An Accurate Evaluation of MSD Log-likelihood and its Application in Human Action Recognition

Nuha Zamzami, and Nizar Bouguila

Concordia Institute for Information Systems Engineering (CIISE),
Concordia University, Montreal, QC., Canada

IEEE GlobalSIP 2019

Summary

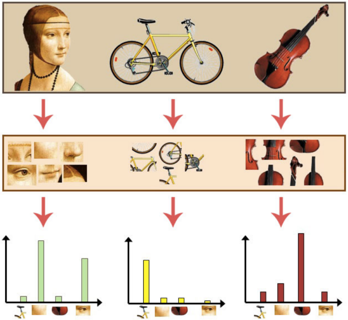
Objectives

1. To propose a new parameterization of the Multinomial Scaled Dirichlet (MSD) log-likelihood function based on a truncated series consisting of Bernoulli polynomials.
2. To adopt the mesh algorithm for computing this log-likelihood used for parameters estimation within the mixture model framework.
3. To propose a model selection approach which is seamlessly integrated with the parameters estimation and avoids several drawbacks of the standard Expectation-Maximization algorithm including the sensitivity to initialization and possible convergence to the boundary of the parameter space.
4. To validate the merits of the proposed approach via a challenging application that involves human action recognition.

Count Data in Computer Vision

Bag of words can be applied to image classification by treating features as visual words. Several challenges include :

- Features appears in **bursts** : if a visual word appears once it is much more likely to appear again.
- Many detected regions are assigned to a single visual word leads to **overdispersion** with respect to the multinomial distribution (*i.e.* the variance of a count variable exceeds its mean).
- Each image represented by a **high-dimensional and sparse** vector of occurrence count of thousands of features.



Modeling Count Data

- A generic multinomial (MN) distribution, which is typically used to model count data, can not handle the burstiness phenomenon given its independency assumption.
- The overdispersion can not be handled as more variation is observed than expected, according to the nominal covariance matrix of the multinomial distribution.
- Researchers found that by extending the multinomial distribution to the Dirichlet-multinomial (DMN) both burstiness and overdispersion phenomena modeling can be addressed.
- Recently, we have introduced an alternative model to DMN, called Multinomial Scaled Dirichlet (MSD) that allow more modeling flexibility in several real-world applications.

The Multinomial Scaled Dirichlet (MSD) Distribution

Define $\mathbf{X} = (x_1, \ldots, x_W)$ as a vector of counts. The Multinomial Scaled Dirichlet (MSD) is the composition of the Multinomial and scaled Dirichlet, with a set of parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_W)$, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_W)$, and it is obtained by integrating over the multinomial, which gives the marginal distribution of $\mathbf{X}$ (Zamzami & Bouguila, 2018) :

$$\mathcal{P}(\mathbf{X}|\alpha, \beta) = \int_{\rho} \mathcal{M}(\mathbf{X}|\boldsymbol{\rho})\mathcal{SD}(\boldsymbol{\rho}|\boldsymbol{\alpha}, \boldsymbol{\beta})d\rho$$
$$= \frac{N!}{\prod_{w=1}^{W} x_w!} \frac{\Gamma(A)}{\Gamma(A+N) \prod_{w=1}^{W} \beta_w^{x_w}} \prod_{w=1}^{W} \frac{\Gamma(\alpha_w + x_w)}{\Gamma(\alpha_w)} \tag{1}$$

where Γ denotes the Gamma function, $N = \sum_{w=1}^{W} x_w$, and $A = \sum_{w=1}^{W} \alpha_w$.

Approximating the Paired Log-Gamma Difference

- The MSD likelihood function is given by :

$$\mathcal{L}(p,\psi,\beta;\mathbf{X}) = -\Big(\ln\Gamma(1/\psi+N) - \ln\Gamma(1/\psi)\Big) - \sum_{w=1}^{W} x_w \ln(\beta_w) + \sum_{w=1}^{W}\left(\ln\Gamma\left(1/\frac{\psi}{p_w}+x_w\right) - \ln\Gamma\left(1/\frac{\psi}{p_w})\right)\right) \quad (2)$$

where $N=\sum_{w=1}^{W} x_w$, $\psi = 1/A$ is the overdispersion parameter, and $p_w = \psi\alpha_w$.

- We use the approximation of the paired log-gamma difference method

$$\ln\Gamma(1/x+y) - \ln\Gamma(1/x) \approx -y\ln x + D_m(x,y) \quad (3)$$

when y is an integer, $|x|, \min(|y-1|,|y|) < 1$, $xy \leq \delta$ and :

$$D_m(x,y) = \sum_{n=2}^{m} \frac{(-1)^n \phi_n(y)}{n(n-1)} x^{(n-1)}$$

where $\phi_n(y) = B_n(y) - B_n$ is the old type Bernoulli polynomial, $B_n(y)$ and B_n indicate the nth Bernoulli polynomial,
and nth Bernoulli number ($B_n = B_n(0)$), respectively.

The Mesh Algorithm for Evaluating the MSD Log-likelihood Function

- First, generate the mesh using :

$$x_w^{(l)} = \lfloor \alpha_w^{(l-1)} \delta \rfloor \tag{4}$$

- Then, we select the level of the mesh L, so it would be the smallest integer satisfying :

$$\sum_{l=1}^{L} x_w^{(l)} \geq x_w \quad \text{for all } w = 1, \ldots, W$$

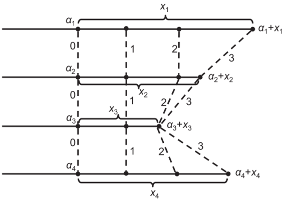


FIGURE – A graphical depiction of the mesh algorithm (Yu & Shaw, 2014).

- Afterwards, we adjust $x_w^{L'_w}$ such that $\sum_{l=1}^{L'_w} x_w^{(l)} = x_w$, and all the remaining $x_w^{(l)}(l > L'_w)$ will be set to zero.
  - With this adjusted mesh, we can use the approximation in (Eq. 2) to compute the MSD log-likelihood as the sum of each $X^{(l)}$ log-likelihood, as :

$$\mathcal{L}(p, \psi, \beta; X^{+}) = \sum_{l=1}^{L} \mathcal{L}(p^{(l-1)+}, \psi^{(l-1)}, \beta^{(l-1)+}; X^{(l)+}) \tag{5}$$

where X^{+} is the vector of non-zero elements in X.

Learning Algorithm

- The algorithm starts with a large number of components and iteratively deletes components as they become irrelevant.
- The model selection is based on the minimum message length (MML) criterion, defined as :

$$\Theta_{MML} = \arg\min_{\Theta} \Big\{ -\ln P(\Theta) - \mathcal{L}(\mathcal{X}, \mathcal{Z}|\Theta) + \frac{1}{2}\ln|\mathbf{I}(\Theta)| + \frac{\mathcal{D}(\Theta)}{2}\Big(1 + \ln\frac{1}{12}\Big)\Big\} \quad (6)$$

where $\mathcal{L}(\mathcal{X}, \mathcal{Z}|\Theta)$ the complete-data log-likelihood, $P(\Theta)$ is the prior distribution, $\mathbf{I}(\Theta)$ is the expected Fisher information matrix, and $\mathcal{D}(\Theta)$ denotes the dimensional of the model.

- We use the component-wise EM procedure (CEM), and any weak component will be annihilated.
- MML criterion is re-evaluated for non-zero components only until the message length difference becomes insignificant.

Results-Frame Level

- Frames extracted from the UCF sports dataset.
- Each extracted frame is treated as an image from which a set of interest points are detected and described using Scale-Invariant Feature Transform (SIFT), then represented as count vectors using Bag of Features (BoF) approach.

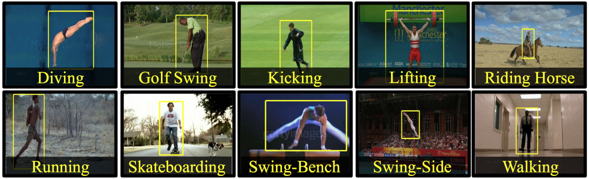


FIGURE – Sample frames from UCF sports dataset.

The average recognition accuracy for UCF sports dataset.

Model	Accuracy
Mixture of Multinomials	61.72±0.02
Mixture of DMN	66.97±0.34
Mixture of MSD (EM)	66.97±0.03
MSD_mesh	**78.60±0.03**

Results-Video Sequences 1

- We used two different subsets from the action recognition dataset of realistic action videos, collected from YouTube, called UCF101.
- Each video is represented as a vector of count data using the extension of Bag of words paradigm to videos.
- Detection of the local neighborhood with a significant variations is via Spatio Temporal Interest Points (STIP), then 3D SIFT descriptor is used.

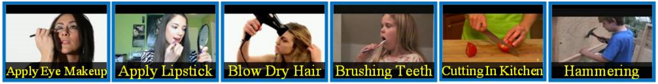


FIGURE – Sample Human-Object Interaction frames from UCF101.

The average recognition accuracy for Human-Object Interaction subset.

Model	Accuracy
Mixture of Multinomials	81.77 ± 0.02
Mixture of DMN	82.29 ± 0.04
Mixture of MSD (EM)	84.90 ± 0.07
MSD_mesh	**87.76 ± 0.02**

Results-Video Sequences 2

- UCF101 gives the largest diversity in terms of actions and with the presence of large variations in camera motion, object appearance and pose, object scale, viewpoint, cluttered background, illumination conditions, etc, making it the most challenging dataset to-date.

FIGURE – Sample Playing Musical Instruments frames from UCF101.

The average recognition accuracy for Playing Musical Instruments subset.

Model	Accuracy
Mixture of Multinomial	81.89±0.07
Mixture of DMN	82.65±0.02
Mixture of MSD (EM)	89.54±0.04
Mixture of MSD (Mesh)	**92.60±0.01**

Conclusion

- The mesh algorithm for the computation of the Multinomial scaled Dirichlet (MSD) log-likelihood function has been proposed to be used within a statistical clustering framework.
- The mesh method is generally more stable and provides an accurate computation of the log-likelihood function leads to a significant improvement in the clustering accuracy.
- The proposed algorithm successfully selected the optimal number of components, that agrees with the prespecified ones for different datasets.
- We validated the proposed framework through human action recognition, and we believe that it could be extended to other applications involve overdispersed count data.

References

N. Zamzami and N. Bouguila (2018)

Text modeling using multinomial scaled dirichlet distributions

Recent Trends and Future Technology in Applied Intelligence. IEA/AIE 2018. Lecture Notes in Computer Science, vol 10868. Springer 2018, 69 – 80.

P. Yu and C. Shaw (2014)

An efficient algorithm for accurate computation of the Dirichlet-multinomial log-likelihood function

Bioinformatics. vol. 30, no. 11, 2014, 1547 – 1554.

M. Figueiredo and A. Jain (2002)

Unsupervised learning of finite mixture models

IEEE Transactions on Pattern Analysis & Machine Intelligence no. 3, 2002, 381 – 396.

* Check to the paper for the complete list of references.

Thank You !

See you at the poster sessions.

For questions ; please contact the authors :

Nuha Zamzami n_zamz@encs.concordia.ca,
Nizar Bouguila nizar.bouguila@concrodia.ca