Time changing baseline predictors

A template for a time sensitive baseline predictor for u's rating of i at day t_{ui} reads:

$$b_{ui} = \mu + b_u(t_{ui}) + b_i(t_{ui})$$

Here, b_u and b_i are real valued functions that change over time. It adequate to split the item biases into time-based bins, using a constant item bias for each time period. The decision of how to split the timeline into bins should balance the desire to achieve finer resolution (hence, smaller bins) with the need for enough ratings per bin (hence, larger bins). A day t is associated with an integer Bin(t) (a number between 1 and 30 in our data), such that the movie bias is split into a stationary part and a time changing part

$$b_i(t) = b_i + b_{i,\,\mathrm{Bin}(t)}$$

One simple modeling choice uses a linear function to capture a possible gradual drift of user bias. For each user u, we denote the mean date of rating by tu. Now, if u rated a movie on day t, then the associated time deviation of this rating is defined as

$$\mathrm{dev}_u(t) = \mathrm{sign}(t - t_u) \cdot |t - t_u|^{\beta}.$$

Here $|t - t_u|$ measures the number of days between dates t and t_u, and the value of β is set by cross validation. A time-dependent user-bias is obtained

$$b_u^{(1)}(t) = b_u + \alpha_u \cdot \mathrm{dev}_u(t)$$

A more flexible parameterization is offered by splines. Let u be a user associated with n_u ratings. We designate k_u time points – $\left\{t_1^u, \cdots, t_{k_u}^u\right\}$ – spaced uniformly across the dates of u's ratings as kernels that control the following function

$$b_u^{(2)}(t) = b_u + \frac{\sum e^{-\sigma|t - t_u|} b_{t_1}^u}{\sum e^{-\sigma|t - t_u|}}$$

The parameters $b_{t_l}^u$ are associated with the control points (or, kernels), and are automatically learned from the data.

Beyond the temporal effects described so far, one can use the same methodology to capture more effects. A primary example is capturing periodic effects.

$$b_i(t) = b_i + b_{i,\,\mathrm{Bin}(t)} + b_{i,\,\mathrm{period}(t)}$$

Another temporal effect within the scope of basic predictors is related to the changing scale of user ratings. While $b_i(t)$ is a user-independent measure for the merit of item i at time t, users tend to respond to such a measure differently. To address this, we add a time-dependent scaling feature to the baseline predictors, denoted by $c_u(t)$. Thus, the baseline predictor

$$b_{ui}(t) = \mu + b_u + \alpha_u \cdot \mathrm{dev}_u(t_{ui}) + b_{u,t_{ui}} + \left(b_i + b_{i,\,\mathrm{Bin}(t_{ui})}\right) \cdot c_u(t_{ui})$$