What can pre-release search traffic profiles tell us?

Oliver Schaer
Nikolaos Kourentzes
Robert Fildes

40th ISMS Marketing Science Conference

15th of June 2018

Marketing Analytics and Forecasting
Lancaster University Management School

Clustering for new product forecasting

Applications

- Model parameter clustering (Goodwin et al. 2013, Bayus, 1993)
- Time series shape clustering (Basallo-Triana et al. 2017)
- Product features clustering (Hu et al. 2018, Baardman et al. 2017)
- Functional clustering (Sood et al. 2009)

For pre-launch forecasting additional information needed, i.e. expert judgment

- Judgmental bias (Belvedere & Goodwin, 2017; Tyebjee, 1987)
- Consumer preferences change during pre-launch phase (Meeran et al. 2017)

Forecasting with pre-launch buzz

The literature reports positive findings when using online buzz for pre-launch forecasting (e.g. Kim and Hanssens 2017; Xiong & Bharadwaj 2014; Kulkarni et al. 2012)

- Predominately regression type models

Why clustering?

- Restricting into clusters increases homogeneity
- More robust to rely on cluster profile means rather individual forecasts?
- Regression based models are not readily applicable to obtain competitor product forecasts

Suggested approach

Create Clusters

Generate features from product characteristics, sales information and search traffic

Train classifier

Classify search traffic and known product features against obtained cluster solutions

Predict new product

Obtain the cluster profile of sales by allocating a new product to their respective clusters

Empirical evaluation

Dataset

- Global physical video game sales of 240 popular video games from VGChartz
- IGN Score and MetaCritic information
- Weekly Google Trends data with game title as keyword

Split data into training and test set (70% / 30%)

Objective is to assess the performance of cluster profiles for predicting the opening week and total life-cycle sales.

Obtaining cluster profiles

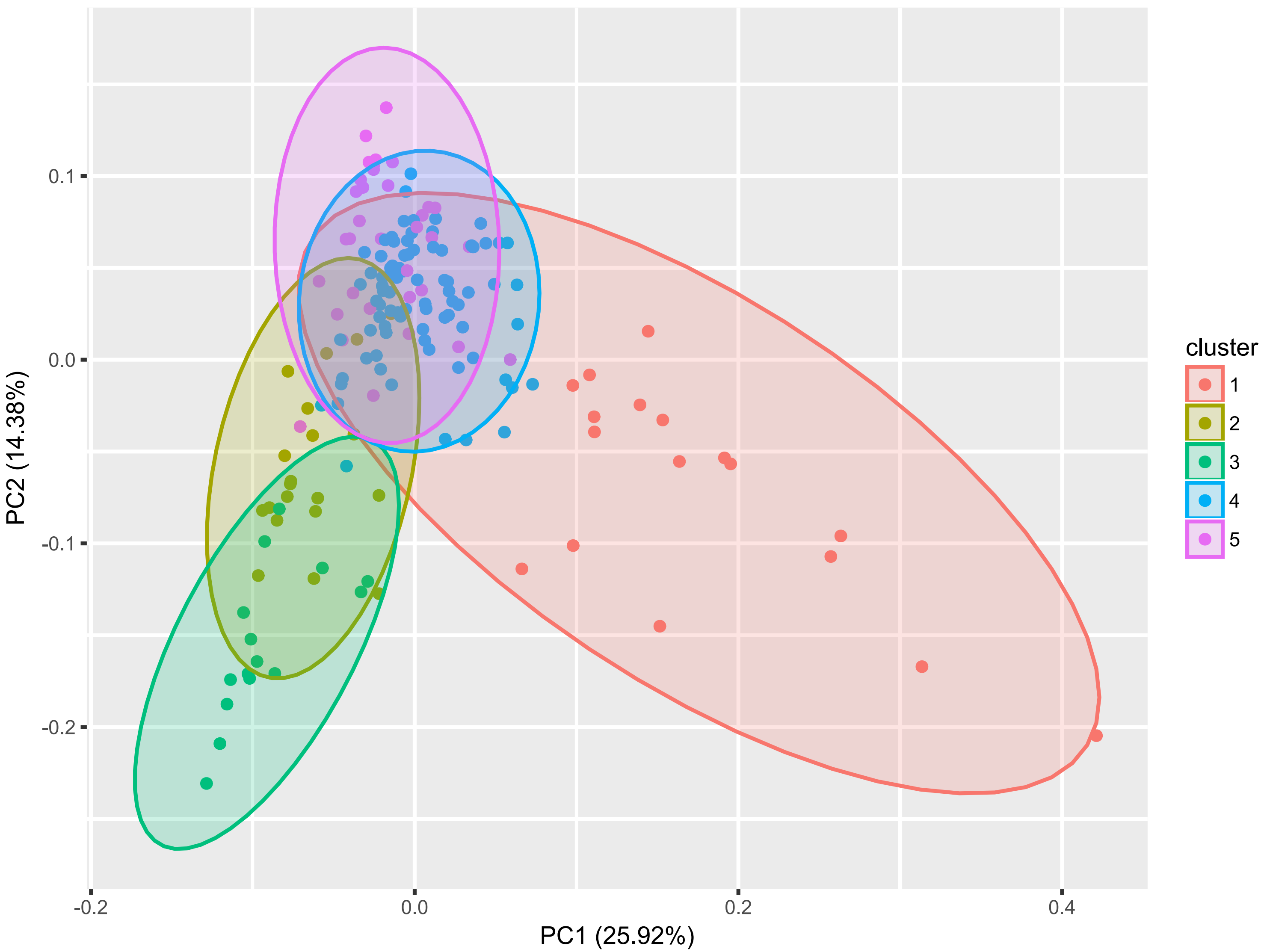
Features

- Product characteristics (publisher, genre, no. release, month)
- Sales information (actuals and time for life-cycle stages, Gompertz model parameters)
- Pre-release search (sum of search as soon signal is available, trend line, bass model parameters)

Estimation

- Mixed-data converted into binaries using Euclidean distances (and weighted product features)
- Various clustering algorithms including k-means, k-medoids, hierarchies and hyperplane cutting

k-means cluster solution



Some insights on cluster profiles (k-means)

Median cluster profiles

	Cluster 1 (n = 30)	Cluster 2 (n = 32)	Cluster 3 (n = 12)	Cluster 4 (n = 3)	Cluster 5 (n = 91)
log(First week sales)	11.370	12.016	15.358	16.326	12.787
log(Total Sales)	13.674	14.087	16.453	17.036	14.568

Some insights on cluster profiles

Median cluster profiles

	Cluster 1 (n = 30)	Cluster 2 (n = 32)	Cluster 3 (n = 12)	Cluster 4 (n = 3)	Cluster 5 (n = 91)
log(First week sales)	11.370	12.016	15.358	16.326	12.787
log(Total Sales)	13.674	14.087	16.453	17.036	14.568
Weeks until sales 25%	4.000	3.000	1.000	1.000	2.000
Weeks until sales 95%	44.500	34.500	19.000	14.000	24.000

Some insights on cluster profiles

Median cluster profiles

	Cluster 1 (n = 30)	Cluster 2 (n = 32)	Cluster 3 (n = 12)	Cluster 4 (n = 3)	Cluster 5 (n = 91)
log(First week sales)	11.370	12.016	15.358	16.326	12.787
log(Total Sales)	13.674	14.087	16.453	17.036	14.568
Weeks until sales 25%	4.000	3.000	1.000	1.000	2.000
Weeks until sales 95%	44.500	34.500	19.000	14.000	24.000
log(Total GT)	4.709	6.167	9.101	9.495	6.661
First signal	24.000	38.500	40.000	40.000	40.000
GT trend slope	2.421	5.864	118.284	200.755	10.779

Some insights on cluster profiles

Median cluster profiles

	Cluster 1 (n = 30)	Cluster 2 (n = 32)	Cluster 3 (n = 12)	Cluster 4 (n = 3)	Cluster 5 (n = 91)
log(First week sales)	11.370	12.016	15.358	16.326	12.787
log(Total Sales)	13.674	14.087	16.453	17.036	14.568
Weeks until sales 25%	4.000	3.000	1.000	1.000	2.000
Weeks until sales 95%	44.500	34.500	19.000	14.000	24.000
log(Total GT)	4.709	6.167	9.101	9.495	6.661
First signal	24.000	38.500	40.000	40.000	40.000
GT trend slope	2.421	5.864	118.284	200.755	10.779
IGN Score	7.975	7.917	8.779	8.000	8.250

Classification task

Random Forest to learn search traffic and product features against the obtained cluster solutions

Forecast exercise

- Benchmark against regression based Random Forest (same inputs on opening week and total life-cycle sales)
- Compare the MAE against the mean cluster profile

Forecasting process

Clustering

Inputs	**Output**
Sales information	K number of clusters
Google Trends information	
Product characteristics	

Classification

Inputs	**Output**
Google Trends information	Predicted cluster allocation
Product characteristics	

Benchmark

Candidate: Clustering	**Benchmark: Regression**
Mean of cluster profile	Product level forecast

MAE of opening week sales and total life-cycle sales

Overall forecasting performance

Relative Mean Absolute Error of clustering over regression

	k-means	k-medoids	Hierarchical	Hyperplanes
First week sales	1.141	1.178	**0.896**	1.709
Total life-cycle sales	**0.768**	0.890	1.750	1.210

Number of clusters, k = 5
All cluster inputs without diffusion model parameters

Conclusion

Modelling implications

- Extending the new product clustering literature by considering pre-launch search profiles.
- Outperforms more common regression based approach

Managerial implications

- More intuitive view on the drivers of the cluster profiles
- With known clusters any new product can be assessed, also those of our competitors

Next steps

Optimise the clustering via cross-validation for

- selecting the best clustering algorithm;
- find optimal number of clusters;
- obtain distance metric settings.
  → to improve the opening week performance

Forecasting competitor success only training clusters with company internal sales information

Assess different lead times of Google Trends

Thank you!

o.schaer@lancaster.ac.uk
@oliverschaer
www.lancaster.ac.uk/lums/cmaf

Marketing Analytics and Forecasting | Lancaster University Management School

References

Baardman, L., Levin, I., Perakis, G. & Singhvi, D. 2017. Leveraging Comparables for New Product New Product Forecasting. Available at SSRN: https://ssrn.com/abstract=3086237

Basallo-Triana, M.J., Rodriguez-Sarasty, J.A. & Benitez-Restrepo, H.D. 2017. Analogue-based demand forecasting of short life-cycle products. A regression approach and a comprehensive assessment. International Journal of Production Research, 55(8), 2336-2350.

Bayus, B.L. 1993. High-Definition Television. Assessing Demand Forecast for Next Generation Consumer Durable. Management Science, 39(11), 1319-1333.

Belvedere, V., Goodwin, P., 2017. The inuence of product involvement and emotion on short-term product demand forecasting. International Journal of Forecasting 33 (3), 652-661.

Goodwin, P., Dyussekeneva, K., Meeran, S., 2013. The use of analogies in forecasting the annual sales of new electronics products. IMA Journal of Management Mathematics 24 (4), 407-423.

Goodwin, P., Meeran, S., Dyussekeneva, K., 2014. The challenges of pre-launch forecasting of adoption time series for new durable products. International Journal of Forecasting 30 (4), 1082-1097.

References

Hu, K., Acimovic, J., Erize, F., Thomas D.J. & Van Mieghem, J.A. 2018. Finalist—2017 M&SOM Practice-Based Research Competition—Forecasting New Product Life Cycle Curves: Practical Approach and Empirical Analysis. Manufacturing & Service Operations Management (in press).

Kim, H., Hanssens, D. M., 2017. Advertising and word-of-mouth effects on pre-launch consumer interest and initial sales of experience products. Journal of Interactive Marketing 37, 57-74.

Kulkarni, G., Kannan, P., Moe, W., 2012. Using online search data to forecast new product sales. Decision Support Systems 52 (3), 604-611.

Sood, A., James, G.M. & Tellis, G.J. Functional Regression: A New Model for Predicting Market Penetration of New Products. Marketing Science, 28(1), 36-51.

Tyebjee, T. T., 1987. Behavioral biases in new product forecasting. International Journal of Forecasting 3 (3), 393-404.

Xiong, G. and Bharadwaj, S., 2014. Prerelease Buzz Evolution Patterns and New Product Performance. Marketing Science, 33 (3), 401-421.

Appendix: cluster profile of sales