An Incremental von Mises Mixture Framework for Modelling Human Activity Streaming Data

Eris Chinellato\textsuperscript{1}, Kanti V. Mardia\textsuperscript{2}, David C. Hogg\textsuperscript{2}, and Anthony G. Cohn\textsuperscript{2}

\textsuperscript{1} Middlesex University
The Burroughs, London, NW4 4BT, UK
e.chinellato@mdx.ac.uk
\textsuperscript{2} University of Leeds
Leeds, LS2 9JT, UK

\textbf{Abstract.} Modelling the time of occurrence of events from data streams is a challenging task, since the underlying distributions can be both cyclic and multimodal. Moreover, in order to avoid the indefinite growth of data storage, historical streaming data has to be represented only with model parameters, discarding the single values. In this work, we introduce an incremental framework for a mixture of circular von Mises distributions to model the time of occurrence of events. Applying our framework to the time of occurrence of human activities, we show that it is able to represent the relevant information of a cyclic data stream by storing only the distribution parameters, highlighting that its use can extend to a number of applications.

\section{Introduction}

When analysing information from data streams, modelling efforts have to satisfy conflicting goals, such as representing the data as faithfully as possible while optimising data storage, as saving all data for an indefinitely long period of time is not sustainable. Among the most common data in streams is the time of occurrence of events, the analysis of which is nevertheless particularly challenging, since underlying distributions are usually both cyclic and multimodal.

The distribution of choice for modelling cyclical data is the von Mises distribution, which wraps around over a specified period [13]. Mixtures of von Mises (VMM) have been used before to model cyclical multimodal data, but all existing approaches require the availability of all data to generate the mixture model. By applying core circular statistics concepts to the standard VMM, we have developed IVMM, an \textit{Incremental von Mises Mixture} model. IVMM fits a cyclical, multimodal distribution over the data, and for the first time this is performed incrementally, so that only summary statistics of the model need to be saved, and not all the data as in standard applications.

Incremental model building, here introduced for the first time for circular processes, is in some cases very important, as efficiency can become an issue when considering long term data from multiple sources. Consider for example data from millions of users of a smartphone application coming in continuously,
and having to be merged and processed together, or people passing in front of multiple sensors in public places. In this work we offer a solution to this issue when it refers to cyclic, multimodal events, and apply it to the time of occurrence of human activities.

We next describe the type of data we are dealing with, and existing related works and techniques. We then introduce our new IVMM framework, and show and discuss its experimental evaluation.

2 Background

In this section we describe in detail the type of data we are dealing with, the task, and review current approaches to the problem.

2.1 The problem and data description

Modelling data such as the daily times of occurrence of a human activity is not straightforward, because the data are typically both cyclic and multimodal. This can be appreciated looking at the example activity in Fig. 1, obtained from the Consolidated Human Activity Database (CHAD). CHAD is a large registry of domestic human activities collected in smart-homes and organised into several datasets. A typical CHAD dataset includes dozens of activities, each with hundreds or thousands of occurrences collected daily, in some cases for more than two years. Each activity instance specifies the location and time of occurrence of the activity, among other information. The time of occurrence is normally rounded to the nearest minute, but some sensors have a one second precision. The plots in Fig. 1 represent the same histogram of the time of occurrence of 515 instances of activity Converse (dataset code: CAA, activity code: 17241), plotted in polar (a) and Cartesian (b) spaces.

On the one hand, the activity frequency distribution covers the 24 hours completely, without any time gaps, and there is no easy and correct way to split the data at a certain time (e.g. midnight, as conventionally done, or any other time). This means that any analysis needs to consider the data as wrapped around in a continuous fashion, otherwise there would be a discontinuity in the model at the splitting time. On the other hand, most activities are expected to have a multimodal nature: they cannot be fitted by a single function with only one maximum, but require instead a mixture of functions, like the data in Fig. 1.

Considering that events can have different periodicity $T$ (e.g. 360 for radial directions, 86400 for time of the day in seconds, 366 for days in a year), we express our problem as the following. Let us consider a set $X$ of $N$ points $x_i, i = 1, ..., N$, coming from a data stream of variable $x$, taking values within the interval $[0, T]$. We want to create and continuously update a model of $x$ using a mixture of distributions, without having to store all data points in memory.

---

Fig. 1. Example of activity time data requiring a cyclic, multimodal fitting. To preserve the proportion between frequencies and plot bar surface, bar height in the polar plot is proportional to the square root of the frequency.

2.2 Existing solutions

Circular statistics techniques are typically applied to geographical, meteorological, zoological and biochemical data (see e.g. [6, 10, 12, 14]). Their use in human studies, particularly for what concerns time periodicity, is virtually unexplored. Typical methods are non-parametric and require the storage of all datapoints [2, 9], whilst here we are more concerned with parametric and incremental methods.

For what concerns the analysis of time of occurrence of human activities, previous works have often employed non-homogeneous Poisson processes, for classifying and predicting daily patterns and day types [8, 15], or for spatio-temporal identification of people for robotics applications [17]. Works that take into account the cyclic nature of data to avoid discontinuities in the models have recently become more common. They mostly use wrapped Gaussian distributions [4, 3], in one case mixed with von Mises [11]. So far, all of them can only be applied in batch to the whole dataset, and not incrementally, so their models cannot be updated dynamically with the data stream, and they require the storage of every single data point.

Different approaches have been proposed for implementing incremental versions of Gaussian mixture models (see e.g. [16, 5]), but they have not been so far extended to the case of von Mises, or any other circular distribution. An incremental update of a von Mises mixture model is introduced in [1], but only for inserting new points into existing models in the mixture, while in this work we test new data batches against all existing models, and are thus able to add new models to the mixture when necessary. The conceptual complexity of allowing a certain flexibility to the mixture while keeping it parametric is testified by the fact that current split-and-merge methods (see e.g. [18]) are not incremental, and require saving all data in order to add models to the mixture.
2.3 Mixture of von Mises fitting

Cyclic processes are better represented, and fitted, on a circular space, following techniques developed in the field of directional statistics [13]. The distribution of choice for fitting circular data is the von Mises, analogue to a normal distribution on the circle. It is typically employed to fit radial directions or periodical events. Multimodal data that cannot be modelled by a single distribution require the fitting of mixtures of von Mises [12, 7]. The blue curves in Fig. 1 show the outcome of fitting a mixture of von Mises distributions over the data, in polar and Cartesian coordinates, respectively. Fig. 1(b) also shows the contribution of each von Mises component.

3 Incremental von Mises Mixture [IVMM]

3.1 The von Mises distribution

A von Mises distribution is described by a preferred direction or mean, $\mu$, and a concentration parameter, $\kappa$:

$$ f(x | \mu, \kappa) = \frac{e^{\kappa \cos(x - \mu)}}{2 \pi I_0(\kappa)}, \quad (1) $$

where $I_0(\kappa)$ is the modified Bessel function of order 0 (the canonical solution of a type of differential equation on purely imaginary numbers). $\mu$ is the average of the vectors of all datapoints, and can be estimated by using either angular, or complex number representations. $\kappa$ is a measure of concentration – reciprocal of dispersion, and thus akin to $1/\sigma^2$ – which cannot be computed analytically, but only estimated by approximation.

Given a dataset of $n$ data points $x_i$, and a selected period $T$, all data points can be represented by angles $\theta_i = 2\pi x_i / T$ on the unit circle, $i = 1, \ldots, n$.

The mean Cartesian coordinates $\bar{C}$ and $\bar{S}$ of all the points are the fundamental descriptors for building the circular statistics.

$$ \bar{C} = \frac{1}{n} \sum_{j=1}^{n} \cos \theta_j, \quad \bar{S} = \frac{1}{n} \sum_{j=1}^{n} \sin \theta_j \quad (2) $$

The mean resultant length $\bar{R}$ and mean direction $\bar{\theta}$ of the dataset can be computed directly from $\bar{C}$ and $\bar{S}$.

$$ \bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2} \quad \bar{\theta} = \arctan \frac{\bar{S}}{\bar{C}} \quad (3) $$

Estimated $\hat{\mu}$ and $\hat{\kappa}$ of a von Mises can be computed through maximum Likelihood Estimation. Considering a samples of $n$ values $\theta_1, \ldots, \theta_n$ originating from a von Mises distribution $vM(\mu, \kappa)$, the Log-likelihood to maximise is:

$$ l(\mu, \kappa; \theta_1, \ldots, \theta_n) = n(- \log 2\pi + \kappa \bar{R} \cos(\bar{\theta} - \mu) - \log I_0(\kappa)) \quad (4) $$
The resultant maximum likelihood estimators are:
$$\hat{\mu} = \bar{\theta}$$  \(5\)
$$A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)} = \bar{R} \quad \rightarrow \quad \hat{\kappa} = A^{-1}(\bar{R})$$  \(6\)
where \(I_1(\kappa)\) is the modified Bessel function of order 1. The preferred direction \(\hat{\mu}\) can thus be computed directly from the sets of \(\theta_i\), while the concentration parameter \(\hat{\kappa}\) can be estimated numerically from \(\bar{R}\) by using one of various existing approximations of the inverse of function \(A(\kappa)\) [13, 7].

3.2 Incremental approach

In order to build an incremental mixture of von Mises it is necessary to test if two samples are likely to belong to the same generating distribution or not. A two samples \(F\) test on the hypothesis of equality of means is performed for this purpose:
$$H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 \neq \mu_2$$  \(7\)

The approximated value of the \(F\) distribution to check is [13]:
$$\frac{(n-2)(R_1 + R_2 - R)}{n - R_1 - R_2} \approx F_{1,n-2}$$  \(8\)
where \(R_1, R_2\) and \(R\) are the resultant lengths of the two samples of size \(n_1\) and \(n_2\), and of the combined sample of size \(n = n_1 + n_2\), respectively.

The above formulation allows us to develop a schema for incrementally updating a mixture of von Mises distributions. The only requirement is that each sample has enough datapoints to be suitably modelled by a von Mises function. Our feature vector for representing a single von Mises function is composed of only three elements: \((n, SC = nC, SS = nS)\). Let us consider an initial von Mises mixture \(vMM\), composed of \(N\) von Mises distributions:
$$vMM = \{vM_i, i = 1, ..., N\}, \quad vM_i = (n_i, SC_i, SS_i)$$  \(9\)

A new data sample received at time \(t\) can again be modelled by a mixture of von Mises:
$$vMM' = \{vM_j, j = 1, ..., N'\}, \quad vM_j = (n_j, SC_j, SS_j)$$  \(10\)

In order to merge the two mixture models, each new \(vM_j\) is tested against each existing \(vM_i\). If we can accept the null hypothesis \(H_0(\mu_i = \mu_j)\), the current \(vM_i\) is updated:
$$vM_i' = (n_i + n_j, SC_i + SC_j, SS_i + SS_j)$$  \(11\)

If, instead, for a certain \(j\) we reject the null hypothesis for all existing models, \(H_1(\mu_i \neq \mu_j), \forall i\), it means that \(vM_j\) cannot be merged with any existing \(vM_i\), and must be added to \(vMM\).

As a last update step, after the model \(vMM\) has been modified according to all \(vM_j\), each pair of distributions \((vM_i, vM_{i+1})\) of \(vMM\) (including \((vM_N, vM_1)\)) are tested for equality against each others, and merged if \(H_0\).
3.3 Using IVMM

The new schema introduces the possibility of testing new samples, or whole new datasets, against the current model. An example of how datasets are represented by gradually increasing, but eventually stabilising, number of von Mises distributions is shown in Fig. 2. Clear continuous lines show the evolution of the number of components for eight different example activities, whilst the thicker black line is the average trend.

![Fig. 2. Evolution of the number of components with number of data samples for eight example von Mises mixtures. Black line is the average trend.](image)

Fig. 3 shows an example of the different outputs obtained by applying the IVMM framework (left) and the standard VMM approach (right). Fitting curves are shown in black, and the contribution of each component in grey. This particular example shows that in some cases the number of components given by IVMM or VMM can be different (here it is 5 and 3, respectively), but there is no implication that either model is better than the other.

4 Experimental validation

The main purpose of the IVMM framework is to produce faithful models of time of occurrence of events when saving all data is not possible or convenient, and standard mixture models cannot thus be applied. The key objective in our experimental validation is then to ensure that information loss in IVMM compared to non incremental methods, particularly standard VMM, is small.

In all experiments we have employed the same values for the fundamental parameters required by the IVMM. The period is in our case always set to 86400, the number of seconds in a day. The maximum number of components allowed for each single VMM is 7 (no limits are set for the total number of components in an incremental model). We have been using the Bayesian Information Criterion
Fig. 3. Comparison between IVMM (left hand side) and VMM (right hand side) fitting (CHAD activity 12100, care of baby).

(BIC) for choosing the best number of von Mises components of each mixture, both for IVMM and VMM. The significance level $\alpha$ to compare with F value in equation 8 is set to 0.1. Different datasets and activities achieve better performance for different values of $\alpha$, but a sensitivity analysis showed that there are no major effects for values between 0.05 and 0.15.

4.1 Testing within a dataset

Even when considering only time of day analysis, the IVMM technique offers a great range of practical applications. For example, it can be used to detect changes in habits, anomalous behaviours or people identities, by assessing the novelty of new data with respect to the accumulated experience.

In order to compare the discrimination capabilities of the models generated by IVMM and VMM, we employed both techniques for identifying atypical data, using only time information, by assessing their response to new samples coming from the same activity or from different ones. We have performed the test on 16 datasets from CHAD, on an average of 13.3 activities per dataset (we omitted all activities with less than 500 points in total, or with less than 3 samples). For each activity in a dataset, we built a full IVMM model, and then tested all samples from all other activities against it, using Equations 7 and 8. We have also tested each activity on its own models, by using a 3-fold validation to avoid testing on data used for building the model.

Fig. 4 shows the percentage of (correctly) passed tests for the same activity (full circles) and that of (wrongly) passed tests for different activities (empty circles). Average passed tests for all activities against themselves was $86.7 \pm 8.0\%$, while average passed tests for other activities was $54.3 \pm 14.9\%$ (experiments (a) and (b), respectively, in Tab. 1). This last figure may appear high as a false
positive, but it is expected, since most activities are consistently overlapping. For example, activities such as personal hygiene (code 14120) or general household (11000) are typically spread throughout the day and show in many cases very little regularity or clear patterns. These can thus be easily confused with many other activities on the base on pure time information. We also included placeholder activities such as uncertain (code U) or missing (code X). Moreover, all samples were drawn regularly (one per month) for each activity, with the consequence that, for some activities/datasets there were many points available and precise models could thus be built, whilst for others there were just a few points, resulting in very generic and sparse models.

Finally, tests performed by changing the order of presentation of samples have shown that, while IVMM is sensitive to sample order for small numbers of samples, models obtained from the same set of samples in different sequences are substantially indistinguishable from each others if the number of samples is large enough, as in the CHAD case.

![Fig. 4. Cross-activity testing on different datasets from CHAD, identified by their three letter codes. Full dots: passed tests on same activity (true positives); empty dots: average passed tests on other activities (false positives).](image)

### 4.2 Comparison with standard VMM

To verify if and how much the incremental approach reduced the predictive capabilities of a standard, batch Mixture of von Mises, we performed the same tests as in experiments (a) and (b) in Tab. 1 on standard VMM, computed
Table 1. IVMM experiments, summary of results. Rec Same Act = Recall on target activity, FP Other Act = False Positives on different activities.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Mean (%)</th>
<th>Std. Dev. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Rec Same Act, IVMM</td>
<td>86.7</td>
<td>8.0</td>
</tr>
<tr>
<td>(b) FP Other Act, IVMM</td>
<td>54.3</td>
<td>14.9</td>
</tr>
<tr>
<td>(c) Rec Same Act, VMM</td>
<td>89.8</td>
<td>5.0</td>
</tr>
<tr>
<td>(d) FP Other Act, VMM</td>
<td>52.6</td>
<td>20.2</td>
</tr>
<tr>
<td>(e) Generalisation, IVMM</td>
<td>80.5</td>
<td>9.2</td>
</tr>
<tr>
<td>(e) Generalisation, VMM</td>
<td>82.9</td>
<td>11.9</td>
</tr>
</tbody>
</table>

over all the available data. The results of tests (c) and (d) in Tab. 1 show that recall increased by about 3%, while false positives decreased by almost 2%. The expected performance loss of IVMM with respect to VMM is thus small. Quite importantly, the standard deviation of results is similar in the two cases, indicating that the incremental framework never carries to catastrophic failures in modelling the data. The opportunity of using VMM or IVMM would thus depend on the application, but it can be argued that the observed information loss is compensated by the increased long term usability of the IVMM approach.

4.3 Generalisation test

Sensitivity for a modelled activity is particularly useful if it generalises over datasets. As a validation test, we have thus performed an additional experiment, in which a mixture model created for an activity based on a certain dataset is tested with samples of the same activity obtained from other datasets, thus implementing a leave-(n-1)-out validation. The average percentage of successful tests over 37 activities (e) was 80.5 ± 9.2, compared to the same set value (a) 86.7 ± 8.0, showing that IVMM models were indeed able to generalise across datasets. Generalisation capabilities of standard VMM were only slightly better at 82.9, and had a higher variance in performance across activities. The detailed results depicted in Fig. 5 show that only for few particularly difficult activities the generalisation performance was below 0.7. Overall, our results show that faithful modelling of time of occurrence can be used to provide a useful prior to a classifier which combines this and other, non-temporal, features.

5 Conclusions

We have introduced in this work a new incremental von Mises mixture model for analysing and modelling time of occurrence of human activities. We have shown that our IVMM framework allows for both high quality representation of the data and efficient compression, being able to generalise over different datasets, with a small amount of information loss compared to standard VMM.

The gain in processing times and storage requirements of IVMM over VMM can be substantial for big datasets. In fact, space and time complexity of IVMM
Fig. 5. Generalisation of activity identification across CHAD datasets. Full dots: passed tests on same dataset; crosses: average passed tests on other datasets.

depend only on the number of models in the mixture. Space requirements are negligible, and processing time is always instantaneous for the small number of models expected to compose a mixture (never more than a dozen in all our experiments). For VMM, space requirements increase linearly with data, and this becomes an issue after a sufficiently long time. Computing time also grows linearly with the amount of data, eventually making online updating of models infeasible. Millions of data items are generated daily by increasingly high numbers of sensors, and on our PC applying a standard von Mises Mixture to one million data points (maximum k = 12) took 16 minutes.

The proposed methodology can be applied to a number of different fields, and the fast growth of big data technologies and global connectivity suggests that tools such as IVMM, able to represent data in compact and efficient ways will be more and more required in the future. We are currently working on an extension of the IVMM approach in order to make it more dynamic and flexible to changing trends in the data. The plan is to make it time-aware, giving higher importance in the model to more recent data. The incremental nature of the approach makes such a goal relatively easy, for example by modifying at each new sampling the importance of each model component according to the last time it has been updated.

References


