Label Noise-Tolerant Hidden Markov Models

Benoît Frénay, Gaël de Lannoy and Michel Verleysen

September 6, 2011
Introduction

For real datasets, perfect labelling is difficult:

- **subjectivity** of the labelling task;
- **lack** of information;
- communication **noise**.

In particular, label noise arise in **biomedical** applications.

Previous works by e.g. Lawrence et al. incorporated a noise model into a generative model for i.i.d. observations (classification).
Label noise in the case of sequential data modelled by HMMs:

- a new label-noise tolerant algorithm is proposed;
- experiments are carried on ECG signals;
- the interest of the proposed approach is shown.
Hidden Markov Models in a Nutshell

**HMM**: description of the relationship between an unobservable sequence of hidden states $S$ and an observable sequence $O$.

![Diagram showing the relationship between hidden states $S_1, S_{t-1}, S_t, S_{t+1}, \ldots, S_T$ and observable sequences $O_1, O_{t-1}, O_{t+1}, \ldots, O_T$.]

**Parameters** $\Theta = (q, a, b)$:

- $q_i$ is the prior of state $i$;
- $a_{ij}$ is the transition probability from state $i$ to state $j$;
- $b_i$ is the observation distributions for state $i$.

Here, $b_i$ are **Gaussian mixture models** (GMMs).
Standard Inference Algorithms for HMMs

Supervised learning:
- assumes the observed labels are correct;
- maximises the likelihood $P(S, O|\Theta)$;
- learns the correct concepts;
- sensitive to label noise.

Baum-Welch algorithm:
- unsupervised, i.e. observed labels are discarded;
- iteratively (i) label samples and (ii) learn a model;
- may learn concepts which differs significantly;
- theoretically insensitive to label noise.
Label Noise Model

Two distinct sequences of states:
- the observed, noisy annotations \( Y \);
- the hidden, true labels \( S \).

The annotation probability is

\[
\begin{align*}
d_{ij} &= \begin{cases} 
1 - p_i & (i = j) \\
p_i / (|S| - 1) & (i \neq j)
\end{cases}
\end{align*}
\]

where \( p_i \) is the expert error probability in \( i \).
Label Noise-Tolerant HMMs

**Compromise** between supervised learning and Baum-Welch.

- assumes the observed labels are **potentially noisy**;
- **maximises** the likelihood $P(Y, O | \Theta)$;
- learns the **correct** concepts;
- **less sensitive** to label noise.
Non-convex function to optimise:

\[
\log P(O, Y|\Theta) = \log \sum_S P(O, Y, S|\Theta),
\]

Solution: **EM algorithm**.

**Expectation step**: estimate the posteriors

\[
\gamma_t(i) = P(S_t = i|O, Y, \Theta^{old})
\]

\[
\epsilon_t(i, j) = P(S_{t-1} = i, S_t = j|O, Y, \Theta^{old})
\]
Maximisation Step (parts of)

Maximisation step for $p_i$:

$$p_i = \frac{\sum_{t \mid Y_t \neq i} \gamma_t(i)}{\sum_{t=1}^{T} \gamma_t(i)}$$

Maximisation step for $\mu_{il}$:

$$\mu_{il} = \frac{\sum_{t=1}^{T} \gamma_t(i, l) o_t}{\sum_{t=1}^{T} \gamma_t(i)}$$

The true labels are estimated and used to compute the parameters.
Application: Electrocardiograms

**Electrocardiograms** (ECGs):

- periodic signal measuring the **electrical activity** of the **heart**;
- **patterns**: P waves, QRS complexes, T waves and B3 baseline;

![ECG Diagram]

**Preprocessing**:

- **filtered** using a 3-30 Hz band-pass filter;
- transformed using a continuous **wavelet transform**;
- dyadic scales from $2^1$ to $2^7$ are kept and normalised.
Experimental Settings

EM algorithms:
- GMM with 5 components;
- EM algorithms are repeated 10 times;

Electrocardiograms:
- a set of 10 artificial ECGs;
- 10 ECGs selected in the sinus MIT-QT database;
- 10 ECGs selected in the arrhythmia MIT-QT database.

Comparison:
- learning with addition of artificial label noise;
- comparison on original signals;
- label noise moves the boundaries of P and T waves.
Results for Artificial ECGs

Supervised learning, \textit{Baum-Welch} and label noise-tolerant.
Results for Sinus ECGs

Supervised learning, Baum-Welch and label noise-tolerant.
Results for Arrhythmia ECGs

Supervised learning, Baum-Welch and label noise-tolerant.
Supervised learning:
• affected by increasing label noise.

Baum-Welch:
• worst results for small levels of noise;
• less affected by the label noise
• better than supervised learning for large levels of noise.

Label-noise tolerant algorithm:
• affected by increasing label noise;
• most often better than Baum-Welch;
• better than supervised learning for large levels of noise.
Conclusion

An *EM* algorithm for *label noise-tolerant HMM inference* is proposed and compared with supervised learning and Baum-Welch.

**Experiments** on *healthy* and *pathological* ECGs signals show:

- all approaches are adversely *impacted* by *label noise*;
- the proposed algorithm can yield *better performances*.

**Future work** includes

- *testing* other types of label noise;
- *comparing* algorithms on other problems.