TECHNICAL REPORT: On Recognition of Seasonal Predictability in SLIGRO product sales

Indrė Žliobaitė

July 20, 2009

Abstract

A summary of new results on SLIGRO sales prediction.

1 The task

We aim to predict one week ahead the quantity of weekly product sales, aggregated over all SLIGRO locations.

2 What's new since ICDM paper

The modifications since ICDM paper are the following:

1. Product categorization into 2 classes: 'predictable' and 'unpredictable' (previously we had 4).
2. Product categorization using the Linear Discriminant Classifier. Reasons: interpretable weight vector (importance of features), unique model for a given dataset, easy to manipulate prior probabilities and costs of missclassification. We train the model assuming equal prior probabilities, but in fact we have about 4 times more of 'unpredictable' products. Then we bias the discrimination threshold towards 'unpredictable' class, using a threshold -0.002 which was learned experimenting with the training data to match the prior probabilities of the classes.
3. We changed normalization and categorization of the time series to more domain driven approach. Now normalization works as follows:

$$x_t^* = \frac{x_t}{\text{mean}(x_1, \dots, x_T)}, \tag{1}$$

   here x_t is the original signal at time t and x_T is the last instance of the training data.

4. simple change detection is introduced (as a result of discussions with Joao). There are a lot of products with 'hills' or 'valleys' right in the middle (see Figure 1). This in fact corresponds to the new year. We guess that they renegotiate contracts with the suppliers. Thus the naive approach is to cut the training history at those points and start training from scratch if the change at the new year is significant. Change detection criterion is the following:

$$K = \frac{2|m_1 - m_2|}{m_1 + m_2}, \tag{2}$$

$$\text{if } K > 1.0 \text{ then a change is detected,} \tag{3}$$

here m_1 and m_2 are the means of the training data coming from different years. The detection threshold is set based on visual inspection, selecting the changing products from the training data. For now we detect only the fist change and automatically assume that there is the second change if there was one.

3 The results

We present two collections of results. The first shows the predictability of a category under new setup under static conditions (i.e. the category is assigned once for a given product). The second illustrates the idea of dynamic category selection online.

We use the same structural features, the same product set and the same split into training and testing as in ICDM paper. The categorization rule (linear discriminant classifier) is trained and selected using 5 fold cross validation on the training data.

3.1 Predictability of a category

For predictability of a category we present three result tables. The first one shows the accuracies for training data. The second shows accuracies for the testing data. The third table shows accuracies after a random categorization.

The final results are averaged over 50 runs. That means the experiment is repeated 50 times, each time random assigning into 5 crossvalidation set is different, therefore we come up with slightly differently parameterized classifiers.

We include 6 weeks moving average as a benchmark to see how the current company method would perform.

The results are provided in Table 1.

We train the linear discriminant on the structural features derived from a full length of the training data. The main delimitation in this setup is that we also use the full length to derive the structural features of the testing data.

The main observation and conclusion is that in 'predictable' class the intelligent method outperforms both naive predictors.

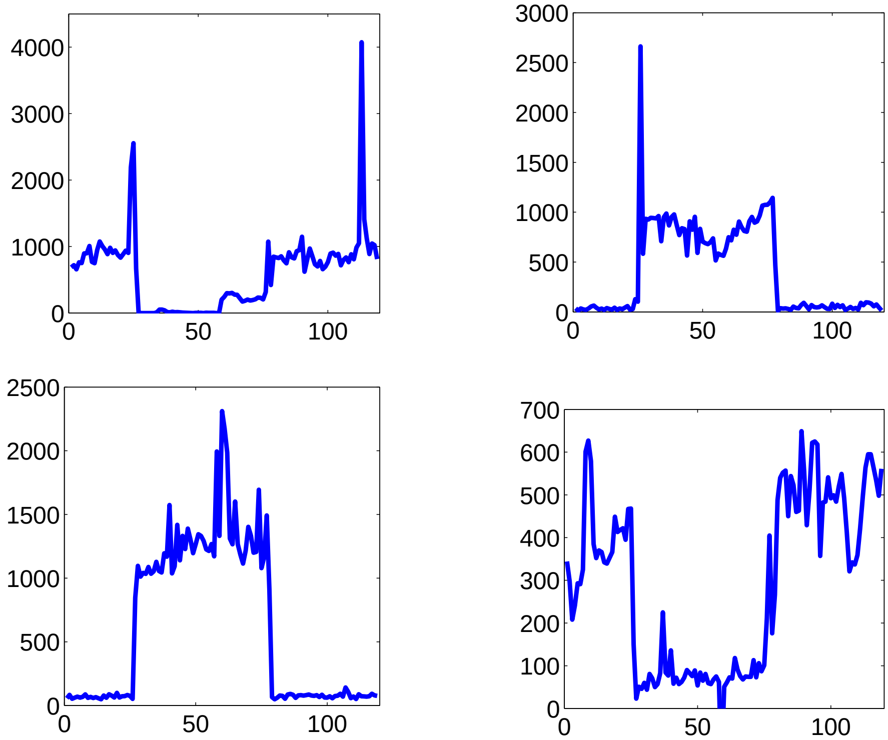

Figure 1: Examples of new year changes in product sales, x axis: time (weeks), y axis quantity (units).

3.2 Dynamic category selection

In this experiment we train the linear discriminant developing the features only from one year sales history to match the testing setup and the reality more closely. For the linear discriminant again 5 fold cross validation is used. Here we increase the threshold of the decision linear discriminant to -0.004 to reduce the number of false positives.

We test three scenarios on the testing data:

1. The product category is assigned once and forever based on the structural features developed from the first year of the sales history.

2. The product category is reassigned every week based on a moving window (one year size) of sales history.

3. The product category is reassigned every week based on incremental learning. Every week a new sales data is added, thus the history window is expanding.

	naive	6 weeks MA	intelligent	average size
Training data				
‘unpredictable’	1.000	1.341	1.730	323
‘predictable’	1.000	0.987	**0.940**	115
Testing data				
‘unpredictable’	1.000	1.383	1.762	80
‘predictable’	1.000	0.970	**0.950**	20
Random clustering				
‘unpredictable’	1.000	1.306	1.604	50
‘predictable’	1.000	1.296	**1.593**	50

Table 1: Predictability: scaled mean absolute errors.

	Testing result
Scenario 1, static	1.0136
Scenario 2, dynamic, fixed window	0.9952
Scenario 3, dynamic, incremental window	0.9953

Table 2: Online categorization: scaled mean absolute errors.

The dominance of scenario 2 would mean that the category is indeed changing during this small period of 2 years. The dominance of scenario 3 would mean that the accuracy of categorization increases when we have more historical samples available.

The results are provided in Table 2. The results are averaged over 100 runs.

Scenario 2 and 3 clearly outperform scenario 1, that implies that online assignment of categories is beneficial. Scenario 2 (windowing) is on average slightly better than scenario 3, that suggests possible concept drift within a given product.

Now let us look how many products have were changing the category. In Figure 2 we plot the bars of predictability for testing products (100 products). On y axes ‘1’ means that the product was assigned to a ‘predictable’ category, ‘0’ means ‘unpredictable’. We average the assignments for a given product over all the 60 testing weeks and then average over 100 runs. Note that the graphs are based on the linear discriminant outputs, not on the ‘ground truth’.

It can be concluded that there are many heterogeneous products thus the system accuracy benefits from online categorization.

4 What products are predictable?

Let us have a closer look, what products are predictable. In Figure **??** we plot the six most accurately predictable products by the intelligent methods. The scaled mean absolute errors are correspondingly $0.61, 0.64, 0.72, 0.78, 0.83, 0.87$

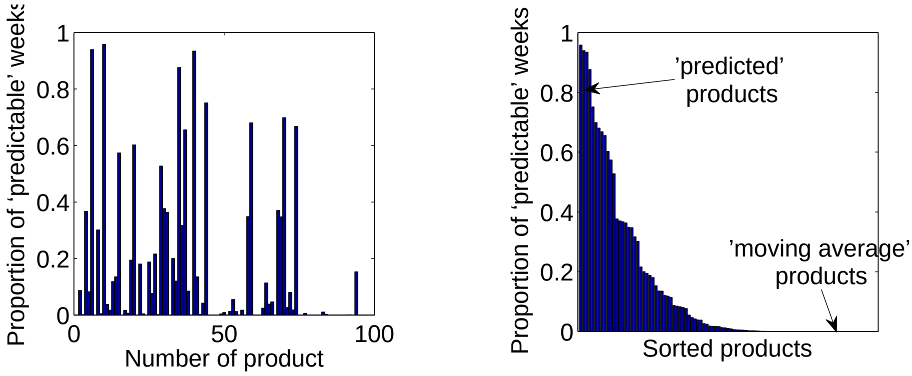

Figure 2: Online categories of the testing products: (a) in order, (b) sorted.

(the product numbers from the testing set are 13,64,29,44,48,65).

Most of the predictable products seem chaotic from visual inspection (except maybe the 5th one, which is rather periodic). However, all the top products seem visually alike in their behavior. To explore the behavior further, in Figure 4 we depict the same best product after normalization of the historical series.

It would be interesting to check, what are these actual products (Jorn?). The results suggest, that there is a room for intelligent data mining methods.

Now let as look what are the least predictable products, i.e. where the intelligent classifier makes the largest error. The six worst are depicted in Figure **??**. The errors are correspondingly $8.50, 4.11, 4.09, 3.53, 3.50, 3.42$ and the product numbers are 82,81,77,97,91,18. At least from the order it seems that there is a consistency with the real product categories (Jorn?). It can be also seen from the graph a in Figure 3 where all the predictability is concentrated on the left side (and we have the products numbered correspondingly to their real categories).

It can be seen that most of the 'troublesome' products have a hill pattern at the new year. As it was discussed, change detection could contribute to dealing with those.

5 Conclusion

- We did not take into account new year changes, which are systematic. Change detection helped to increase accuracy a bit. Currently very basic approach is implemented, it might be worth exploring this direction. We also saw that the 'hill' peaks form the least predictable group of products.
- The test results show that it is possible to predict the categories accurately enough. Still we have too little training data to learn the structural features. One year is not enough to observe annual seasonality.

- We showed that online reassignment of the categories help to increase prediction accuracy. Therefore, it is likely that each product is not homogeneous and it would be worth exploring the regions of predictability within the product to employ online classifier switching.
- We spoten a tendency of the predictable and unpredictable products to be clustered correspondingly to the real categories (to be checked which).

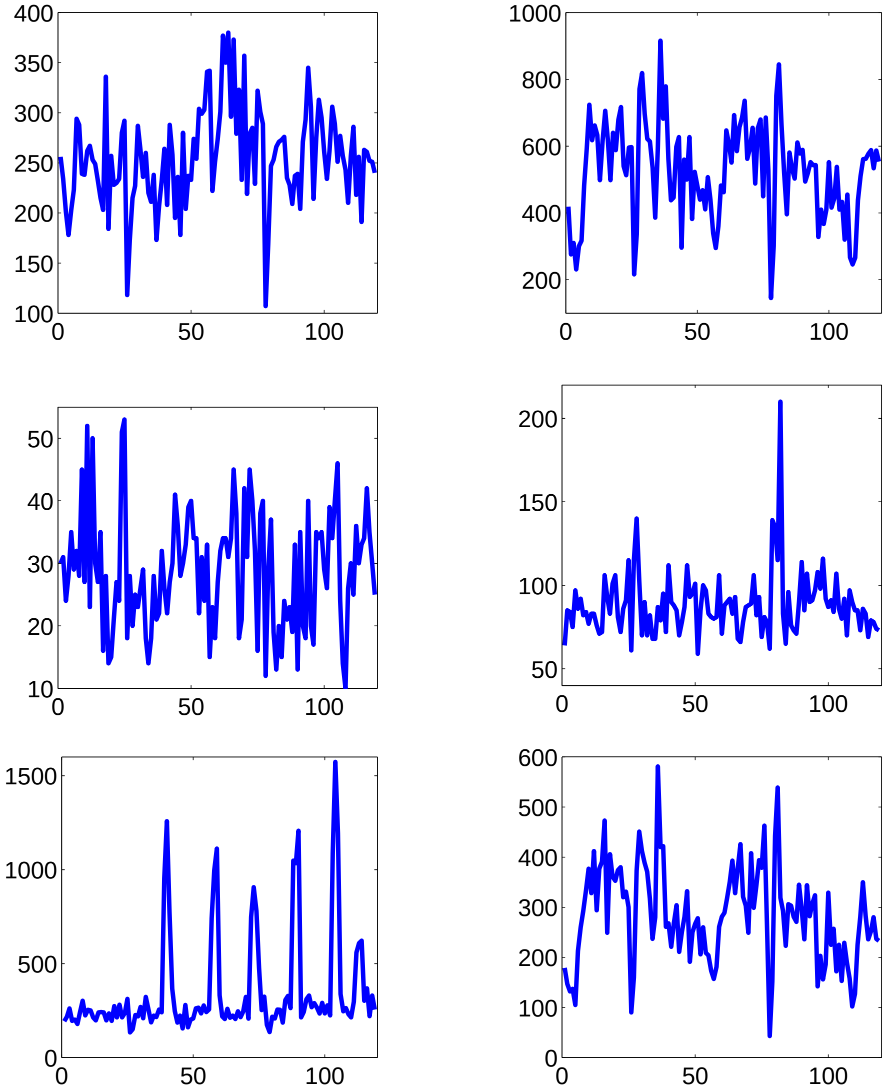

Figure 3: The top 6 *predictable* products.

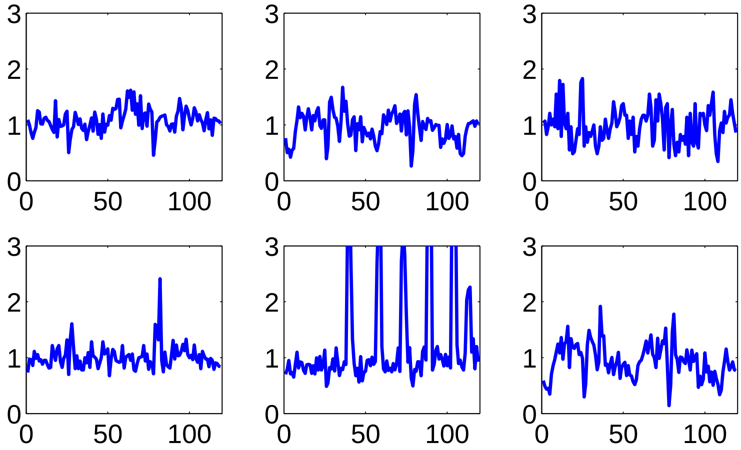

Figure 4: *Predictable* products scaled.

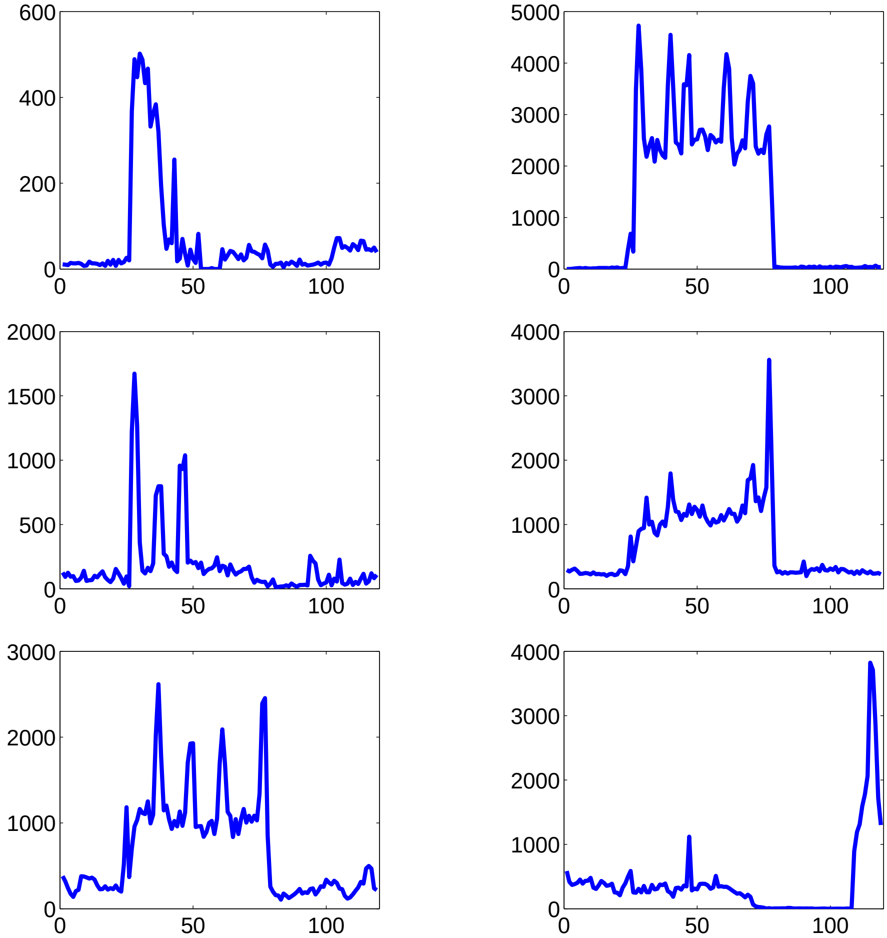

Figure 5: The 6 most *unpredictable* products.