LEARNING DYADIC DATA AND PREDICTING UNACCOMPLISHED CO-OCCURRENT VALUES BY MIXTURE MODEL

Nguyen L1\* and Lanuza MH2

*1Loc Nguyen’s Academic Network, Vietnam*

*2Philippine Normal University, Manila, Philippines*

*\*Email: ng\_phloc@yahoo.com*

**Abstract**

Dyadic data which is also called co-occurrence data (COD) contains co-occurrences of objects where these objects are indexed and grouped into two finite sets. It is necessary to model dyadic data by applied mathematical tools because dyadic data analysis is interesting and important to many applications relating to indexed two-dimensional data such as image processing and recommendation collaborative filtering. Fortunately, finite mixture model is a solid statistical model to learn and make inference on dyadic data because mixture model is built smoothly and reliably by expectation maximization (EM) algorithm which is suitable to inherent spareness of dyadic data. This research summarizes mixture models for dyadic data, in which there are three well-known models such as symmetric mixture model (SMM), asymmetric mixture model (AMM), and product-space mixture model (PMM) which are described by beautiful mathematical proofs and explanations derived from EM algorithm. Objects in traditional dyadic data are indexed as categories and so their potential real values are concerned because of potential applications and extensions of dyadic data analysis. For instance, when each co-occurrence in dyadic data is associated with a real value, there are many unaccomplished values because a lot of co-occurrences are inexistent. In the research, these unaccomplished values are estimated as mean (expectation) of random variable given partial probabilistic distributions inside dyadic mixture model. This estimation result is solid due to support of EM algorithm.

**Keywords:** dyadic data, co-occurrence data, expectation maximization (EM) algorithm, mixture model.

**1. Introduction**

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* | Θ) where Θ is parameter. Given sample = {*Y*1, *Y*2,…, *YN*} whose all *Yi* (s) are mutually independent and identically distributed (iid), it is required to estimate Θ based on whereas *X* is unknown. Expectation maximization (EM) algorithm is applied to solve this problem when only is 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 Θ(*t*), the two steps are described as follows:

*Table 1.1: E-step and M-step of EM algorithm.*

|  |  |  |
| --- | --- | --- |
| *E-step*:  The expectation *Q*(Θ | Θ(*t*)) is determined based on current parameter Θ(*t*), according to equation 1.1 (Nguyen, 2020, p. 50).   |  |  | | --- | --- | |  | (1.1) |   *M-step*:  The next parameter Θ(*t*+1) is a maximizer of *Q*(Θ | Θ(*t*)) with subject to Θ. Note that Θ(*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 Θ(*t*) = Θ(*t*+1) = Θ\*. Note, the estimate Θ\* is result of EM. The EM algorithm shown in Table 1.1 is also called general EM or GEM.

Especially, the random variable *X* represents latent class or latent component of random variable *Y*. Suppose *X* is discrete and ranges in {1, 2,…, *K*}. As a convention, let *k*=*X*. Note, because all *Yi* (s) are iid, let random variable *Y* represent every *Yi*. The so-called probabilisticfinitemixture model is represented by the PDF of *Y*, as follows:

|  |  |
| --- | --- |
|  | (1.2) |

Where,

Note, the superscript “*T*” denotes transpose operator for vector and matrix. The *Q*(Θ | Θ(*t*)) is re-defined for finite mixture model as follows (Nguyen, 2020, p. 79):

|  |  |
| --- | --- |
|  | (1.3) |

Where,

|  |  |
| --- | --- |
|  | (1.4) |

An interesting application of finite mixture model is soft clustering. Traditional clustering methods assign a fixed cluster to every data point in sample, which means that every data point belongs exactly to one cluster. Soft clustering is more flexible when every data point belongs to more than one cluster and the degree of assignment is represented by a probability. It is easy to recognize that when mixture model is applied into soft clustering, latent class *k* represents a cluster.

Every observation in ordinary sample is univariate or multivariate but there is a case that ordinary sample becomes dyadic sample related to two sets of objects, which causes some modifications of mixture model. *Dyadic data* which is also called co-occurrence data (COD) contains co-occurrent events of objects. It is necessary to obtain statistical models to represent dyadic data and fortunately, finite mixture model is the one. Recall that EM is applied to learn mixture model. The next section focuses on mixture model for dyadic data.

**2. Mixture models for dyadic data**

Given two finite sets = {*x*1, *x*2,…, *xN*) and = {*y*1, *y*2,…, *yM*) with note that *xi* (s) and *yj* (s) represent -objects and -objects, respectively; exactly, they are names of objects. The numbers of -objects and -objects are =*N* and =*M*, respectively. For example, in information retrieval, *xi* (s) are documents and *yj* (s) are keywords. Hence, *xi* and *yj* are not evaluated as numbers. An observational pair (*xi*, *yj*) is called a *co-occurrence* of *xi* and *yj*. Dyadic data or COD contains these co-occurrences with note that a co-occurrence (*xi*, *yj*) can exist more than one time. So, each co-occurrence (*xi*, *yj*) is indexed by an index *r*. As a result, each co-occurrence is denoted by the triple (*xi*, *yj*, *r*) and we have (Hofmann & Puzicha, 1998, p. 1):

|  |  |
| --- | --- |
|  | (2.1) |

Where,

Of course, the size of is . As a convention, *xi*(*r*) and *yj*(*r*) indicate that -object and -object at the *r*th co-occurrence are *xi* and *yj*, respectively. Thus, the triplet (*xi*, *yj*, *r*) can be denoted as (*xi*(*r*), *yj*(*r*), *r*). For example, suppose = {*x*1, *x*2, *x*3) and = {*y*1, *y*2), and dyadic data of 4 co-occurrences, = {(*x*1, *y*1, 1), (*x*1, *y*1, 2), (*x*1, *y*2, 3), (*x*1, *y*1, 4)}, we observe that *x*1 and *y*1 occur together three times at *r*=1, *r*=2, and *r*=4 where as *x*1 and *y*2 occur together one time at *r*=3. In the first co-occurrence (*x*1, *y*1, 1), the notation *x*1(1) indicate that the -object at this co-occurrence is *x*1. In the third co-occurrence (*x*1, *y*2, 3), the notation *y*2(3) indicate that the -object at this co-occurrence is *y*2.

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

|  |  |
| --- | --- |
|  | (2.2) |

Where,

As a convention, *Z*(*r*) or *z*(*r*) indicates that the associative value at *r*th co-occurrence is *Z*=*z*. Thus, the quadruplet (*xi*, *yj*, *Z*, *r*) can be denoted as (*xi*(*r*), *yj*(*r*), *Z*(*r*), *r*). For example, suppose = {*x*1, *x*2, *x*3) and = {*y*1, *y*2), and dyadic sample of 4 co-occurrences, = {(*x*1, *y*1, 6, 1), (*x*1, *y*1, 8, 2), (*x*1, *y*2, 7, 3), (*x*1, *y*1, 9, 4)}, we observe that *x*1 and *y*1 occur together three times at *r*=1, *r*=2, and *r*=4 where as *x*1 and *y*2 occur together one time at *r*=3. Moreover, at *r*=1, *r*=2, *r*=3, and *r*=4, associative values are *Z*(1)=6, *Z*(2)=7, *Z*(3)=8, and *Z*(4)=9, respectively. Valued dyadic data is special case of dyadic data. As a convention, dyadic data is default if there is no additional information.

Given fixed *xk*, let be the -partitioned subset of which contains co-occurrences whose -objects are fixed at *xk* (Hofmann & Puzicha, Statistical Models for Co-occurrence Data, 1998, p. 1). Note, can be empty. The size of is .

|  |  |
| --- | --- |
|  | (2.3) |

Dyadic data is partitioned into subsets .

Given fixed *yl*, let be the -partitioned subset of which contains co-occurrences whose -objects are fixed at *yl*. Note, can be empty. The size of is .

|  |  |
| --- | --- |
|  | (2.4) |

Dyadic data is partitioned into subsets .

Given fixed *xk* and fixed *yl*, let be the subset of the which contains co-occurrences whose -objects and -objects are fixed at *xk* and *yl*. Note, can be empty. The size of is .

|  |  |
| --- | --- |
|  | (2.5) |

Let *n*(*xi*) and *n*(*yj*) denote the number of *xi* and the number of *yj*, respectively.

|  |  |
| --- | --- |
|  | (2.6) |

Let *n*(*xi*, *yj*) denote the number of co-occurrences (*xi*, *yj*).

|  |  |
| --- | --- |
|  | (2.7) |

Let *n*(*xi*|*yj*) and *n*(*yj*|*xi*) denote the frequency of *xi* given *yj* and the frequency of *yj* given *xi*, respectively.

|  |  |
| --- | --- |
|  | (2.8) |

For example, suppose = {*x*1, *x*2, *x*3) and = {*y*1, *y*2), and dyadic data of 4 co-occurrences, = {(*x*1, *y*1, 1), (*x*1, *y*1, 2), (*x*1, *y*2, 3), (*x*1, *y*1, 4)}, we have = {(*x*1, *y*1, 1), (*x*1, *y*1, 2), (*x*1, *y*2, 3), (*x*1, *y*1, 4)}, = = Ø, = {(*x*1, *y*1, 1), (*x*1, *y*1, 2), (*x*1, *y*1, 4)}, = {(*x*1, *y*2, 3)}, = = {(*x*1, *y*1, 1), (*x*1, *y*1, 2), (*x*1, *y*1, 4)}, = {(*x*1, *y*2, 3)}, = = = = Ø, *n*(*x*1) = 1, *n*(*x*2) = *n*(*x*3) = 0, *n*(*y*1) = 3, *n*(*y*2) = 1, *n*(*x*1, *y*1) = 3, *n*(*x*1, *y*2) = 1, *n*(*x*2, *y*1) = *n*(*x*2, *y*2) = *n*(*x*3, *y*1) = *n*(*x*3, *y*2) = 0, *n*(*x*1 | *y*1) = 1, *n*(*x*1 | *y*2) = 1, *n*(*x*2 | *y*1) = *n*(*x*2 | *y*2) = *n*(*x*3 | *y*1) = *n*(*x*3 | *y*2) = 0, *n*(*y*1 | *x*1) = 3/4, *n*(*y*2 | *x*1) = 1/4.

Suppose each co-occurrence (*xi*, *yj*) belongs to a latent variable *C* and *C* has *K* values *ck* (s). These values *ck* (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 *ck* = *k* where *k* = 1, 2,…, *K*. The random variable *C* has discrete distribution such that every value has an associated probability *αk*. Of course, there are *K* probabilities *αk* (s). There are three kinds of dyadic mixture model for dyadic data such as symmetric mixture model (SMM), asymmetric mixture model (AMM), and product-space mixture model (PMM). This section only explains these models when they were introduced by Hofmann and Puzicha (Hofmann & Puzicha, Statistical Models for Co-occurrence Data, 1998).

The mixture model of dyadic data is called symmetric mixture model (SMM) if *αk* (s) are independent from both *xi* and *yj*. SMM is defined as follows (Hofmann & Puzicha, Statistical Models for Co-occurrence Data, 1998, p. 2):

|  |  |
| --- | --- |
|  | (2.9) |

Where *αk* is the probability of aspect *k*. Note, *P*(.) denote probability.

The is the probability of *xi* given aspect *k*.

The is the probability of *yj* given aspect *k*.

This implies that *xi* and *yj* are mutually independent in SMM.

The joint probability of *xi*, *yj*, and *k* is:

The parameter of SMM is Θ = (*αk*, *pi*|*k*, *qj*|*k*)*T* in which there are *K*( + + 1) partial parameters *αk*, *pi*|*k*, and *qj*|*k*. Note,

By applying GEM, given dyadic sample , at the *t*th iteration of GEM, given current parameter Θ(*t*) = (*αk*(*t*), *pi*|*k*(*t*), *qj*|*k*(*t*))*T*, the conditional expectation *Q*(Θ|Θ(*t*)) is:

|  |  |
| --- | --- |
|  | (2.10) |

Where,

|  |  |
| --- | --- |
|  | (2.11) |

Note, *n*(*xi*, *yj*) is the number of co-occurrences (*xi*, *yj*) in , which is specified by equation 2.7. Please refer to equation 1.4 to comprehend equation 2.11. Because there are three constraints

We use Lagrange duality method to maximize *Q*(Θ|Θ(*t*)). The Lagrange function *la*(Θ, *λ* | Θ(*t*)) is sum of *Q*(Θ|Θ(*t*)) and these constraints, as follows:

Note, *λ* = (*λ*1, *λ*2, *λ*3)*T* where *λ*1≥0, *λ*2≥0, and *λ*3≥0 are called Lagrange multipliers. Of course, *la*(Θ, *λ* | Θ(*t*)) is function of Θ and *λ*. The next parameters Θ(*t*+1) that maximizes *Q*(Θ|Θ(*t*)) at M-step of some *t*th iteration is solution of the equation formed by setting the first-order partial derivatives of Lagrange function regarding Θ and *λ* to be zero.

The first-order partial derivative of Lagrange function regarding *αk* is:

Setting this partial derivative to be zero, we obtain:

Summing the equation above over *K* aspects {1, 2,…, *K*}, we have:

This means the next parameters *αk*(*t*+1) is:

|  |  |
| --- | --- |
|  | (2.12) |

The first-order partial derivative of Lagrange function regarding *pi*|*k* is:

Setting this partial derivative to be zero, we obtain:

Summing the equation above over , we have:

This means the next parameters *pi*|*k*(*t*+1) is:

|  |  |
| --- | --- |
|  | (2.13) |

Similarly, the next parameters *qj*|*k*(*t*+1) is:

|  |  |
| --- | --- |
|  | (2.14) |

The two steps of GEM algorithm for SMM at some *t*th iteration are shown in Table 2.1.

*Table 2.1: E-step and M-step of GEM algorithm for SMM.*

|  |
| --- |
| *E-step*:  The conditional probability *P*(*k* | *xi*, *yj*, Θ(*t*)) is calculated based on current parameter Θ(*t*) = (*αk*(*t*), *pi*|*k*(*t*), *qj*|*k*(*t*))*T*, according to equation 2.11.  *M-step*:  The next parameter Θ(*t*+1) = (*αk*(*t*+1), *pi*|*k*(*t*+1), *qj*|*k*(*t*+1))*T*, which is a maximizer of *Q*(Θ | Θ(*t*)) with subject to Θ, is calculated by equation 2.12, equation 2.13, and equation 2.14. |

GEM algorithm converges at some *t*th iteration. At that time, Θ*\** = Θ(*t*+1) = Θ(*t*) is the SMM itself. When SMM is applied into soft clustering, dyadic data is clustered according to blocks and each *αk* is coverage ratio of cluster *k* (aspect *k*).

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

|  |  |
| --- | --- |
|  | (2.15) |

The *αk*|*i* is the probability of aspect *k* given *xi*.

Where *pi* is the probability of *xi*.

The *qj*|*k* is the conditional probability of *yj* given aspect *k*. Suppose *yj* is independent from *xi* given *k*, we have:

Note, *qj*|*i* is the conditional probability of *yj* given *xi*, which is defined as follows:

The joint probability of *xi*, *yj*, and *k* is:

The parameter of AMM is Θ = (*αk*|*i*, *pi*, *qj*|*k*)*T* in which there are *K*( + ) + partial parameters *αk*|*i*, *pi*, and *qj*|*k*. Note,

By applying GEM, given dyadic sample , at the *t*th iteration of GEM, given current parameter Θ(*t*) = (*αk*(*t*), *pi*|*k*(*t*), *qj*|*k*(*t*))*T*, the conditional expectation *Q*(Θ|Θ(*t*)) is:

|  |  |
| --- | --- |
|  | (2.16) |

Where,

|  |  |
| --- | --- |
|  | (2.17) |

Please refer to equation 1.4 to comprehend equation 2.17. Because there are three constraints

We use Lagrange duality method to maximize *Q*(Θ|Θ(*t*)). The Lagrange function *la*(Θ, *λ* | Θ(*t*)) is sum of *Q*(Θ|Θ(*t*)) and these constraints, as follows:

Note, *λ* = (*λ*1, *λ*2, *λ*3)*T* where *λ*1≥0, *λ*2≥0, and *λ*3≥0 are called Lagrange multipliers. Of course, *la*(Θ, *λ* | Θ(*t*)) is function of Θ and *λ*. The next parameters Θ(*t*+1) that maximizes *Q*(Θ|Θ(*t*)) at M-step of some *t*th iteration is solution of the equation formed by setting the first-order partial derivatives of Lagrange function regarding Θ and *λ* to be zero.

The first-order partial derivative of Lagrange function regarding *αk*|*i* is:

Setting this partial derivative to be zero, we obtain:

Summing the equation above over *K* aspects {1, 2,…, *K*}, we have:

This means the next parameters *αk*|*i*(*t*+1) is:

|  |  |
| --- | --- |
|  | (2.18) |

The first-order partial derivative of Lagrange function regarding *pi* is:

Setting this partial derivative to be zero, we obtain:

Summing the equation above over , we have:

This means the next parameters *pi*(*t*+1) is:

|  |  |
| --- | --- |
|  | (2.19) |

The first-order partial derivative of Lagrange function regarding *qj*|*k* is:

Setting this partial derivative to be zero, we obtain:

Summing the equation above over , we have:

This means the next parameters *qj*|*k*(*t*+1) is:

|  |  |
| --- | --- |
|  | (2.20) |

The two steps of GEM algorithm for AMM at some *t*th iteration are shown in Table 2.2.

*Table 2.2: E-step and M-step of GEM algorithm for AMM.*

|  |
| --- |
| *E-step*:  The conditional probability *P*(*k* | *xi*, *yj*, Θ(*t*)) is calculated based on current parameter Θ(*t*) = (*αk*|*i*(*t*), *pi*(*t*), *qj*|*k*(*t*))*T*, according to equation 2.17.  *M-step*:  The next parameter Θ(*t*+1) = (*αk*|*i*(*t*+1), *pi*(*t*+1), *qj*|*k*(*t*+1))*T*, which is a maximizer of *Q*(Θ | Θ(*t*)) with subject to Θ, is calculated by equation 2.18, equation 2.19, and equation 2.20. |

GEM algorithm converges at some *t*th iteration. At that time, Θ*\** = Θ(*t*+1) = Θ(*t*) is the AMM itself. When AMM is applied into soft clustering, dyadic data is clustered vertically (horizontally) and each *αk*|*i* is coverage ratio of cluster *k* (aspect *k*) according to *xi*. Soft clustering with AMM is also called one-side clustering.

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

|  |  |
| --- | --- |
|  | (2.21) |

For every *k* belongs to {1, 2,…, *K*}, there always exists a respective pair: and . However, for each or each , there are many respective *k*.

|  |  |
| --- | --- |
|  | (2.22) |

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

|  |  |
| --- | --- |
|  | (2.23) |

As usual, *αk* is the probability of aspect *ck* but is the probability of *xi* given of *k* and is the probability of *yj* given of *k*.

The joint probability of *xi*, *yj*, and *k* is:

The parameter of PMM is Θ = (*αk*, , )*T* in which there are *K* + + partial parameters *αk*, , and . Note,

Learning PMM is like learning SMM and so it is not necessary to duplicate the expansion of *Q*(Θ|Θ(*t*)). The two steps of GEM algorithm for PMM at some *t*th iteration are shown in Table 2.3.

*Table 2.3: E-step and M-step of GEM algorithm for PMM.*

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *E-step*:  The conditional probabilities *P*(*k* | *xi*, *yj*, Θ(*t*)), *P*( | *xi*, *yj*, Θ(*t*)), and *P*( | *xi*, *yj*, Θ(*t*)) are calculated based on current parameter Θ(*t*) = , according to equation 2.24, equation 2.25, and equation 2.26.   |  |  | | --- | --- | |  | (2.24) | |  | (2.25) | |  | (2.26) |   Please refer to equation 1.4 to comprehend equation 2.24.  *M-step*:  The next parameter Θ(*t*+1) = , which is the maximizer of *Q*(Θ | Θ(*t*)) with subject to Θ, is calculated by equation 2.27, equation 2.28, and equation 2.29.   |  |  | | --- | --- | |  | (2.27) | |  | (2.28) | |  | (2.29) | |

GEM algorithm converges at some *t*th iteration. At that time, Θ*\** = Θ(*t*+1) = Θ(*t*) is the PMM itself. When PMM is applied into soft clustering, dyadic data is clustered in checked (stripped) and each *αk* is coverage ratio of cluster *k* (aspect *k*) but such cluster *k* corresponds to a pair of cluster and cluster . Soft clustering with PMM is also called two-side clustering.

**3. Predicting unaccomplished co-occurrent values**

This section is the main subject of this research in which some extensions of dyadic mixture models are used to predict unaccomplished values in valued dyadic data. When is valued dyadic data in which every co-occurrence (*xi*, *yj*) is associated with value *z* from random variable *Z* then, SMM is reformed as follows:

|  |  |
| --- | --- |
|  | (3.1) |

AMM is reformed as follows:

|  |  |
| --- | --- |
|  | (3.2) |

PMM is reformed as follows:

|  |  |
| --- | --- |
|  | (3.3) |

Where *fk*(*Z*|*φk*) is the *k*th PDF of *Z* corresponding to the aspect *k*, in which *φk* is parameter of *fk*(*Z*|*φk*). Of course, the parameter Θ now must include all *φk*. It is possible to consider that

Moreover, *Z* is only dependent on *k*.

Note, suppose *xi* and *yj* (as well as *yj* given *xi*) are independent from *Z* given aspect *k*, which is the hint to reform these models.

For example, within SMM, the joint PDF of *xi*, *yj*, *Z*, and *k* is:

Within AMM, the joint PDF of *xi*, *yj*, *Z*, and *k* is:

Within PMM, the joint PDF of *xi*, *yj*, *Z*, and *k* is:

Here it is only necessary to estimate *φk* because how to estimate other partial parameters was mentioned in section 2. By reforming the conditional expectation *Q*(Θ|Θ(*t*)), it is easy to find out that the next parameter *φk*(*t*+1) is solution of following equation:

|  |  |
| --- | --- |
|  | (3.4) |

Where *P*(*k* | *xi*(*r*), *yj*(*r*), Θ(*t*)) is specified by equation 2.11, equation 2.17, and equation 2.24 for SMM, AMM, and PMM, respectively. Especially, if *fk*(*Z*|*φk*) distributed normally, the next parameter *φk*(*t*+1) = (*μk*(*t*+1), Σ*k*(*t*+1))*T* containing mean *μk*(*t*+1) and covariance matrix Σ*k*(*t*+1) is calculated as follows:

|  |  |
| --- | --- |
|  | (3.5) |

Where *P*(*k* | *xi*(*r*), *yj*(*r*), Θ(*t*)) is specified by equation 2.11, equation 2.17, and equation 2.24 for SMM, AMM, and PMM, respectively. Please refer to (Nguyen, 2020, pp. 83-84) to comprehend equation 3.5.

In valued dyadic sample , many co-occurrences (*xi*, *yj*) are not existent and thus, it is required to predict or estimate *Z* value of inexistent co-occurrence (*xi*, *yj*). This *Z* value is called unaccomplished co-occurrent value or unaccomplished associative value. A so-called expected co-occurrent (EC) method is used to estimate *Z*. Firstly, it is necessary to define the conditional PDF of *Z* given *xi* and *yj*. According to Bayes’ rule, we have:

|  |  |
| --- | --- |
|  | (3.6) |

Then, *Z* value of inexistent co-occurrence (*xi*, *yj*) is estimated by an estimate which is the expectation of *Z* given the conditional PDF *f*(*Z* | *xi*, *yj*, Θ), as follows:

|  |  |
| --- | --- |
|  | (3.7) |

In short, EC method is specified by equation 3.6 and equation 3.7. Now we expend the two equations for SMM, AMM, and PMM. The conditional PDF *f*(*Z* | *xi*, *yj*, Θ) of SMM is:

|  |  |
| --- | --- |
|  | (3.8) |

Following is the proof of equation 3.8.

Similarly, the conditional PDF *f*(*Z* | *xi*, *yj*, Θ) of AMM is:

|  |  |
| --- | --- |
|  | (3.9) |

The conditional PDF *f*(*Z* | *xi*, *yj*, Θ) of PMM is:

|  |  |
| --- | --- |
|  | (3.10) |

Obviously, equation 3.8, equation 3.9, and equation 3.10 are extensions of equation 3.6.

The estimate for SMM is:

|  |  |
| --- | --- |
|  | (3.11) |

The estimate for AMM is:

|  |  |
| --- | --- |
|  | (3.12) |

The estimate for PMM is:

|  |  |
| --- | --- |
|  | (3.13) |

Where *Ek*(*Z*|*φk*) is expectation of *Z* given the *k*th PDF of *Z*:

|  |  |
| --- | --- |
|  | (3.14) |

If *fk*(*Z*|*φk*) is multinormal PDF with mean *μk* and covariance matrix Σ*k* then, we have *Ek*(*Z*|*φk*) = *μk*. Note, equation 3.11, equation 3.12, and equation 3.13 are extensions of equation 3.7.

Hofmann’s research (Hofmann, Latent Semantic Models for Collaborative Filtering, 2004) is different from EC method when Hofmann assumed that *fk*(*Z*|*φk*) is dependent on both *k* and *xi* so that *fk*(*Z*|*φk*) is replaced by .

Hofmann also assumed that (Hofmann & Puzieha, Latent Class Models for Collaborative Filtering, 1999, p. 690)

The sign “” indicates the proportion. Therefore, according to Hofmann, the conditional PDF *f*(*Z* | *xi*, *yj*, Θ) was defined as follows:

|  |  |
| --- | --- |
|  | (3.15) |

The estimate is still calculated by equation 3.7 except that *f*(*Z* | *xi*, *yj*, Θ) was defined by equation 3.15. As a result, equation 3.15 is the real mixture model of Hofmann in (Hofmann, Latent Semantic Models for Collaborative Filtering, 2004) and then Hofmann applied EM algorithm to learn parameters *αk*, *qj*|*k*, and *φik*. Therefore, Hofmann’s mixture model in (Hofmann, Latent Semantic Models for Collaborative Filtering, 2004) is not mixture models of co-occurrences (*xi*, *yj*) specified by equation 2.9 (SMM), equation 2.15 (AMM), and 2.23 (PMM). Hofmann’s mixture model is appropriate to collaborative filtering.

**4. Conclusions**

Essentially, learning dyadic data with models such as SMM, AMM, and PMM is unsupervised learning and it is easy to apply these models into soft clustering. Predicting or estimating unaccomplished values is essential to make a weighted sum of centroids over all clusters. Currently, an unaccomplished value is estimated based on pre-knowledge of an existent pair of two objects (-object and -object). As a result, an estimate is fixed if the two objects are fixed. In future, we try to find out another method to take advantages of more than two existent objects with a set of values. Combination of dyadic mixture model and regression model is a candidate method but how to prove and explain it is still fuzzy problem.

**Declaration of Interest Statement**

An original version of this paper was published as the third chapter in the book “Some Applications of Expectation Maximization Algorithm” by Eliva Press (https://www.elivapress.com/en/book/book-5132639477) available at https://www.researchgate.net/publication/359661325\_Some\_Applications\_of\_Expectation\_Maximization\_Algorithm

**References**

Hofmann, T. (2004, January). Latent Semantic Models for Collaborative Filtering. *ACM Transactions on Information Systems (TOIS), 22*(1), 89-115. doi:10.1145/963770.963774

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., & Puzieha, J. (1999). Latent Class Models for Collaborative Filtering. In T. Dean (Ed.), *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI '99)* (pp. 688-693). San Francisco, CA, USA: Morgan Kaufmann. Retrieved from https://dl.acm.org/citation.cfm?id=687583

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-learning-from-dyadic-data

Nguyen, L. (2020). *Tutorial on EM algorithm.* MDPI. Preprints. doi:10.20944/preprints201802.0131.v8