Project, Statistical Methods in Applied Computer Science, Fall 2016, Due Jan. 26, 2017

December 20, 2016

This project should be done in pairs. Describe your results in a short report using at most 5 pages.

The train: Imagine a model railway with a single train. You know the map of the tracks including the position of all the switches, but you do not know current states of the switches, or where the train is currently located. Each switch has three connections: $0, L, R$. If the train comes from the direction of L or R, it always leaves in the direction 0. If the train comes from the direction 0, it will leave in either direction L or R, depending on the state of the switch. The switch has prior probability $1/2$ for each direction, but will remain the same throughout the train run.

You are receiving a stream of signals from the train, each signal specifying the direction in which train has passed a switch: $0L, 0R, L0$, or $R0$; you do not know, however, which switch the train has passed. Also, the sensors are noisy, and with a certain probability p, the train reports a random other signal than real direction in which it passed the switch.

We are given G, which is undirected and all vertices have degree 3. At each vertex the edges are labeled $0, L$, and R (an edge can have different labels at a pair of neighboring vertices). So a vertex is a switch. Start positions and, as mentioned above, switch settings have uniform priors.

A switch setting is a function $\sigma : V(G) \to \{L, R\}$, which has the natural interpretation. By a position we mean a pair (v, e), where $v \in V(G)$ and $e \in E(G)$, with the interpretation that the train has passed the vertex v and exited through the edge e, so it is actually an edge together with a direction of travel.

Below we give a DP algorithm for $p(s, O|G, \sigma)$. We will estimate $p(\sigma|G, O)$ using MCMC and then $p(s|G, O)$ using

$$p(s|G, O) = \sum_{\sigma} p(s, \sigma|G, O) = \sum_{\sigma} p(s|\sigma, G, O)p(\sigma|G, O)$$

$$= \sum_{\sigma} p(s, O|G, \sigma)p(\sigma|G, O)/p(O|G, \sigma)$$

The probability p is 0.05. Given G, σ, $s \in V(G)$ (s is a stop position), $O \in L, R, 0$ (observed switch signals), we can compute $p(s, O|G, \sigma)$ using DP. The formulation actually induces a HMM. The natural way to compute this probability is to do DP and in each step compute the probability of going from some position s' to s in t steps and observing $o_1 \ldots, o_t$ when the switch settings are σ. By doing this for all stop positions s and then summing out both the start and the stop position or only the stop we obtain $p(O|G, \sigma)$ and $p(s', O|G, \sigma)$, respectively, in time $O(N^2T)$, where $N = |V(G)|$.

The states in our HMM are positions. The transition probabilities are always 1, i.e., given how we enter a vertex it is uniquely determined how we exit (since switches are fixed). Also, when passing a switch the correct direction of the label of the position is emitted with probability $1-p$ and any different direction is emitted with probability $p/2$.

Let $c(s,t)$ be computed as below (we want $c(s,t)$ to be the probability of going from some position s' to $s = (v,e)$ in t steps and observing $o_1 \dots, o_t$). Let $f = (u,v)$ and $g = (w,v)$ be the two edges that are incident with v but different from e.

- $c(s,0) = 1/N$
- If e is labeled 0 and $o_t = 0$, then $c(s,t) = [c((u,f),t-1) + c((w,g),t-1)](1-p)$.
- If e is labeled 0 and $o_t \neq 0$, then $c(s,t) = [c((u,f),t-1) + c((w,g),t-1)]p$.
- If e is labeled L, the switch at v is set to L, $o_t = L$, and f is labeled 0 at v (i.e., w.l.o.g. assume the latter), then $c(s,t) = c((u,f),t-1)(1-p)$.
- If e is labeled L, the switch at v is set to L, $o_t \neq L$, and f is labeled 0 at v, then $c(s,t) = c((u,f),t-1)p$.
- If e is labeled R, the switch at v is set to R, $o_t = R$, and f is labeled 0 at v, then $c(s,t) = c((u,f),t-1)(1-p)$.
- If e is labeled R, the switch at v is set to R, $o_t \neq R$, and f is labeled 0 at v, then $c(s,t) = c((u,f),t-1)p$.
- If e is labeled L and the switch at v is set to R, then $c(s,t) = 0$.
- If e is labeled R and the switch at v is set to L, then $c(s,t) = 0$.

Generate graphs G, e.g., by taking a $n \times n$ grid an shortcutting degree 2 vertices and splitting degree 4 vertices, so that each degree 4 vertex gives rise to 2 degree 3 vertices. Assign labels randomly or in a specific fashion. Generate synthetic data and analyse it. Report convergence, how well you do etcetera.