

# Попытки учесть каннибализацию

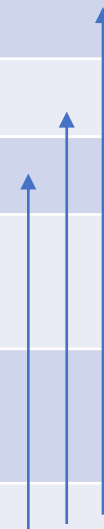
Д. Звежинский, 12.12.2018

[Dmitry.Zvezhinsky@sas.com](mailto:Dmitry.Zvezhinsky@sas.com)

# Иерархическое прогнозирование

- Краткая суть:
  - Выбирается иерархия прогнозирования (ИП) – в ней смешиваются иерархии товаров и магазинов.
  - На \*каждом\* уровне этой иерархии \*каждый\* ВР прогнозируется своей моделью (на основании только своей истории и независимых переменных). При этом независимые переменные могут содержать коллективную информацию с уровней ниже.

Уровень ИП	Целевая переменная	Предикторы
Все тов-все маг	СУММА(спрос)	СУММА(флаг промо) - ????, СРЕДНЕЕ(цена) - ????, СРЕДНЕЕ(температура) - ????
Группа тов-область	СУММА(спрос)	...
Группа тов-город	СУММА(спрос)	...
Категор тов-город	СУММА(спрос)	СУММА(флаг промо), СРЕДНЕЕ(цена), СРЕДНЕЕ(температура)
Категор тов-маг	СУММА(спрос)	СУММА(флаг промо), СРЕДНЕЕ(цена), СРЕДНЕЕ(температура)
Товар-маг	спрос	Флаг промо, цена, температура



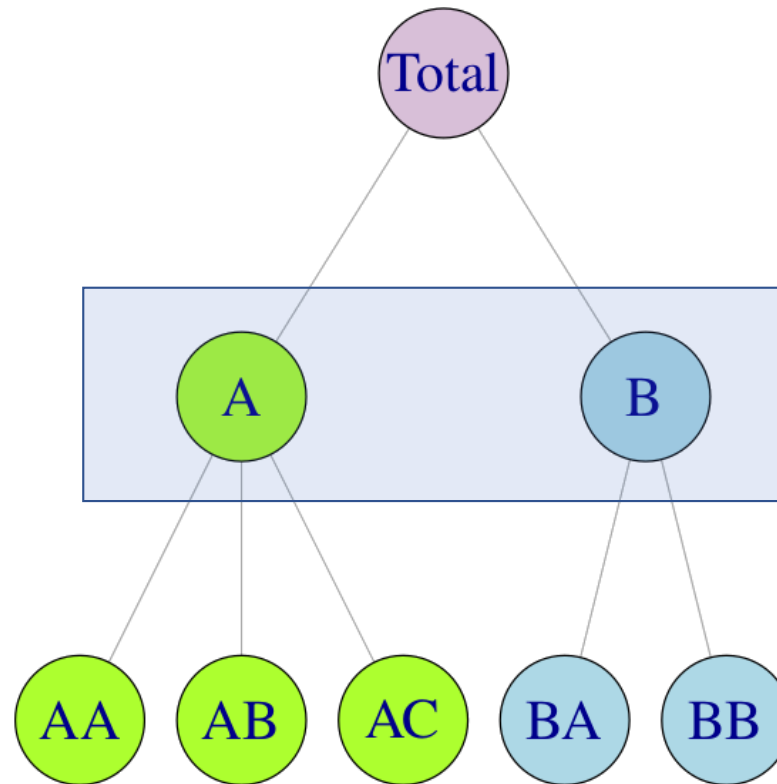
# Иерархическое прогнозирование

- Продолжение...

- Выбирается уровень согласования: именно прогнозу на этом уровне мы доверяем больше всего. Прочие прогнозы согласуются относительно этого уровня. Также выбирается метод вычисления поправок [1].

Прогноз ДО согласования

	1 янв	2 янв	3 янв	4 янв
TOT	10	12	8	9
A	5	4	3	4
B	4	5	3	4
AA	1	2	99	2
AB	2	3	3	2
AC	3	1	1	0
BA	1	2	5	2
BB	2	1	2	1



Прогноз ПОСЛЕ согласования  
(метод Forecast proportions)

	1 янв	2 янв	3 янв	4 янв
TOT	9	9	6	8
A	5	4	3	4
B	4	5	3	4
AA	0.833333	1.333333	2.883495	2
AB	1.666667	2	0.087379	2
AC	2.5	0.666667	0.029126	0
BA	1.333333	3.333333	2.142857	2.666667
BB	2.666667	1.666667	0.857143	1.333333

# Иерархическое прогнозирование - проблемы

- Видно, что нельзя просто так «поднимать» те же предикторы с нижних уровней ИП при помощи агрегатов, т.к. теряется их смысл. Скорее всего, на верхних уровнях должны быть другие предикторы.
- Выбор иерархии прогнозирования (ИП) и уровня согласования – нетривиальные вопросы
- Обоснованный выбор наиболее точного метода согласования осуществляется при помощи кросс-валидации (пакет HTS для R). На практике такая кросс-валидация неосуществима из-за объема данных.

# Иерархическое прогнозирование - проблемы

- Учет каннибализации в рамках ИП - через модели прогнозирования спроса. Однако, сами модели для индивидуальных рядов могут не учитывать какие-то значимые драйвера каннибализации (из-за автоматического выбора моделей).
- Группы взаимно каннибализирующих рядов:
  - Фиксируются выбором уровня согласования в ИП
  - Привязаны к неизменной ИП – не факт, что она подходит для этого наилучшим образом

# Учет каннибализации (обзор статьи\*)

- Ординальные атрибуты продуктов (тип/размер/объем упаковки, цена) – драйверы каннибализации.
- Каннибализация происходит внутри определённых групп (product group) при вводе новых товаров.
- Данные:
  - Data for the case study relate to sales of a consumer beverage product, in terms of weekly sales volume over 120 weeks and dollar value from eight to ten stores across the USA. They covered between 400 and 600 varieties of the product in each store, and between 35 and 45 brands. In addition, each brand had between two and six product families, and each product family had between two and four product groups with multiple instances of products within a product group.

\* Marketing Intelligence & Planning Vol. 23 No. 5, 2005 pp. 470-485

# Учет каннибализации

- Фаза 1. Какие товары будут «жертвами»?
- Фазы 2,3. Ранжирование «жертв» по схожести с новым товаром.
- Фаза 4. Вклад каннибализации (loss of sales) из-за нового товара для  $i$ -го старого товара:

$$Y_i = \sum_{j=1}^{\text{атр.товаров}} w_j \left(1 - \frac{f_j}{\psi_j}\right)$$

$w$  = вес атрибута  $j$  (сумма весов нормирована на 1)

$N_j$  = число различных значений атрибута  $j$

$f_j = (1/N_j) * (N_j - (\text{ранг атр. } j \text{ нового товара} - \text{ранг атр. } j \text{ товара } i))$

$\psi_j$  = число товаров с тем же атр.  $j$ , что и у нового товара.

# Учет каннибализации

- Сила эффекта: с ростом числа схожих атрибутов между товарами ожидается бОльшее падение продаж старых товаров,
- Падение продаж пропорционально:  
(число знач. атр.  $j$  – (ранг атр.  $j$  нового товара – ранг атр.  $j$  товара  $i$ ))  
*(attribute level similarity),*
- Падение продаж обратно пропорционально числу товаров с одинаковым значением атрибута (таким же, как у нового товара).

При этом:

- Поправка  $Y$  не меняется со временем и не зависит от истории спроса.
- Не зависит от интересных драйверов (promo-активность, цены), т.к. главный драйвер - ввод нового товара.
- Простые зависимости от атрибутов (нет кросс-эффектов).



# Учет каннибализации

- Выводы:
  - Схожесть между товарами можно вычислить по рангам атрибутов. Это помогает выделить группу с наибольшей вероятностью каннибализации.
    - Не все атрибуты можно ранжировать
    - Атрибуты = ручной ввод
    - Атрибуты – это разреженные данные. Для разных групп товаров могут быть свои атрибуты.
    - А можно ли вычислить схожесть по истории продаж/спроса?
  - Чем больше товаров в группе, которую каннибализирует новый товар, тем меньше эффект на каждый товар.
    - Спорно.
  - Обратите внимание на достигнутый эффект:

It was observed that in this case study the proposed model reduced the over-prediction of ARIMA by between 28 percent and 46 percent

# Домашнее задание

- Методы нахождения групп взаимно каннибализирующих товаров (Наталья):
  - Тест Грэйнджера на наличие причинности во временных рядах.  
([https://en.wikipedia.org/wiki/Granger\\_causality](https://en.wikipedia.org/wiki/Granger_causality)) Подходит ли он нам?  
(кажется, нет)
  - DTW (<https://www.cs.unm.edu/~mueen/DTW.pdf>),
  - Прочие метрики схожести временных рядов (Time series similarity):  
<https://quant.stackexchange.com/questions/848/time-series-similarity-measures>