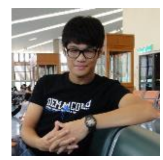
A Multimodality Approach to Predicting the Popularity of Sneakers

Mei-Chen Yeh* and Shao-Ting Yang
Department of Computer Science and Information Engineering
National Taiwan Normal University

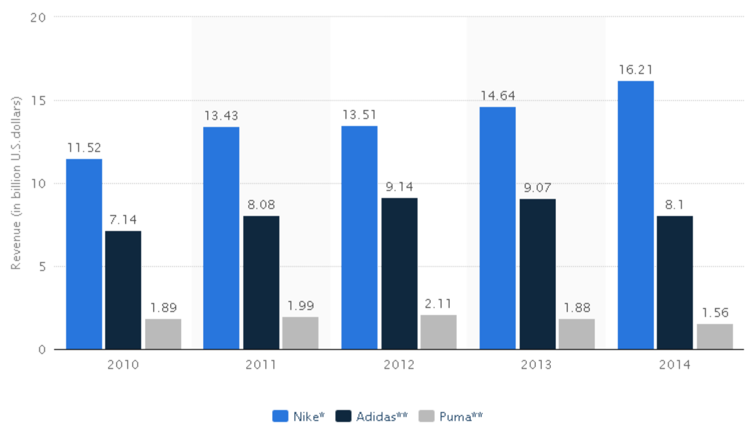
June 6, 2015

Intelligent fashion analysis




Revenue from footwear segment (in billion U.S. dollars)



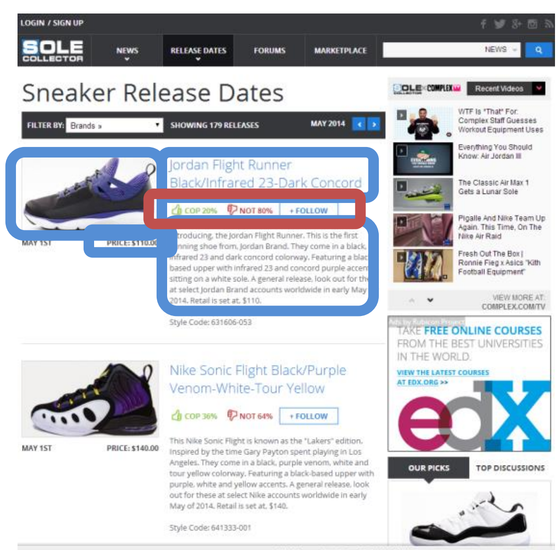
source: http://www.statista.com/

Sneaker design

Predicting the popularity of sneakers

- Objective
  - Rate a sneaker in terms of popularity
- Applications
  - Fashion analysis
  - Product search: relevance and quality
  - Product recommendation

solecollector.com

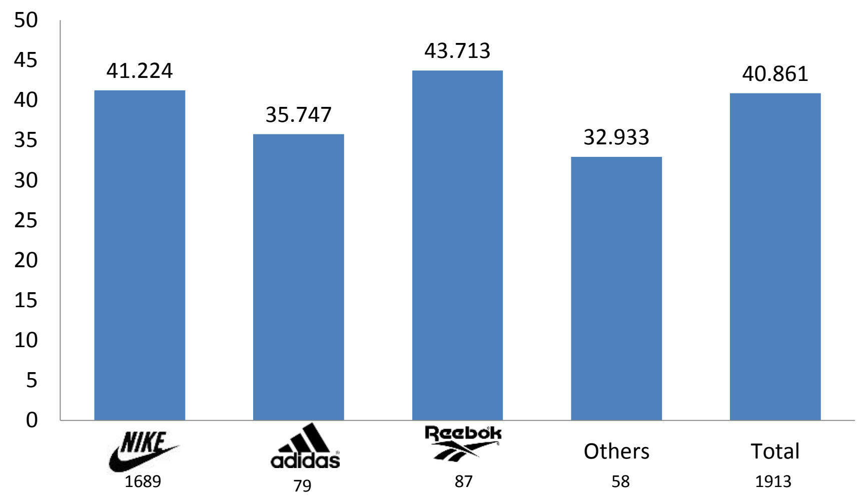


Dataset

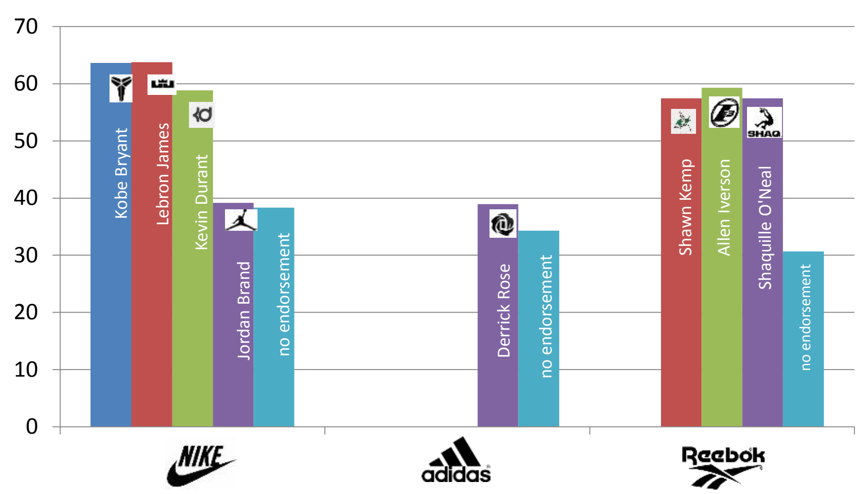
- Collected from solecollector.com
- 1913 products released from December 2010 to October 2013
- Population data (no sampling)

Sneaker design vs. COP score

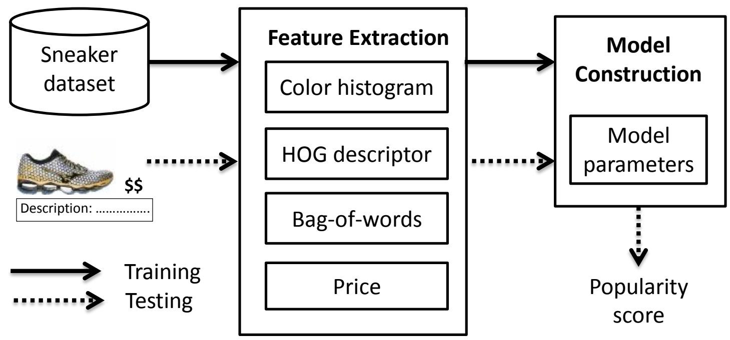
Brand vs. COP score



Product endorser vs. COP score



A multimodality approach



Feature extraction

Textual features

- Sneakers name & description: Bag-of-words model (873-d)
- Price (scalar)

Visual features

- Color: Histogram of pixels in HSV space (256-d)
- Shape: Histogram of oriented gradients (40-d)

Prediction model construction

- A regression problem:

  Given a set of training samples $D = \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)\}$, where $\boldsymbol{x}_i = (\boldsymbol{x}^T, x^P, \boldsymbol{x}^C, \boldsymbol{x}^S)$ and y_i is the desired score, find a solution for unknown model parameters $\boldsymbol{\vartheta}$ that minimizes the distortion between the measured and predicted COP scores.

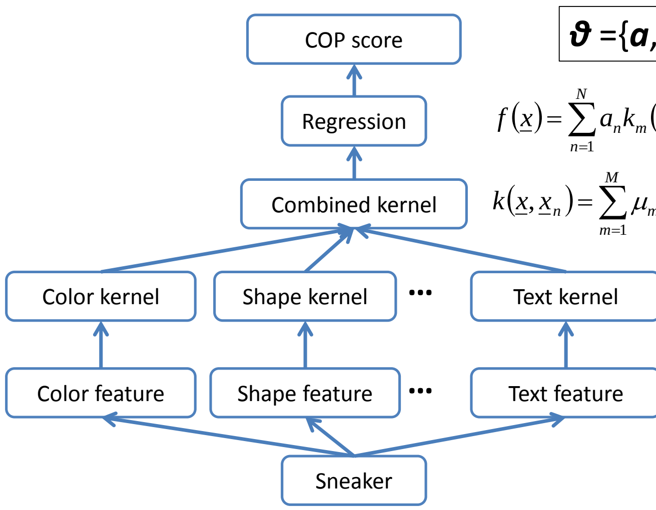
Kernel fusion + regression

Multiple-kernel support vector regression

$$\boxed{\boldsymbol{\vartheta} = \{\boldsymbol{a}, b, \boldsymbol{\mu}\}}$$

$$f(\underline{x}) = \sum_{n=1}^{N} a_n k_m(\underline{x}, \underline{x}_n) + b$$

$$k(\underline{x}, \underline{x}_n) = \sum_{m=1}^{M} \mu_m k_m(\underline{x}, \underline{x}_n)$$



Kernel functions

Textual features

- Bag-of-words model

chi-square kernel

- Price (scalar)

$$k_p\left(x^P, x_n^P\right) = 1 \Big/ \left(\frac{\left| x^P - x_n^P \right|}{c} + 1 \right)$$

Visual features

- Color: Histogram of pixels in HSV space (256-d)

radial basis function kernel

- Shape: Histogram of oriented gradients (40-d)

histogram intersection kernel

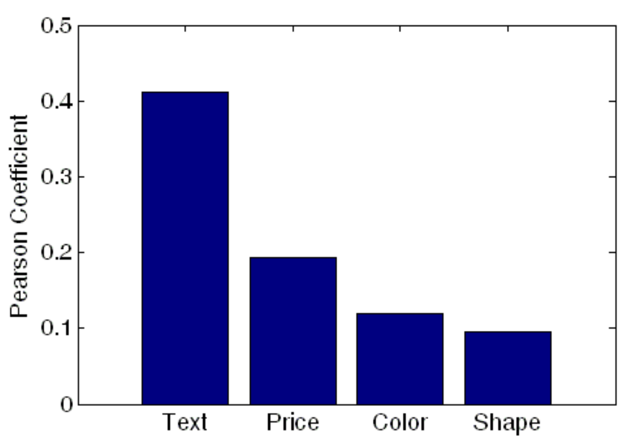
Research questions

- Which feature mostly affects the COP score?
- Does using kernel fusion result in a better prediction performance?
- Do customized kernels perform better than Radial Basis Function (RBF) kernels?

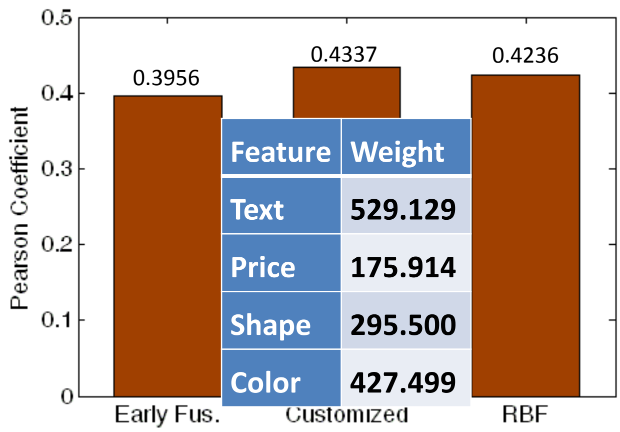
Experiment: Setup

- Dataset
  - 1913 products collected from solecollector.com
  - Training: 1703 (between Dec. 2010 and Aug. 2013)
  - Testing: 210 (between Sep. 2013 and Oct. 2013)
- Evaluation metrics
  - Pearson coefficient

Results: feature comparison



Results: model comparison



Feature	Weight
Text	529.129
Price	175.914
Shape	295.500
Color	427.499

Conclusion

- We obtained a satisfactory prediction result by using a few different facets to describe sneakers.
- Textual features play an important role.
- Compared with an early fusion approach, we show that a late fusion approach is more effective.
- Customized kernels in general perform better than the RBF kernels.

A Multimodality Approach to
Predicting the Popularity of Sneakers

More information:
http://www.csie.ntnu.edu.tw/~myeh/