Some Initial Ideas

We have a bilingual conversation between (at least) two talkers. Considering only one talker, we observe over the stream of words occurring at times $i \in \{1, 2, 3, \ldots\}$ the use of either language $\mathbf{L} = \{L_1, L_2, \ldots\}$ where $L_i \in \{0, 1\}$. Concurrent with the (dependent) variable L_i we observe a vector of covariates $\mathbf{X}_i$ that includes part-of-speech, the language spoken by the conversational partner, and so forth. The goal is to predict *change points,* or the values of i where $L_i \neq L_{i+1}$.

Simplest Case

We model the L_is as a logistic autoregressive (AR) model of (presently) unknown order n. For now, assume $n = 1$. Then for the AR(1) model

$$Z_i \mid L_{i-1} = \beta_0 + \beta L_{i-1} + \alpha_0 Z_{i-1} + \alpha' \mathbf{X}_i + \epsilon_i,$$

where β_0 is a constant intercept, β is the coefficient weighting the previous value of L, α_0 is the autoregressive coefficient, α is a vector of covariate regression weights, and $\epsilon_i \sim \mathcal{N}(0, \sigma)$. Letting

$$l(x) = \frac{1}{1 + \exp(-x)},$$

the logistic function, then the transition probability matrix describing the change point probabilities at time i is

$$\mathbf{P}_i = \left[\begin{array}{cc} l\left(Z_i \mid L_{i-1} = 0\right) & 1 - l\left(Z_i \mid L_{i-1} = 0\right) \\ 1 - l\left(Z_i \mid L_{i-1} = 1\right) & l\left(Z_i \mid L_{i-1} = 1\right) \end{array} \right].$$

While the transition probabilities change with i, the regression coefficients do not. Diebold et al. (1994) suggest that estimation of such model parameters should be accomplished using the expectation maximization (EM) algorithm.

This proposal assumes a Markov-like structure with time-varying transition probabilities. Because these probabilities are a function of all previous L_is (through the AR model), it is not Markovian. I think this is a good thing but it might introduce some difficulties. I don't know offhand what those difficulties might be.

Less Simple Models

I had originally imagined a (meta)model in which a conversation was a path through a network of concepts C_j, $j = 1, 2, \ldots$, each C_j requiring a sequence of words W_k, $k = 1, 2, \ldots, n_j$, to express. This in turn could be represented as another path through two overlapping semantic networks, one for $L1$ and the other for $L2$, where the concept C_j

“induces” (restricts) the set of possible words that can be chosen. I thought that one important factor determining a code switch at any point in the sequence would be the distances between the current location in the word path and the next desired concept-induced word in the two semantic networks: there should be a strong tendency to select the closest word, regardless of the network to which that word belonged.

Such a hierarchical structure would require a lot more effort to model. Would we want to take a graph theory approach? Could we still use a logistic AR model? Could we think about it as a hidden Markov model with those concepts C_j as the latent Markov variable? What would the word-level effect be; could we condense it down to simply part-of-speech? I have no ideas here.

My second thoughts are focused on that hierarchical structure. Going back to the simple AR(1) model, if we go Bayesian we could model both talkers in the conversation simultaneously, which would be cool and way easier than the hidden (non)Markov approach. This is what I would vote for at this time, unless Jorge thinks all this is crazy.

My third thoughts are that I have no idea what I'm doing. I don't know the current state of modeling in this area or what factors are considered the most important. But I hope this can serve as a starting point for discussion.

Reference

Diebold, Francis X., Lee, Joon-Haeng and Weinbach, Gretchen C. (1994): Regime Switching with Time-Varying Transition Probabilities, in C. Hargreaves (Ed), *Non-Stationary Time Series Analysis and Cointegration* (pp. 284-302). Oxford University Press.