_j)P(k|x_i, y_j, \Theta^{(n)})}{\sum_{i=1}^{\lvert \mathcal{X} \rvert} \sum_{j=1}^{\lvert \mathcal{Y} \rvert} n(x_i, y_j)P(k|x_i, y_j, \Theta^{(n)})}
$$

(1.11)

Where the conditional probability $P(k|x_i, y_j, \Theta^{(n)})$ of AMM is calculated at E-step as follows:

$$
P(k|x_i, y_j, \Theta^{(n)}) = \frac{\alpha_{k|i}^{(n)}p_{i}^{(n)}q_{j|k}^{(n)}}{\sum_{i=1}^{\lvert \mathcal{X} \rvert} \alpha_{i|j}^{(n)}p_{i}^{(n)}q_{j|k}^{(n)}}
$$

(1.12)

Product-space mixture model (PMM) is derived from SMM with a minor change that the aspect set \( \{1, 2, \ldots, K\} \) is Cartesian product of \( \mathcal{X} \)-aspect set \( \{1, 2, \ldots, K_X\} \) and \( \mathcal{Y} \)-aspect set \( \{1, 2, \ldots, K_Y\} \). In other words, the aspect space is still symmetric but is checked (stripped) according to two directions \( \mathcal{X} \) and \( \mathcal{Y} \).

$$
\{1, 2, \ldots, K\} = \{1, 2, \ldots, K_X\} \times \{1, 2, \ldots, K_Y\}
$$

(1.13)

For every \( k \) belongs to \( \{1, 2, \ldots, K\} \), there always exists a respective pair: \( k_X \in \{1, 2, \ldots, K_X\} \) and \( k_Y \in \{1, 2, \ldots, K_Y\} \). However, for each \( k_X \) or each \( k_Y \), there are many respective \( k \).

$$
k \sim \{k_X, k_Y\}
$$

(1.14)

The sign “\( \sim \)” denotes correspondence. PMM is defined as follows (Hofmann & Puzicha, Statistical Models for Co-occurrence Data, 1998, p. 4):

$$
P(x_i, y_j | \Theta) = \sum_{k=1}^{K} \alpha_k p_{i|k_X} q_{j|k_Y}
$$

(1.15)

By applying EM, given dyadic sample \( \mathcal{S} \), at some \( n \)th iteration, given current parameter $\Theta^{(n)} = (a_{ij}^{(n)}, p_{i}^{(n)}, q_{jk}^{(n)})^T$, the next parameter $\Theta^{(n+1)} = (a_{ij}^{(n+1)}, p_{i}^{(n+1)}, q_{jk}^{(n+1)})^T$ of PMM is calculated at M-step as follows:

$$
a_{k}^{(n+1)} = \frac{\sum_{i=1}^{\lvert \mathcal{X} \rvert} \sum_{j=1}^{\lvert \mathcal{Y} \rvert} n(x_i, y_j)P(k|x_i, y_j, \Theta^{(n)})}{\sum_{i=1}^{\lvert \mathcal{X} \rvert} \sum_{j=1}^{\lvert \mathcal{Y} \rvert} n(x_i, y_j)}
$$

(1.16)

$$
p_{i|k_X}^{(n+1)} = \frac{\sum_{j=1}^{\lvert \mathcal{Y} \rvert} n(x_i, y_j)P(k_X|x_i, y_j, \Theta^{(n)})}{\sum_{i=1}^{\lvert \mathcal{X} \rvert} \sum_{j=1}^{\lvert \mathcal{Y} \rvert} n(x_i, y_j)P(k_X|x_i, y_j, \Theta^{(n)})}
$$

(1.17)

$$
d_{j|k_Y}^{(n+1)} = \frac{\sum_{i=1}^{\lvert \mathcal{X} \rvert} n(x_i, y_j)P(k_Y|x_i, y_j, \Theta^{(n)})}{\sum_{i=1}^{\lvert \mathcal{X} \rvert} \sum_{j=1}^{\lvert \mathcal{Y} \rvert} n(x_i, y_j)P(k_Y|x_i, y_j, \Theta^{(n)})}
$$

(1.18)

Where (Nguyen, Learning Dyadic Data and Predicting Unaccomplished Co-Occurrent Values by Mixture Model, 2020, p. 10),
\[
P(k \mid x_i, y_j, \Theta^{(t)}) = \frac{\alpha_k^{(t)} p_{i[kx]}^{(t)} q_j^{(t)}_{k[y]}}{\sum_{k} \alpha_k^{(t)} p_{i[kx]}^{(t)} q_j^{(t)}_{k[y]}} \tag{1.19}
\]

\[
P(k_X \mid x_i, y_j, \Theta^{(t)}) = \sum_{k : k_x = k} P(k \mid x_i, y_j, \Theta^{(t)}) \tag{1.20}
\]

\[
P(k_Y \mid x_i, y_j, \Theta^{(t)}) = \sum_{k : k_y = k} P(k \mid x_i, y_j, \Theta^{(t)}) \tag{1.21}
\]

SMM, AMM, and PMM are defined in case of traditional dyadic data. The next section is main subject of this research where the so-called attributed dyadic data (ADD) is modeled by conditional mixture model (CMM).

2. Learning attributed dyadic data by conditional mixture model

In dyadic data \(S\), if each co-occurrence of \(x_i\) and \(y_j\) is associated with a value \(z\) (Hofmann, Puzicha, & Jordan, Learning from Dyadic Data, 1998, p. 1), the triple \((x_i, y_j, z, r)\) becomes the quadruplet \((x_i, y_j, z, r)\) which is called valued co-occurrence of \(x_i\) and \(y_j\). The value \(z\) is called associative value or co-occurrence value. If \(z\) is value of a variable \(Z\) then, \(Z\) is called associative variable. As a result, the sample \(S\) is called valued dyadic data. Note, \(Z\) can be univariate or multivariate (vector).

\[
S = \{(x_i, y_j, Z, r) : 1 \leq r \leq |S|\} \tag{2.1}
\]

Where,

\[
\begin{align*}
x_i & \in X = \{x_1, x_2, ..., x_{|X|}\} \\
y_j & \in Y = \{y_1, y_2, ..., y_{|Y|}\}
\end{align*}
\]

As a convention, \(Z(r)\) or \(z(r)\) indicates that the associative value at \(r^{th}\) co-occurrence is \(Z=z\). Thus, the quadruplet \((x_i, y_j, Z, r)\) can be denoted as \((x(r), y(r), Z(r), r)\).

An extension of valued dyadic data is called attributed dyadic data in every \(x_i\) has an attribute \(X_i\) and every \(y_j\) has an attribute \(Y_j\) with constraint that all \(X_i\) (s) are iid and all \(Y_j\) (s) are iid. Of course, these attributes are considered as random variables. Let \(X\) and \(Y\) be random variable representing every \(X_i\) and every \(Y_j\), respectively. Note, \(X\) and \(Y\) can be univariate or multivariate (vector), which are called attribute variable. As a result, the sample \(S\) is called attributed dyadic data (ADD).

\[
S = \{(x_i, X, y_j, Y, Z, r) : 1 \leq r \leq |S|\}
\]

As a convention, \(X_r\) and \(Y_r\) indicate that \(X\)-object attribute and \(Y\)-object attribute at the \(r^{th}\) co-occurrence, respectively whereas \(Z\) indicates associative variable at the \(r^{th}\) co-occurrence. The attributed dyadic data \(S\) is represented as follows:

\[
S = \{(X_r, Y_r, Z_r) : 1 \leq r \leq |S|\} \tag{2.2}
\]

Thus, each co-occurrence in attributed dyadic data is denoted as a triplet \((X_r, Y_r, Z_r)\). The \(X\)-object and \(Y\)-object of \(X_r\) and \(Y_r\) are denoted \(X(r)\) and \(Y(r)\) which are some \(x_i\) and \(y_j\), respectively. Here it is required to extends SMM, AMM, and PMM to represent ADD.

The joint PDF of \(X\)-object attribute \(X = (x_1, x_2, ..., x_n)^T\), \(Y\)-object attribute \(Y = (y_1, y_2, ..., y_m)^T\), aspect \(k\), and associative variable \(Z = (z_1, z_2, ..., z_o)^T\) given conditional variable \(W = (w_1, w_2, ..., w_p)^T\) is:

\[
f(X, Y, k, Z | W) = f(X, Y | k, Z, W) f(k, Z | W)
\]

Please distinguish partial random variable \(x_i\) and \(y_j\) here from object names \(x_i\) and \(y_j\). Dimensions of \(X\), \(Y\), \(Z\), and \(W\) are \(n\), \(m\), \(a\), and \(b\), respectively. We have:

\[
f(X, Y, k, Z | W) = f(X, Y | k, Z, W) f(k, Z | W)
\]

(Suppose the proposed model is symmetric such that \(f(X, Y | k) = f(X | k) f(Y | k)\))
\[ f(X|k)f(Y|k)f(k,Z|W) = f(k|W)f(X|k)f(Y|k)f(Z|k,W) \]

Thus, the joint PDF of \( X \)-object attribute \( X \), \( Y \)-object attribute \( Y \), aspect \( k \), and associative variable \( Z \) given conditional variable \( W \) is defined as follows:

\[
 f(X, Y, k, Z|W, \Theta) = f_k(W|\alpha_k)g_k(X|\beta_k)h_k(Y|\gamma_k)v_k(Z|W, \theta_k) \tag{2.3}
\]

Where,

\[
 f_k(W|\alpha_k) = f(k|W)
\]

\[
 g_k(X|\beta_k) = f(X|k)
\]

\[
 h_k(Y|\gamma_k) = f(Y|k)
\]

\[
 v_k(Z|W, \theta_k) = f(Z|k,W)
\]

Of course, \( \alpha_k \), \( \beta_k \), \( \gamma_k \), and \( \theta_k \) are partial parameters of \( f_k(W|\alpha_k) \), \( g_k(X|\beta_k) \), \( h_k(Y|\gamma_k) \), and \( v_k(Z|W, \theta_k) \), respectively. These functions are PDFs. The whole parameter is \( \Theta = (\alpha_k, \beta_k, \gamma_k, \theta_k)^T \). The PDF \( f_k(W|\alpha_k) \) implies distribution of aspect \( k \) given conditional \( W \). The whole PDF \( v_k(Z|W, \theta_k) \) implies distributions of attributes with regard to \( X \)-object, \( Y \)-object, and aspect \( k \). The PDF \( v_k(Z|W, \theta_k) \) is conditional PDF of \( Z \) given \( W \) with regard to aspect \( k \); later on we will note that it is more useful if it is considered as regression model.

According to Bayes’ rule, the conditional probability of \( k \) given \( X \)-object attribute \( X \), \( Y \)-object attribute \( Y \), associative variable \( Z \), and conditional variable \( W \) is:

\[
 P(k|X, Y, Z, W, \Theta) = \frac{f_k(W|\alpha_k)g_k(X|\beta_k)h_k(Y|\gamma_k)v_k(Z|W, \theta_k)}{\sum_{l=1}^{K}f_l(W|\alpha_l)g_l(X|\beta_l)h_l(Y|\gamma_l)v_l(Z|W, \theta_l)} \tag{2.4}
\]

Symmetric model (SMM) for attributed dyadic data is called symmetric attributed mixture model (SAMM), which is defined based on the joint PDF \( f(X, Y, k, Z|W, \Theta) \) and \( K \) aspects \( \{1, 2, \ldots, K\} \) as follows:

\[
 f(X, Y, k, Z|W, \Theta) = \sum_{k=1}^{K} f(X, Y, k, Z|W, \Theta)
\]

\[
 = \sum_{k=1}^{K} f_k(W|\alpha_k)g_k(X|\beta_k)h_k(Y|\gamma_k)v_k(Z|W, \theta_k) \tag{2.5}
\]

Recall that the parameter of SMM is \( \Theta = (\alpha_k, \beta_k, \gamma_k, \theta_k)^T \). Obviously, SAMM is an extension of conditional mixture model (CMM) (Nguyen, Conditional Mixture Model and Its Application for Regression Model, 2020) when \( f_k(W|\alpha_k) \) and \( v_k(Z|W, \theta_k) \) are based on the condition \( W \). SAMM is symmetric, which correspond to SMM. Here I focus on building up SAMM by mixture model and EM. Asymmetric model and product-space model for ADD will be mentioned later.

SAMM represented by equation 2.5 will be more specific and useful if \( f_k(W|\alpha_k), g_k(X|\beta_k), \) and \( h_k(Y|\gamma_k) \) are multinormal PDFs and \( v_k(Z|W, \theta_k) \) is regressive PDF which represents regression model. Means and covariance matrices of \( f_k(W|\alpha_k), g_k(X|\beta_k), \) and \( h_k(Y|\gamma_k) \) are \( \alpha_k = (\mu_{ak}, \Sigma_{ak})^T, \beta_k = (\mu_{bk}, \Sigma_{bk})^T, \) and \( \gamma_k = (\mu_{yk}, \Sigma_{yk})^T \).

\[
 f_k(W|\alpha_k) = (2\pi)^{-\frac{b}{2}}|\Sigma_{ak}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(W - \mu_{ak})^T\Sigma_{ak}^{-1}(W - \mu_{ak})\right) \tag{2.6}
\]

\[
 g_k(X|\beta_k) = (2\pi)^{-\frac{n}{2}}|\Sigma_{bk}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(X - \mu_{bk})^T\Sigma_{bk}^{-1}(X - \mu_{bk})\right) \tag{2.7}
\]

\[
 h_k(Y|\gamma_k) = (2\pi)^{-\frac{m}{2}}|\Sigma_{yk}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(Y - \mu_{yk})^T\Sigma_{yk}^{-1}(Y - \mu_{yk})\right) \tag{2.8}
\]

Note, the superscript “\(^T\)” denotes transposition operator for vector and matrix. The regressive PDF of \( Z \) given \( W \) denoted \( v_k(Z|W, \theta_k) \) is:
\[ v_k(Z | W, \theta_k) = (2\pi)^{-\frac{a}{2}} |\Sigma_{\theta_k}|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} (Z - \omega_k W)^T \Sigma_{\theta_k}^{-1} (Z - \omega_k W) \right) \]  

(2.9)

Where,

\[ \theta_k = (\omega_k, \Sigma_{\theta_k})^T \]

\[ \omega_k = \begin{pmatrix} \omega_{k10} & \omega_{k11} & \omega_{k12} & \cdots & \omega_{k1b} \\ \omega_{k20} & \omega_{k21} & \omega_{k22} & \cdots & \omega_{k2b} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \omega_{ka0} & \omega_{ka1} & \omega_{ka2} & \cdots & \omega_{kab} \end{pmatrix} = \begin{pmatrix} \omega_{k1}^T \\ \omega_{k2}^T \\ \vdots \\ \omega_{ka}^T \end{pmatrix} \]

(2.10)

\[ \omega_{kj} = (\omega_{kj0}, \omega_{kj1}, \omega_{kj2}, \ldots, \omega_{kj2b})^T \]

Note, \(a\) and \(b\) are dimensions of \(Z\) and \(W\). Mean and covariance matrix of \(Z\) given \(W\) are \(\omega_k W\) and \(\Sigma_{\theta_k}\), respectively. The partial parameter \(\omega_k\) is called regressive coefficient matrix, which is an \(a \times b\) matrix having \(a\) rows and \(b\) columns. Note, the product \(\omega_k W\) is:

\[ \omega_k W = \begin{pmatrix} \omega_{k10} + \sum_{l=1}^{b} \omega_{k1l} \omega_{l1} \\ \omega_{k20} + \sum_{l=1}^{b} \omega_{k2l} \omega_{l1} \\ \vdots \\ \omega_{ka0} + \sum_{l=1}^{b} \omega_{kal} \omega_{l1} \end{pmatrix} \]

(2.11)

The equation 2.11 also specifies multivariate regression function. Of course, equation 2.11 implies that:

\[ \omega_{kj}^T W = \omega_{kj0} + \sum_{l=1}^{b} \omega_{kjl} \omega_{l1} \]

By applying EM algorithm, given attributed dyadic sample \(S\), at the \(t\)th iteration of GEM, given current parameter \(\Theta^{(t)} = (\alpha^{(t)}, \beta^{(t)}, \gamma^{(t)}, \theta^{(t)})\), the conditional expectation \(Q(\Theta | \Theta^{(t)})\) of SAMM specified by equation 2.5 is:

\[ Q(\Theta | \Theta^{(t)}) = \sum_{r=1}^{|S|} \sum_{k=1}^K P(k | X_r, Y_r, Z_r, W_r, \Theta^{(t)}) \]

\[ \times \log(f_k(W_r | \alpha_k) g_k(X_r | \beta_k) h_k(Y_r | \gamma_k) v_k(Z_r | W_r, \theta_k)) \]

Note, all \(X_r\) (s) are iid represented by \(X\), all \(Y_r\) (s) are iid represented by \(Y\), all \(Z_r\) (s) are iid represented by \(Z\), and all \(W_r\) (s) are iid represented by \(W\). The \(X\)-object and \(Y\)-object of \(X_r\) and \(Y_r\) are denoted \(X(r)\) and \(Y(r)\) which are some \(x_i\) and \(y_j\), respectively. In short, we obtain:

\[ Q(\Theta | \Theta^{(t)}) = \sum_{r=1}^{|S|} \sum_{k=1}^K P(k | X_r, Y_r, Z_r, W_r, \Theta^{(t)}) \left( \log(f_k(W_r | \alpha_k)) + \log(g_k(X_r | \beta_k)) \right) \]

(2.12)

\[ + \log(h_k(Y_r | \gamma_k)) + \log(v_k(Z_r | W_r, \theta_k)) \]

Following equation 2.4, the conditional probability \(P(k | X_r, Y_r, Z_r, W_r, \Theta^{(t)})\) is calculated at E-step as follows:
\[ P(k|X_r, Y_r, Z_r, W_r, \Theta^{(t)}) = \frac{f_k(W_r \alpha_k^{(t)})g_k(X_r \beta_k^{(t)})h_k(Y_r | \gamma_k^{(t)})v_k(Z_r | W_r, \theta_k^{(t)})}{\sum_{i=1}^{K} f_i(W_r \alpha_i^{(t)})g_i(X_r \beta_i^{(t)})h_i(Y_r | \gamma_i^{(t)})v_i(Z_r | W_r, \theta_i^{(t)})} \] (2.13)

The next parameter \( \theta^{(t+1)} = (\alpha_k^{(t+1)}, \beta_k^{(t+1)}, \gamma_k^{(t+1)}, \theta_k^{(t+1)})^T \) that maximizes \( Q(\Theta|\Theta^{(t)}) \) at M-step of some \( \mathfrak{m} \)th iteration is solution of the equation formed by setting the first-order partial derivatives of \( Q(\Theta|\Theta^{(t)}) \) regarding \( \Theta = (\alpha_k, \beta_k, \gamma_k, \theta_k) \) to be zero. Because \( f_i(W|\alpha_k), g_k(X|\beta_k), \) and \( h_k(Y|\gamma_k) \) distribute normally, by referring to (Nguyen, 2020, p. 2), the next parameters \( \alpha_k^{(t+1)} = (\mu_{\alpha_k}^{(t+1)}, \Sigma_{\alpha_k}^{(t+1)})^T, \beta_k^{(t+1)} = (\mu_{\beta_k}^{(t+1)}, \Sigma_{\beta_k}^{(t+1)})^T, \) and \( \gamma_k^{(t+1)} = (\mu_{\gamma_k}^{(t+1)}, \Sigma_{\gamma_k}^{(t+1)})^T \) for SAMM are calculated at M-step as follows:

\[
\mu_{\alpha_k}^{(t+1)} = \frac{\sum_{r=1}^{[S]} P(k|X_r, Y_r, Z_r, W_r, \Theta^{(t)})W_r}{\sum_{r=1}^{[S]} P(k|X_r, Y_r, Z_r, W_r, \Theta^{(t)})} \\
\Sigma_{\alpha_k}^{(t+1)} = \frac{\sum_{r=1}^{[S]} P(k|X_r, Y_r, Z_r, W_r, \Theta^{(t)}) \left( W_r - \mu_{\alpha_k}^{(t+1)} \right) \left( W_r - \mu_{\alpha_k}^{(t+1)} \right)^T}{\sum_{r=1}^{[S]} P(k|X_r, Y_r, Z_r, W_r, \Theta^{(t)})} \\
\mu_{\beta_k}^{(t+1)} = \frac{\sum_{r=1}^{[S]} P(k|X_r, Y_r, Z_r, W_r, \Theta^{(t)})X_r}{\sum_{r=1}^{[S]} P(k|X_r, Y_r, Z_r, W_r, \Theta^{(t)})} \\
\Sigma_{\beta_k}^{(t+1)} = \frac{\sum_{r=1}^{[S]} P(k|X_r, Y_r, Z_r, W_r, \Theta^{(t)}) \left( X_r - \mu_{\beta_k}^{(t+1)} \right) \left( X_r - \mu_{\beta_k}^{(t+1)} \right)^T}{\sum_{r=1}^{[S]} P(k|X_r, Y_r, Z_r, W_r, \Theta^{(t)})} \\
\mu_{\gamma_k}^{(t+1)} = \frac{\sum_{r=1}^{[S]} P(k|X_r, Y_r, Z_r, W_r, \Theta^{(t)})Y_r}{\sum_{r=1}^{[S]} P(k|X_r, Y_r, Z_r, W_r, \Theta^{(t)})} \\
\Sigma_{\gamma_k}^{(t+1)} = \frac{\sum_{r=1}^{[S]} P(k|X_r, Y_r, Z_r, W_r, \Theta^{(t)}) \left( Y_r - \mu_{\gamma_k}^{(t+1)} \right) \left( Y_r - \mu_{\gamma_k}^{(t+1)} \right)^T}{\sum_{r=1}^{[S]} P(k|X_r, Y_r, Z_r, W_r, \Theta^{(t)})} \\
\] (2.14)

Because \( v_k(Z \mid W, \theta_k) \) is regressive PDF known as adaptive regression model (ARM), by referring to (Nguyen, 2020, pp. 5-6), the next parameter \( \theta_k^{(t+1)} = (\omega_{k}^{(t+1)}, \Sigma_{\theta_k}^{(t+1)})^T \) is calculated as follows:

\[
\omega_{k}^{(t+1)} = \left( W^T U_k^{(t)} \right)^{-1} W^T V_{k}^{(t)} \\
\Sigma_{\theta_k}^{(t+1)} = \frac{\sum_{r=1}^{[S]} P(k|X_r, Y_r, Z_r, W_r, \Theta^{(t)}) \left( Z_r - \omega_{k}^{(t+1)} W \right) \left( Z_r - \omega_{k}^{(t+1)} W \right)^T}{\sum_{r=1}^{[S]} P(k|X_r, Y_r, Z_r, W_r, \Theta^{(t)})} \\
\] (2.15)

Where,

\[
U_k^{(t)} = \begin{pmatrix}
\begin{bmatrix} u_{11}^{(t)}(k) & u_{12}^{(t)}(k) & \cdots & u_{1l}^{(t)}(k) \\
u_{21}^{(t)}(k) & u_{22}^{(t)}(k) & \cdots & u_{2l}^{(t)}(k) \\
\vdots & \vdots & \ddots & \vdots \\
u_{s1}^{(t)}(k) & u_{s2}^{(t)}(k) & \cdots & u_{sl}^{(t)}(k)
\end{bmatrix}
\end{pmatrix} \\
\] (2.16)

\[
U_r^{(t)} = w_{rl}^{(t)}P(k|X_r, Y_r, Z_r, W_r, \Theta^{(t)}), \forall r = \frac{1}{1}, \frac{l}{1}, \frac{b}{1}
\] (2.17)
\[ V_{kj}^{(t)} = \begin{pmatrix} v_{0j}^{(t)}(k) \\ v_{1j}^{(t)}(k) \\ \vdots \\ v_{kj}^{(t)}(k) \end{pmatrix}, \forall j = 1, a \] (2.20)

\[ v_{ij}^{(t)}(k) = z_{ij} P(k | x_i, y_i, z_i, w_i, \Theta^{(t)}), \forall r = 1, |\mathcal{S}|, j = 1, a \]

Note, \( a \) and \( b \) are dimensions of \( Z \) and \( W \). The product \( \omega_k^{(t+1)}W \) is calculated by following equation 2.11 and equation 2.17. The conditional probability \( P(k | x_i, y_i, z_i, w_i, \Theta^{(t)}) \) is calculated at E-step according to equation 2.13. Moreover, \( w_i \) and \( z_{ij} \) are extracted from \( x_i \) and \( z_i \) of \( \mathcal{S} \) where \( W_r = (w_{r1}, w_{r2}, \ldots, w_{rb}) \) and \( Z_r = (z_{r1}, z_{r2}, \ldots, z_{rm})^T \). Let,

\[
W = \begin{pmatrix}
1 & w_{11} & w_{12} & \cdots & w_{1b} \\
1 & w_{21} & w_{22} & \cdots & w_{2b} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & w_{|\mathcal{S}|1} & w_{|\mathcal{S}|2} & \cdots & w_{|\mathcal{S}|b}
\end{pmatrix}
\]

\[
Z = \begin{pmatrix}
z_{11} & z_{12} & \cdots & z_{1a} \\
z_{21} & z_{22} & \cdots & z_{2a} \\
\vdots & \vdots & \ddots & \vdots \\
z_{|\mathcal{S}|1} & z_{|\mathcal{S}|2} & \cdots & z_{|\mathcal{S}|a}
\end{pmatrix}
\]

Asymmetric model is not appropriate to ADD because the PDF of aspect \( k, f_i(W|\alpha_k) \) becomes impractical if it depends on both \( W \) and \( X \) as \( f_i(W | X, \alpha_k) \). However, product-space model (PMM) for ADD is still practical. PMM for attributed dyadic data is called product-space attributed mixture model (PAMM), which is defined as follows:

\[ f(X, Y, Z|W, \Theta) = \sum_{k=1}^{K} f_k(W|\alpha_k)g_k(X|\beta_{kX})h_k(Y|\gamma_{kY})v_k(Z|W, \theta_k) \] (2.22)

The parameter of PAMM is \( \Theta = (\alpha_k, \beta_{kX}, \gamma_{kY}, \theta_k)^T \) where \( \beta_{kX} = (\mu_{\beta kX}, \Sigma_{\beta kX})^T \) and \( \gamma_{kY} = (\mu_{\gamma kY}, \Sigma_{\gamma kY})^T \). Recall that the aspect set \( \{1, 2, \ldots, K\} \) in product-space model is Cartesian product of \( X \)-aspect set \( \{1, 2, \ldots, K_X\} \) and \( Y \)-aspect set \( \{1, 2, \ldots, K_Y\} \). For every \( k \) belongs to \( \{1, 2, \ldots, K\} \), there always exists a respective pair: \( k_X \in \{1, 2, \ldots, K_X\} \) and \( k_Y \in \{1, 2, \ldots, K_Y\} \). However, for each \( k_X \) or each \( k_Y \), there are many respective \( k \). Of course, PAMM is an extension of CMM.

At M-step of some \( t^{th} \) iteration, the next parameters \( \omega_k^{(t+1)} = (\mu_{\omega k}^{(t+1)}, \Sigma_{\omega k}^{(t+1)})^T \) and \( \Theta^{(t+1)} = (\mu_{\Theta}^{(t+1)}, \Sigma_{\Theta}^{(t+1)})^T \) for PAMM are as same as the ones for SAMM but the next parameters \( \beta_{kX}^{(t+1)} \) and \( \gamma_{kY}^{(t+1)} \) for PAMM are estimated particularly as follows:

\[ \mu_{\beta kX}^{(t+1)} = \frac{\sum_{r=1}^{|\mathcal{S}|} P(k | x_r, y_r, z_r, w_r, \Theta^{(t)}) x_r}{\sum_{r=1}^{|\mathcal{S}|} P(k | x_r, y_r, z_r, w_r, \Theta^{(t)})} \]

\[ \Sigma_{\beta kX}^{(t+1)} = \frac{\sum_{r=1}^{|\mathcal{S}|} P(k | x_r, y_r, z_r, w_r, \Theta^{(t)}) (x_r - \mu_{\beta kX}^{(t+1)}) (x_r - \mu_{\beta kX}^{(t+1)})^T}{\sum_{r=1}^{|\mathcal{S}|} P(k | x_r, y_r, z_r, w_r, \Theta^{(t)})} \] (2.23)
\[ \mu_{Y|k}^{(t+1)} = \frac{\sum_{r=1}^{[S]} P(k_{ij}|X_r, Y_r, Z_r, W_r, \Theta^{(t)})Y_r}{\sum_{r=1}^{[S]} P(k_{ij}|X_r, Y_r, Z_r, W_r, \Theta^{(t)})} \]

\[ \Sigma_{Y|k}^{(t+1)} = \frac{\sum_{r=1}^{[S]} P(k_{ij}|X_r, Y_r, Z_r, W_r, \Theta^{(t)}) \left( Y_r - \mu_{Y|k}^{(t+1)} \right)^T \left( Y_r - \mu_{Y|k}^{(t+1)} \right)}{\sum_{r=1}^{[S]} P(k_{ij}|X_r, Y_r, Z_r, W_r, \Theta^{(t)})} \]

Where,
\[ P(k_{X}|X_r, Y_r, Z_r, W_r, \Theta^{(t)}) = \sum_{k:k_{X}=k} P(k|X_r, Y_r, Z_r, W_r, \Theta^{(t)}) \]
\[ P(k_{ij}|X_r, Y_r, Z_r, W_r, \Theta^{(t)}) = \sum_{k:k_{ij}=k} P(k|X_r, Y_r, Z_r, W_r, \Theta^{(t)}) \]

Of course, the conditional probability \( P(k | X_r, Y_r, Z_r, W_r, \Theta^{(t)}) \) is calculated at E-step according to equation 2.13. Please see equation 1.20 and equation 1.21 to understand equation 2.25 and equation 2.26.

CMM (s) for ADD such as SAMM and PAMM can be used to estimate an unknown associative value \( Z \) given attribute variables \( (X, Y) \) and conditional variable \( W \). Let \( \hat{Z} \) be the estimate of \( Z \).
\[ \hat{Z} = E(Z|X, Y, W, \Theta) = \int_Z Zf(Z|X, Y, W, \Theta) \, dZ \]  
(2.27)

It is easy to calculate the conditional PDF of \( Z \) given \( X, Y, \) and \( W \) according to CMM as follows:
\[ f(Z|X, Y, W, \Theta) = \frac{f(X, Y, Z|W, \Theta)}{\int_Z f(X, Y, Z|W, \Theta) \, dZ} \]
\[ = \frac{\sum_{k=1}^{K} f_k(W|\alpha_k)g_k(X|\beta_k)h_k(Y|\gamma_k)v_k(Z|W, \theta_k)}{\sum_{k=1}^{K} f_k(W|\alpha_k)g_k(X|\beta_k)h_k(Y|\gamma_k)} \]  
(2.28)

The estimate \( \hat{Z} \) is resolved:
\[ \hat{Z} = \frac{\sum_{k=1}^{K} f_k(W|\alpha_k)g_k(X|\beta_k)h_k(Y|\gamma_k)E_k(Z|W, \theta_k)}{\sum_{k=1}^{K} f_k(W|\alpha_k)g_k(X|\beta_k)h_k(Y|\gamma_k)} \]  
(2.29)

Where,
\[ E_k(Z|W, \theta_k) = \int_Z Zv_k(Z|W, \theta_k) \, dZ \]

When \( v_k(Z|W, \theta_k) \) is a regressive PDF, the estimate \( \hat{Z} \) given \( X, Y, \) and \( W \) is calculated smoothly as follows:
\[ \hat{Z} = \frac{\sum_{k=1}^{K} f_k(W|\alpha_k)g_k(X|\beta_k)h_k(Y|\gamma_k)\omega_kW}{\sum_{k=1}^{K} f_k(W|\alpha_k)g_k(X|\beta_k)h_k(Y|\gamma_k)} \]
(2.30)

The product \( \omega_kW \) is calculated by following equation 2.11. In general, equation 2.30 is the ultimate estimation formula of CMM (s) for ADD.

3. Conclusions
ADD is flexible and covers most of structures / forms of dyadic data. CMM (s) for ADD such as SAMM and PAMM help researchers to model ADD in a flexible, solid, reliable manner. For example, if aspects are independent from conditional variable, each PDF \( f_k(W|\alpha_k) \) is reduced into a discrete probabilistic parameter. If each regressive PDF \( v_k(Z|W, \theta_k) \) is not formed in favor of regression model, it must be formed as probabilistic distribution like conditional multinormal PDF. An interesting application of SAMM is to build up a unified estimation model of content-based filtering, collaborative filtering, and context-awarded filtering, in
which attributes along with $g_k(X|\beta_k)$ and $h_k(Y|\gamma_k)$ are responsible for content-based filtering whereas associative variable along with $v_k(Z | W, \theta_k)$ and $f_k(W|\alpha_k)$ are responsible for collaborative filtering and context-awarded filtering. The attributes $X$ and $Y$ represent information about users and items in rating data with note that users and items are known as objects. The associative variable $Z$ represents rating values in rating data. The conditional variable $W$ represents contexts. Equation 2.30 is the ultimate formula of the unified estimation model. I hope that researchers will concern such proposed model because it is not realized yet when I compose this paper.

References
Hofmann, T., & Puzicha, J. (1998). Statistical Models for Co-occurrence Data. Massachusetts Institute of Technology, Artificial Intelligence Laboratory. MIT Publisher. Retrieved from https://dspace.mit.edu/bitstream/handle/1721.1/7253/AIM-1625.pdf?sequence=2
Hofmann, T., Puzicha, J., & Jordan, M. I. (1998). Learning from Dyadic Data. In M. J. Kearns, S. A. Solla, & D. A. Cohn (Ed.), Advances in Neural Information Processing Systems 11 (NIPS 1998). 11, pp. 466-472. Denver: MIT Press. Retrieved from https://papers.nips.cc/paper/1503
Nguyen, L. (2020, October 28). Conditional Mixture Model and Its Application for Regression Model. Preprints. doi:10.20944/preprints202010.0550.v2
Nguyen, L. (2020, November 2). Learning Dyadic Data and Predicting Unaccomplished Co-Occurrent Values by Mixture Model. Preprints, 14. doi:10.20944/preprints202011.0038.v1
Nguyen, L. (2020). Tutorial on EM tutorial. MDPI. Preprints. doi:10.20944/preprints201802.0131.v8