Using mutual proximity for novelty detection in audio music similarity

Arthur Flexer[*] and Dominik Schnitzer

Austrian Research Institute for Artificial Intelligence
Freyung 6/6, Vienna, Austria

Abstract. Mutual proximity rescales distance spaces to avoid negative effects of the curse of dimensionality. It results in probabilistic estimates of the proximity of data objects. We use these probabilities directly for novelty detection, i.e. the automatic identification of unknown data not covered by training data (e.g. a new genre in genre classification). Comparing this new approach with a distance based detection method we demonstrate improved performance on a standard music data set.

1 Introduction

Novelty detection is the identification of new or unknown data that a machine learning system is not aware of during training (see [3] for a review). It is a fundamental requirement for every good machine learning system to automatically identify data from regions not covered by the training data since in this case no reasonable decision can be made. In the case of music information retrieval (MIR), an application scenario is the rejection of songs from a previously unseen genre in genre classification. Another example is the automatic rejection of songs from play-lists because they do not fit the overall flavor of the majority of the list. Only little research on novelty detection in MIR so far exists [1, 5]

We compare two reject rules for novelty detection based on spectral similarity of songs and evaluate them within a genre classification context. One method [6] is based on density information about the training data captured in the distance matrix. Our proposed method rescales a distance space to probabilities of mutual proximity (MP) [4] which allows for convenient thresholding to detect novel data due its probabilistic interpretation.

2 Data and Methods

Data: For our experiments we use the "GTZan" collection consisting of $N = 1000$ audio tracks (each $30\,$s length) evenly spread over $G = 10$ music genres [7]. Every track is divided into overlapping frames for which 20 MFCCs are being computed which are modeled via a single Gaussian with full covariance matrix.

To compute a distance value between two Gaussians the symmetrized Kullback-Leibler divergence is used [2].

Ratio-reject: The first reject rule is comparing the distance between a test object x and its nearest neighbor in the training set $NN^{tr}(x)$, and the distance between this $NN^{tr}(x)$ and its nearest neighbor in the training set $NN^{tr}(NN^{tr}(x))$ [6]. The object is regarded as novel if the first distance is much larger than the second distance. Using the following ratio

$$\rho(x) = \frac{\|d(x, NN^{tr}(x))\|}{\|d(NN^{tr}(x), NN^{tr}(NN^{tr}(x)))\|} \tag{1}$$

we reject x if:

$$\rho(x) > E[\rho(x^{tr})] + s * std(\rho(x^{tr})) \tag{2}$$

with $E[\rho(x^{tr})]$ being the mean of all quotients $\rho(x^{tr})$ inside the training set and $std(\rho(x^{tr}))$ the corresponding standard deviation (assuming a normal distribution for $\rho(x^{tr})$). Parameter s can be used to change the threshold for rejection.

MP-reject: Mutual Proximity (MP) [4] rescales the original distance space so that two objects sharing similar nearest neighbors are more closely tied to each other, while two objects with dissimilar neighborhoods are repelled from each other. MP has been devised to counter the negative effects of hubness in high dimensional data spaces (i.e. data objects which are abnormally close to very many other data objects). For MP-reject we exploit the fact that MP rescales distances to probabilities which enables comparability and simple thresholding. MP reinterprets the distance of two objects as a mutual proximity in terms of their distribution of distances. To compute MP, we assume that the distances $D_{x,i=1..N}$ from an object x to all other objects in our data set follow a certain probability distribution, thus any distance $D_{x,y}$ can be reinterpreted as the probability of y being the nearest neighbor of x, given their distance $D_{x,y}$ and the probability distribution $P(X)$:

$$P(X > D_{x,y}) = 1 - P(X \leq D_{x,y}) = 1 - \mathcal{F}_x(D_{x,y}). \tag{3}$$

MP is then defined as the probability that y is the nearest neighbor of x given $P(X)$ and x is the nearest neighbor of y given $P(Y)$:

$$MP(D_{x,y}) = P(X > D_{x,y} \cap Y > D_{y,x}). \tag{4}$$

To compute MP in our experiments we assume that the distances $D_{x,i=1..N}$ follow a Gaussian distribution. We reject x if:

$$1 - MP(D_{x,y}) > p \tag{5}$$

with y being the closest neighbor of x in the training set and p a probability threshold for rejection.

3   Results

To evaluate the two novelty detection approaches we first set aside all songs belonging to a genre `g` as novel songs which yields data sets `novel` and `data` (all songs not belonging to genre g). Then we do a ten-fold crossvalidation (`c=1 ...10`) using `data` and `novel`: we randomly split `data` into `train` and `test` fold with `train` always consisting of 90% and `test` of 10% of `data`. We compute the percentage of `novel` songs which are rejected as being novel (`novel_reject`) and do the same for the `test` songs (`test_reject`). Last we compute the `accuracy` of the nearest neighbor classification on `test` data that has not been rejected as being novel. The evaluation procedure gives $G \times C$ (10×10) matrices of `novel_reject`, `test_reject` and `accuracy` for each parameterization of the novelty detection approaches.
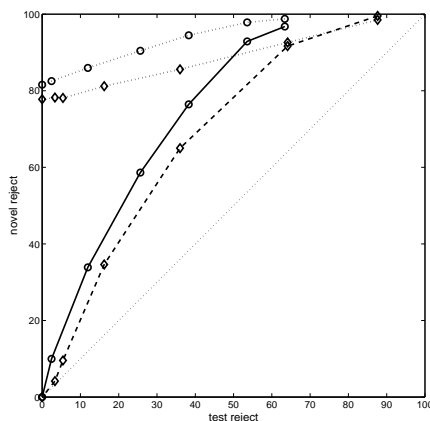
Fig. 1. Mean `test_reject` vs. `novel_reject` (circles, solid line for MP-reject; diamonds, dashed line for Ratio-reject) and `accuracy` (dotted lines, circles for MP-reject, diamonds for Ratio-reject) for s=inf (no rejection), 3, 2, 1, .5, 0, -.5 and p=1 (no rejection), .1, .05, .03, .02, .01, .001.

The results for novelty detection are given in Fig. 1 as Receiver Operating Characteristic (ROC) curves. To obtain an ROC curve the fraction of false positives (object is not novel but it is rejected, in our case `test_reject`) is plotted versus the fraction of true positives (object is novel and correctly rejected, in our case `novel_reject`). An ROC curve shows the trade-off between how sensitive and how specific a method is. Any increase in sensitivity will be accompanied by a decrease in specificity. If a method becomes more sensitive towards novel objects it will reject more of them but at the same it will also become less specific and also falsely reject more non-novel objects. Consequently, the closer a curve follows the left-hand border and then the top border of the ROC space, the better the method is. We plot the mean `test_reject` versus the mean `novel_reject` for falling numbers of s (Ratio-reject) and p (Ratio-reject). In addition the mean `accuracy` for each of the different values of s and p are depicted as separate curves. All means are computed across all 10×10 corresponding values.

The ROC curve for MP-reject is always above the curve for Ratio-reject indicating better performance both in terms of sensitivity and specificity. In addition the accuracy after novelty rejection is also always better with MP-reject. For instance, with the rejection threshold set to $p = .01$, MP-reject is able to correctly reject 93% `novel` data while falsely rejecting 54% of `test` data and achieving an `accuracy` of 98%. To achieve a comparable rejection rate of 92% of `novel` data one has to set the threshold for Ratio-reject to $s = 0$. At this rejection level one already falsely rejects 64% of `test` data and achieves an `accuracy` of only 93%.

4   Conclusion

We presented a new method for the important but often ignored task of novelty detection in audio music similarity. Our proposed method MP-reject was shown to improve the ability to reject novel data when compared to a classic distance based method. Both methods use distance information only and do not rely on any form of (genre) label information. Genre classification accuracy was used only to evaluate the level of success in rejecting novel data. Of course this is just a first attempt and future work should compare more data sets, features and competing methods. Of additional interest is the fact that the method at the heart of MP-reject, mutual proximity, has been developed to avoid a certain negative aspect of the curse of dimensionality named hubness [4]. Due to a principal problem of measuring distances in high dimensional spaces, hub objects emerge which are close to many other data points while orphan objects are far away from most other objects. It seems clear that hub objects should be hard to reject as being novel while orphans will probably always be rejected. Future work will investigate whether the improvement shown by MP-reject is on account of the changed role of hubs and orphans due to the application of mutual proximity.

References

1. Flexer A., Pampalk E., Widmer G.: Novelty detection based on spectral similarity of songs, in Proc. of the 6th Int. Conf. on Music Information Retrieval, 2005.
2. Mandel M., Ellis D.: Song-level features and support vector machines for music classification, in Proc. of the 6th Int. Conf. on Music Information Retrieval, 2005.
3. Markou M., Singh S.: Novelty detection: a review-part 1: statistical approaches, Signal Processing, 83(12):2481-2497, 2003.
4. Schnitzer D., Flexer A., Schedl M., Widmer G.: Local and Global Scaling Reduce Hubs in Space, Journal of Machine Learning Research, 13(Oct):2871-2902, 2012.
5. Sturm B.L.: Music genre recognition with risk and rejection, in Proceedings of International Conference on Multimedia and Expo, 2013.
6. Tax D.M.J., Duin R.P.W.: Outlier detection using classifier instability, in Proc. Joint IAPR Int. Workshop SSPR'98 and SPR'98, Springer, pp. 593-601, 1998.
7. Tzanetakis G., Cook P.: Musical genre classification of audio signals, IEEE Transactions on Speech and Audio Processing, Vol. 10, Issue 5, 293-302, 2002.