Incomplete Label Multi-Task Ordinal Regression for Spatial Event Scale Forecasting (Supplemental Material)

Theorem 1's Proof

Theorem 1. *The optimal solution for matrix V can be obtained by computing its column vectors in order as follows:*

$$V_{\cdot,1} = y^{(2)}_{\cdot,1}/\rho + \Theta_{\cdot,1}$$

$$V_{\cdot,i} = \begin{cases} (\beta V_{\cdot,i-1} L + y^{(2)}_{\cdot,i} + \rho(\Theta_{\cdot,i} + V_{\cdot,i-1}) + y^{(3)}_{\cdot,i-1})(\beta L + 2\rho I)^{-1} & V_{\cdot,i} < V_{\cdot,i-1} \\ (\beta V_{\cdot,i-1} L + y^{(2)}_{\cdot,i} + \rho\Theta_{\cdot,i} + y^{(3)}_{\cdot,i-1})(\beta L + \rho I)^{-1} & V_{\cdot,i} \geq V_{\cdot,i-1} \end{cases}$$

Where $L = R^T R$ *and* $i = 2...k-1$.

Proof. Recall that the problem of update V in matrix multiplication format is as follows:

$$\underset{V}{\operatorname{argmin}} \frac{\beta}{2} \sum_{i=1}^{\mathcal{S}} \sum_{j=2}^{k-1} \left|\left|\left(V_{i,j} - V_{i,j-1}\right) R^T\right|\right|_2^2 + trace\left(y^{(2)}(\Theta - V)^T\right) + \frac{\rho}{2}||\Theta - V||_2^2 + \sum_{i=2}^{k-1} y^{(3)}_{\cdot,i}(V_{\cdot,i-1} - V_{\cdot,i})^T + \frac{\rho}{2} \sum_{i=2}^{k-1} ||max(V_{\cdot,i-1} - V_{\cdot,i}, 0)||_2^2$$

Where $R_{i,i} = 1$ and $R_{i,adj(i)} = -\frac{1}{N_i}$, for $i = 1...\mathcal{S}$. N_i is the total number of neighbors of task i.

For the sub-problem of solving $V_{\cdot,1}$, the analytical solution is fairly straight forward, since it is not involved in the max operator:

$$V_{\cdot,1} = y^{(2)}_{\cdot,1}/\rho + \Theta_{\cdot,1}$$

For each $V_{\cdot,i}(i > 1)$ there is a max operator, thus the derivative with respect to $V_{\cdot,i}(i > 1)$ lies on two situations as follows:

$$\begin{cases} \beta(V_{\cdot,i} - V_{\cdot,i-1})L - y^{(2)}_{\cdot,i} + \rho(V_{\cdot,i} - \Theta_{\cdot,i}) - y^{(3)}_{\cdot,i-1} + \rho(V_{\cdot,i} - V_{\cdot,i-1}) & V_{\cdot,i} < V_{\cdot,i-1} \\ \beta(V_{\cdot,i} - V_{\cdot,i-1})L - y^{(2)}_{\cdot,i} + \rho(V_{\cdot,i} - \Theta_{\cdot,i}) - y^{(3)}_{\cdot,i-1} & V_{\cdot,i} \geq V_{\cdot,i-1} \end{cases}$$

Where $L = R^T R$ and $i = 2...k-1$.

The above equations demonstrate that the analytical solution of $V_{\cdot,i}$ relies on $V_{\cdot,i-1}$. However, as we can obtain analytical solution for $V_{\cdot,1}$, we can get the analytical solution of $V_{\cdot,i}$ consecutively in ascending order. The analytical solution can therefore be computed as shown in Theorem 1

□

Baselines and parameters setting

1) *SVC1V1* (Support Vector Classifier with OneVsOne) is a C-Support Vector Classifier with OneVsOne binary decomposition. The kernel function is set to linear and the regularization parameter C is set based on 10 fold cross validation.

2) *SVC1VA* (Support Vector Classifier with OneVsAll) is a C-Support Vector Classifier with OneVsAll binary decomposition. These two methods represent the main approaches generally used to apply SVM to multi-class problems. The kernel function is set to linear and the regularization parameter C is set based on 10 fold cross validation.

3) *SVMOP* (Support Vector Machines with OrderedPartitions) is a C-Support Vector Classifier with OrderedPartitions binary decomposition. Note that this is a ordinal binary decomposition that takes class order information into consideration. The kernel function is set to linear and the regularization parameter C is set based on 10 fold cross validation.

4) *POM* (Proportional Odds Model) is a threshold based ordinal regression model. It has no tuning parameter.

Top Features Selected by MITOR

Table 1 shows the specific features (i.e., keywords) selected by the MITOR-I and MITOR-II models for the civil unrest and influenza outbreak datasets; these models both utilize group sparsity $\ell_{2,1}$ norms on the feature vector. According to Table 1, MITOR-I and MITOR-II select features that are very relevant to civil unrest and influenza, and this selection is stable and consistent across different countries. This demonstrates the effectiveness of including the group sparsity $\ell_{2,1}$ norm in proposed model.

Table 1: Top 10 static features (keywords) selected by MITOR-I and MITOR-II for both the civil unrest and influenza outbreak datasets (all keywords have been translated to English using Google Translate)

Civil unrest								U.S. flu	
Argentina	Brazil	Chile	Colombia	Mexico	Paraguay	Uruguay	Venezuela	2011-2012	2013-2014
petition	syndicalism	vigil	presence	pronounce	authorization	aggravate	process	aching	doc
political	work	police	traitors	wage demand	vice president	rob	participants	fly	banget
salary increase	war	violence	protection	government	development	official	privacy	antibiotics	activity
order	pose	crisis	burden	atrocity	raw materials	to catch	convulsed	doc	appetite
human rights	mobilization	victim	agrarian reform	protest	criminal act	manifestation	to give up	careful	battling
interest	may plaza	sanctions	pro family	police	victims	make a rally	war	meds	ache
wage demand	injured	hurt	angry	investigation	expensive	policy	authorize	activity	struck
terrorists	enemy	ethnicity	police	retirees	to present	to threaten	delinquency	ache	aching
discrimination	conspiracy	prohibition	refineries	paramilitaries	refuge	justice	conflict	achy	sinus
attendees	humiliate	protest	damaged	protection	precious metal	wage increase	unrest	struck	achy