THE CONTEXT: GENERATIVE EMBEDDINGS

- Classification of structured objects (e.g., shapes) is typically addressed with generative models (able to deal with non vectorial representations).
- Discriminative classifiers (e.g., SVM) typically outperform generative models, but prefer a vectorial representation.
- Generative embeddings represent hybrid generative-discriminative approaches, which exploit a learned generative model to map a possibly non vectorial object into a vector space, where discriminative classifiers can be used.

Using a generative embedding involves three steps:

(i) define and learn the generative model used to build the embedding;
(ii) define the mapping from object space to the generative embedding space;
(iii) discriminatively learn a (maybe kernel) classifier on the adopted feature space.

NOTE The literature on generative embeddings is essentially focused on step (i) and (ii), usually adopting some standard off-the-shelf tool (e.g., an SVM with a linear or RBF kernel) for step (iii).

THE PROPOSAL

Here we follow a different route, testing the recently proposed non-extensive information theoretic kernels on several Hidden Markov Models-based generative embeddings.

The Information Theoretic Kernels

Given two probability measures \( p_1 \) and \( p_2 \) representing two objects, we tested several information theoretic kernels (ITKs) [Martins et al. 09]:

1. \( k_1^n \) - Jensen-Shannon kernel:
\[
k_1^n(p_1, p_2) = \ln(2) - JS(p_1, p_2),
\]
with \( JS(p_1, p_2) \) being the Jensen-Shannon divergence
\[
JS(p_1, p_2) = \frac{H(p_1) + H(p_2)}{2} - \frac{H(p_1 \oplus p_2)}{2},
\]
\( H(p) \) is the usual Shannon entropy.

2. \( k_2^n \) - Jensen-Tsallis (JT) kernel:
\[
k_2^n(p_1, p_2) = \ln_q(2) - T_q(p_1, p_2),
\]
where \( \ln_q(x) = (x^{1/q} - 1)/(1 - q) \) is the \( q \)-logarithm.
\[
T_q(p_1, p_2) = S_q\left(\frac{p_1 + p_2}{2}\right) - \frac{S_q(p_1) + S_q(p_2)}{2^q},
\]
is the Jensen-Tsallis \( q \)-difference, and \( S_q(r) \) is the Jensen-Tsallis entropy, defined, for a multinomial \( r = (r_1, \ldots, r_k) \), with \( r_i \geq 0 \) and \( \sum r_i = 1 \), as
\[
S_q(r_1, \ldots, r_k) = \frac{1}{q - 1} \left( 1 - \sum r_i^q \right).
\]

3. \( k_3^n \) and \( k_4^n \) two versions of the Jensen-Tsallis kernel applicable to unnormalized measures (see the paper for more details).

Experimental Evaluation

Details

- Tests on a 2D shape recognition task (shapes are characterized with a sequence of curvature values), with the Chicken Pieces Database (446 silhouettes of chicken pieces - 5 classes). Accuracies computed with averaged hold out CV (10 repetitions)
- 3-state HMMs with Gaussian emission densities
- SVM with IT kernels on generative embeddings
- \( C \) of SVMs and \( q \) of the information theoretic kernels were optimized by 10-fold cross validation (CV) on the training set

Results

![Classification accuracies](image)

(Left) Classification accuracies. "(S)" refer to experiments where the embeddings were standardized (centered and scaled to unit variance). (Right) SVM accuracies with several kernels for the Transition Embedding, as a function of \( q \).

Comparative Analysis

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Accuracy (%)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-NN + Euclidean edit distance</td>
<td>96.47</td>
<td>[19]</td>
</tr>
<tr>
<td>1-NN + approximated cyclic distance</td>
<td>96.78</td>
<td>[18]</td>
</tr>
<tr>
<td>KNN + cyclic string edit distance</td>
<td>74.3</td>
<td>[19]</td>
</tr>
<tr>
<td>SVM + Edit distance-based kernel</td>
<td>91.11</td>
<td>[21]</td>
</tr>
<tr>
<td>1-NN + multi-based features</td>
<td>76.05</td>
<td>[6]</td>
</tr>
<tr>
<td>1-NN + HMM-based distance</td>
<td>77.3</td>
<td>[6]</td>
</tr>
<tr>
<td>SVM + HMM-based entropic features</td>
<td>81.22</td>
<td>[21]</td>
</tr>
<tr>
<td>SVM + HMM-based Top Kernel</td>
<td>80.88</td>
<td>[22]</td>
</tr>
<tr>
<td>SVM + HMM-based FESS embedding + rbf</td>
<td>83.30</td>
<td>[22]</td>
</tr>
<tr>
<td>SVM + HMM-based non linear Marginalized Kernel</td>
<td>85.5</td>
<td>[6]</td>
</tr>
<tr>
<td>SVM + HMM-based clustered Fisher kernel</td>
<td>85.38</td>
<td>[6]</td>
</tr>
</tbody>
</table>

Comparative kernels on the Chicken database.

Acknowledgements: We acknowledge financial support from the FET programme within the EU FP7, under the SIMBAD project (Contract 213250) and from Fundação para a Ciência e Tecnologia (FCT) grant PTDC/EIA-TEL/72572/2006.