Service Computation 2015 - Tutorial

Machine Learning and Software Quality Assurance

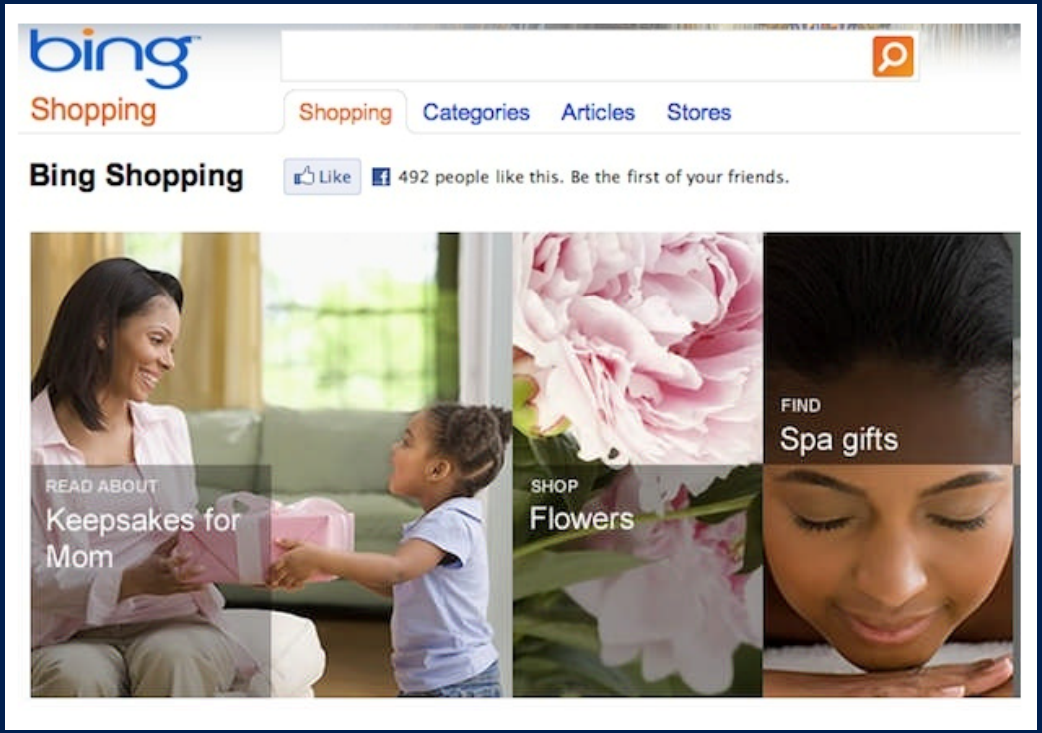
Marcelo De Barros, Principal Group Engineering Manager
Bing Experiences
Microsoft Corporation

This course is about case studies!

- **Case Study 1**: Finding outliers thru image analysis
  - Image Processing
  - **Unsupervised Clustering** Techniques
  - Statistics and some simple math
- **Case Study 2**: Finding "image-not-available" images
  - Image Processing
  - Utilization of a **Neural Networks**
  - Performance & Optimization
- **Case Study 3**: Mimicking production patterns
  - Use of **Markov Chains** to model user-behavior

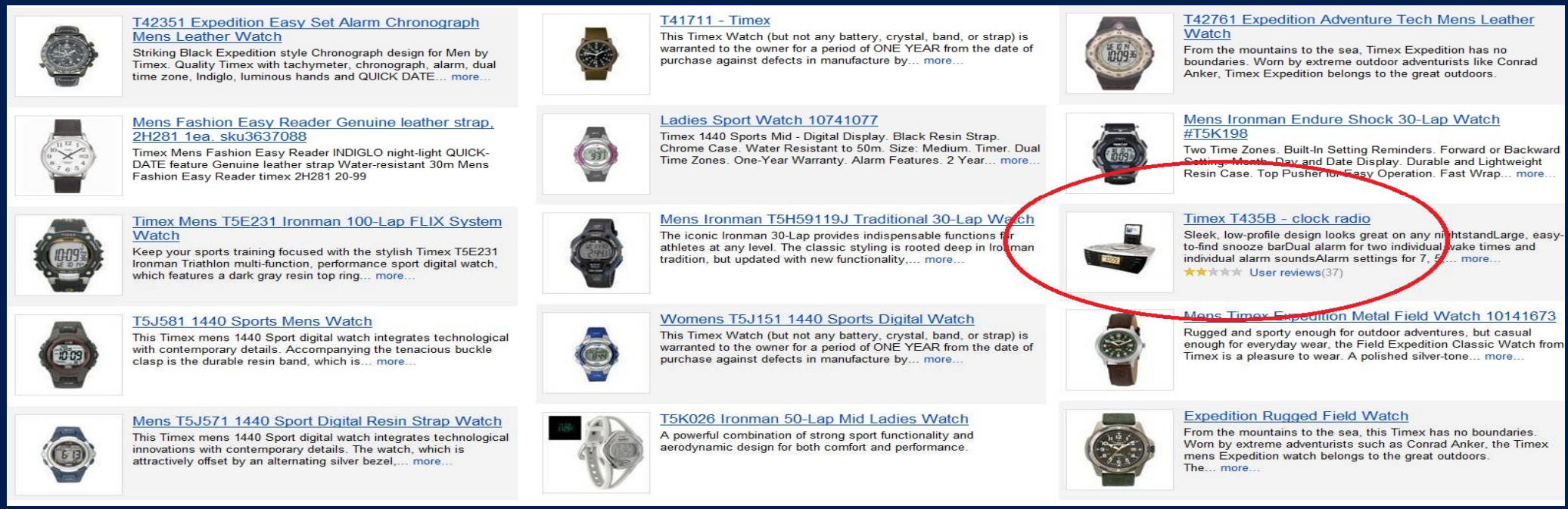
Case Study 1:

Detecting bugs on Bing Shopping/Images via clustering techniques



Example: query=*timex watch*

- What's wrong?



T42351 Expedition Easy Set Alarm Chronograph Mens Leather Watch
Striking Black Expedition style Chronograph design for Men by Timex. Quality Timex with tachymeter, chronograph, alarm, dual time zone, Indiglo, luminous hands and QUICK DATE... more...

Mens Fashion Easy Reader Genuine leather strap, 2H281 1ea. sku3637088
Timex Mens Fashion Easy Reader INDIGLO night-light QUICK-DATE feature Genuine leather strap Water-resistant 30m Mens Fashion Easy Reader timex 2H281 20-99

Timex Mens T5E231 Ironman 100-Lap FLIX System Watch
Keep your sports training focused with the stylish Timex T5E231 Ironman Triathlon multi-function, performance sport digital watch, which features a dark gray resin top ring... more...

T5J581 1440 Sports Mens Watch
This Timex mens 1440 Sport digital watch integrates technological with contemporary details. Accompanying the tenacious buckle clasp is the durable resin band, which is... more...

Mens T5J571 1440 Sport Digital Resin Strap Watch
This Timex mens 1440 Sport digital watch integrates technological innovations with contemporary details. The watch, which is attractively offset by an alternating silver bezel,... more...

T41711 - Timex
This Timex Watch (but not any battery, crystal, band, or strap) is warranted to the owner for a period of ONE YEAR from the date of purchase against defects in manufacture by... more...

Ladies Sport Watch 10741077
Timex 1440 Sports Mid - Digital Display. Black Resin Strap. Chrome Case. Water Resistant to 50m. Size: Medium. Timer. Dual Time Zones. One-Year Warranty. Alarm Features. 2 Year... more...

Mens Ironman T5H59119J Traditional 30-Lap Watch
The iconic Ironman 30-Lap provides indispensable functions for athletes at any level. The classic styling is rooted deep in Ironman tradition, but updated with new functionality,... more...

Womens T5J151 1440 Sports Digital Watch
This Timex Watch (but not any battery, crystal, band, or strap) is warranted to the owner for a period of ONE YEAR from the date of purchase against defects in manufacture by... more...

T5K026 Ironman 50-Lap Mid Ladies Watch
A powerful combination of strong sport functionality and aerodynamic design for both comfort and performance.

T42761 Expedition Adventure Tech Mens Leather Watch
From the mountains to the sea, Timex Expedition has no boundaries. Worn by extreme outdoor adventurists like Conrad Anker, Timex Expedition belongs to the great outdoors.

Mens Ironman Endure Shock 30-Lap Watch #T5K198
Two Time Zones. Built-In Setting Reminders. Forward or Backward Setting. Month, Day and Date Display. Durable and Lightweight Resin Case. Top Pusher for Easy Operation. Fast Wrap... more...

Timex T435B - clock radio
Sleek, low-profile design looks great on any nightstandLarge, easy-to-find snooze barDual alarm for two individual wake times and individual alarm soundsAlarm settings for 7, 5... more...
User reviews(37)

Mens Timex Expedition Metal Field Watch 10141673
Rugged and sporty enough for outdoor adventures, but casual enough for everyday wear, the Field Expedition Classic Watch from Timex is a pleasure to wear. A polished silver-tone... more...

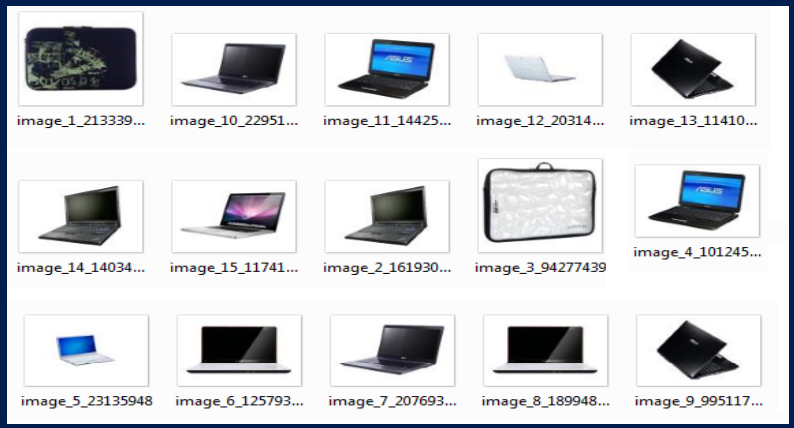
Expedition Rugged Field Watch
From the mountains to the sea, this Timex has no boundaries. Worn by extreme adventurists such as Conrad Anker, the Timex mens Expedition watch belongs to the great outdoors. The... more...

A Machine Learning Solution using Unsupervised Clustering!

- Understand the pattern of images using:
  1. Average and Standard Deviation Images
  2. A Penalty Function based on pixel proximity
  3. Variation of *Chauvenet's Criterion* to detect outliers

The Solution

1. Average and Standard Deviation Images



For each color (R,G,B) in each pixel, calculate:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + \cdots + x_n)$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

Average Image = μ

Higher Average Image = $\mu + \sigma$

Lower Average Image = $\mu - \sigma$

Lower Average Image <=  **Average Image** <=  **Higher Average Image**

This defines the pattern/cluster!

The Solution

2. A Penalty Function based on pixel proximity

For each Image X:

$$\text{Penalty}(\text{Pixel}(i)) = \begin{cases} \textbf{Pixel(i)} - \textbf{(Avg(i)+Std(i))}, & \textit{if Pixel(i)} > \textit{Avg(i)+Std(i)} \\ \textbf{(Avg(i)-Std(i))} - \textbf{Pixel(i)}, & \textit{if Pixel(i)} < \textit{Avg(i)-Std(i)} \\ \textbf{0}, & \textit{otherwise} \end{cases}$$

$$\text{Penalty}(X) = \sum \text{Penalty}(\text{Pixel}(i))$$

The Solution

2. A Penalty Function based on pixel proximity

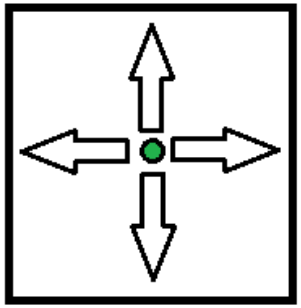
For each Image X:

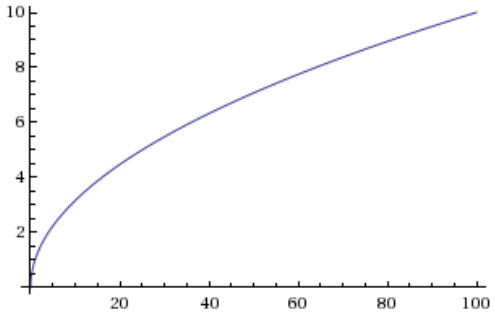
$$\text{Penalty}(\text{Pixel}(i)) = \begin{cases} \textbf{Pixel(i)} - \textbf{(Avg(i)+Std(i))}, & \textit{if Pixel(i)} > \textit{Avg(i)+Std(i)} \\ \textbf{(Avg(i)-Std(i))} - \textbf{Pixel(i)}, & \textit{if Pixel(i)} < \textit{Avg(i)-Std(i)} \\ \textbf{0}, & \textit{otherwise} \end{cases}$$

$$\text{Penalty}(X) = \sum \text{Penalty}(\text{Pixel}(i))$$

Variation:

Amortize the penalty based on the SQRT(Euclidian distance) of the Pixel to the center of the image

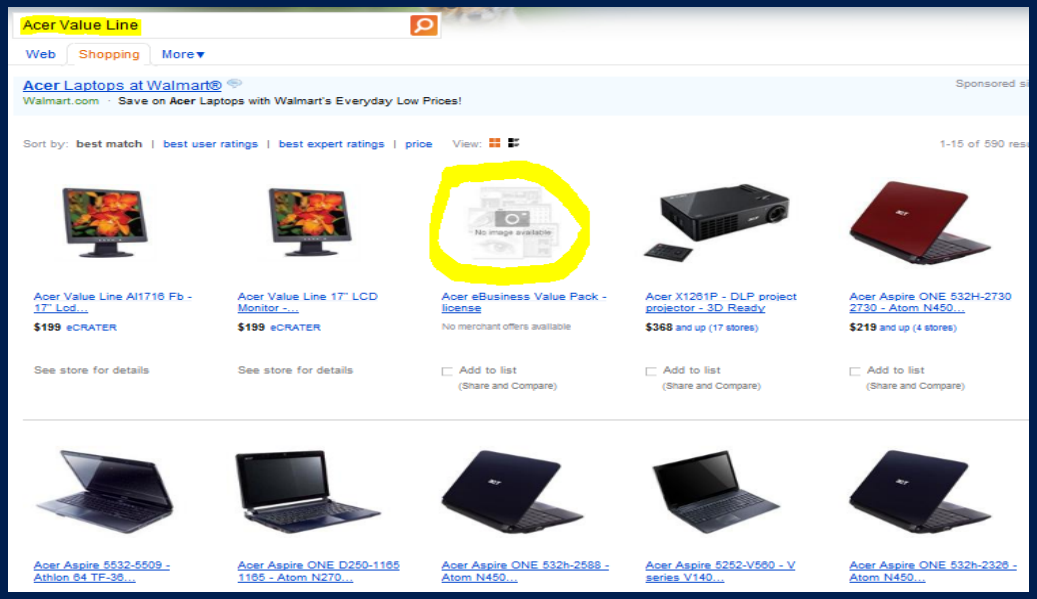


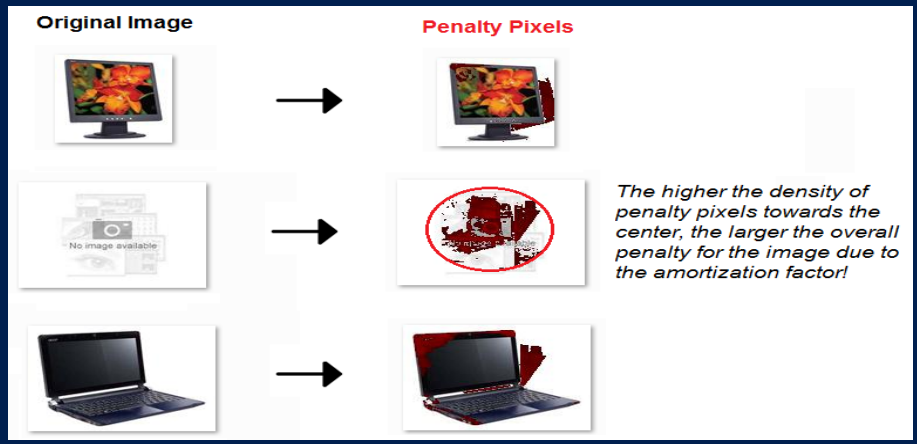


The Solution

2. A Penalty Function based on pixel proximity

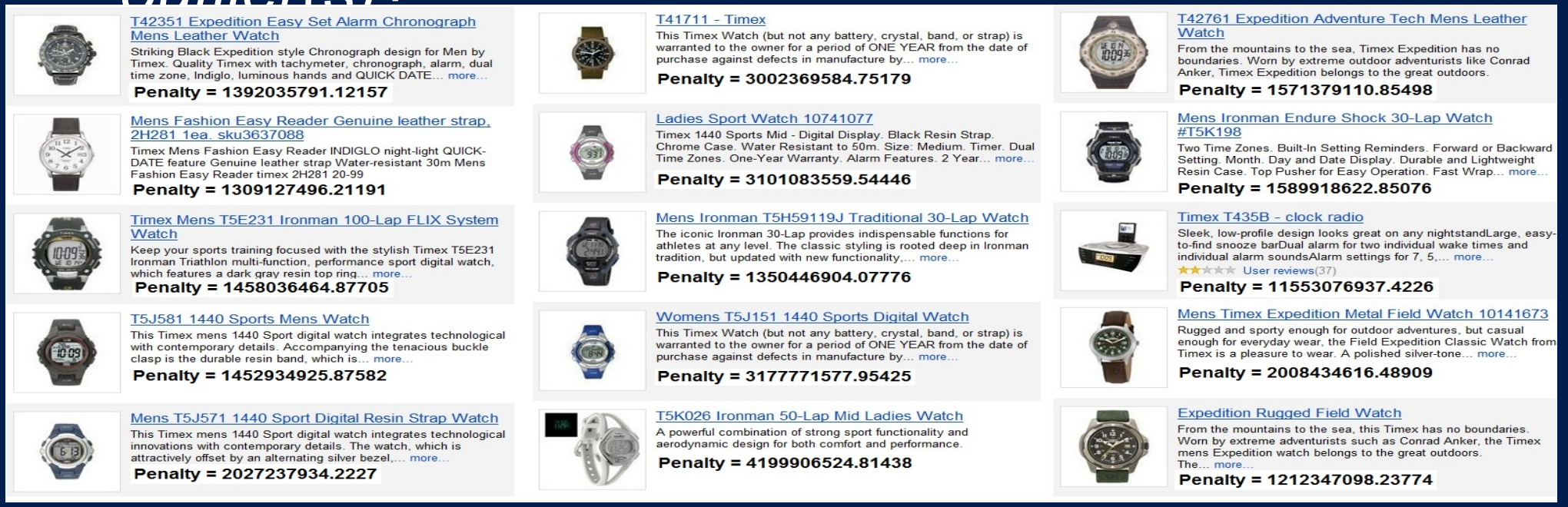
Visualizing the penalty





The Solution

- Now we have one number (penalty) associated with each image. How to find the *outlier(s)*?

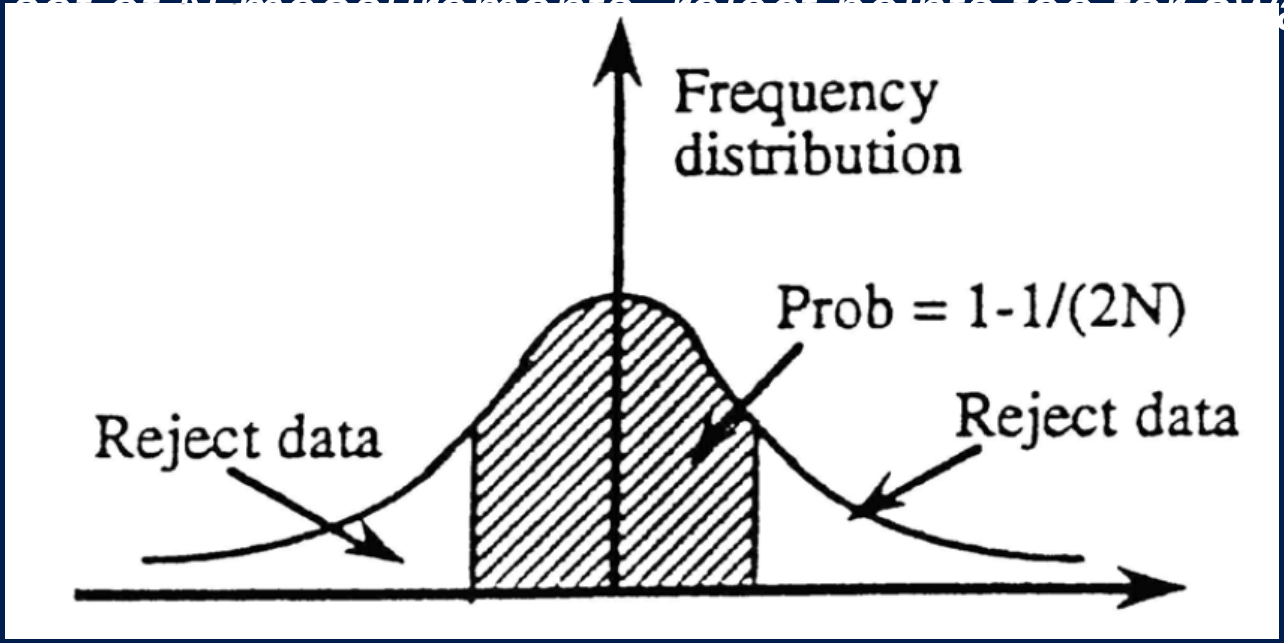


The Solution

- We now have a different (simpler) problem: given a set of numbers, how to find the outlier?
  - In some cases it is easy:
    - 1, 2, 4, 2, 3, 1, 2^{32}
  - But what about this set?
    - 1, 2, 4, 2, 3, 1, 7
- There are known techniques to statically determine outliers, although none of them are 100% accurate

The Solution

- Chauvenet's Criterion (http://www.ohio.edu/people/bayless/seniorlab/chauvenet.pdf)
- *Given a set of N measurements, reject points too far away from the mean*



The Solution

3. Variation of *Chauvenet's Criterion* to find outliers

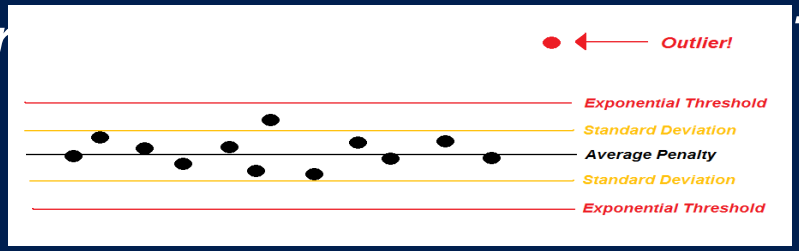
Principles:

i. The smaller the ratio Std/Avg, the narrower the margin for outliers

ii. The larger the ratio Std/Avg, the broader the margin for outliers

Variation implemented uses exponential distance based on the Penalty(X):

- Δ = *Std/Avg*
- ***T = (2$^{\Delta}$ + SEED)*Std + Avg***
- ***Outlier{X > T}***



SEED is configurable
Default: 0.0

Higher Seed	*Higher Precision*
Lower Seed	*Higher Recall*

The Solution - Lemmas

Lemma 1: **if the images are all identical, no outliers will be detected:**

Proof**:**

- *Std = 0 (identical images)*
- *Δ = Std/Avg ➔ Δ = 0*
- *T = (2^{Δ} + SEED)*Std + Avg ➔ T = (1 + SEED)*0 + Avg ➔ T = Avg*
- *Outliers = {Image X | Penalty(X) >= Avg}*
- *Penalty for any image = 0, hence*
- *Outliers = {Image X | 0 >= Avg}, since Avg > 0*
- *Outliers = {}*
- ***Therefore, no outliers will be detected!***

The Solution - Lemmas

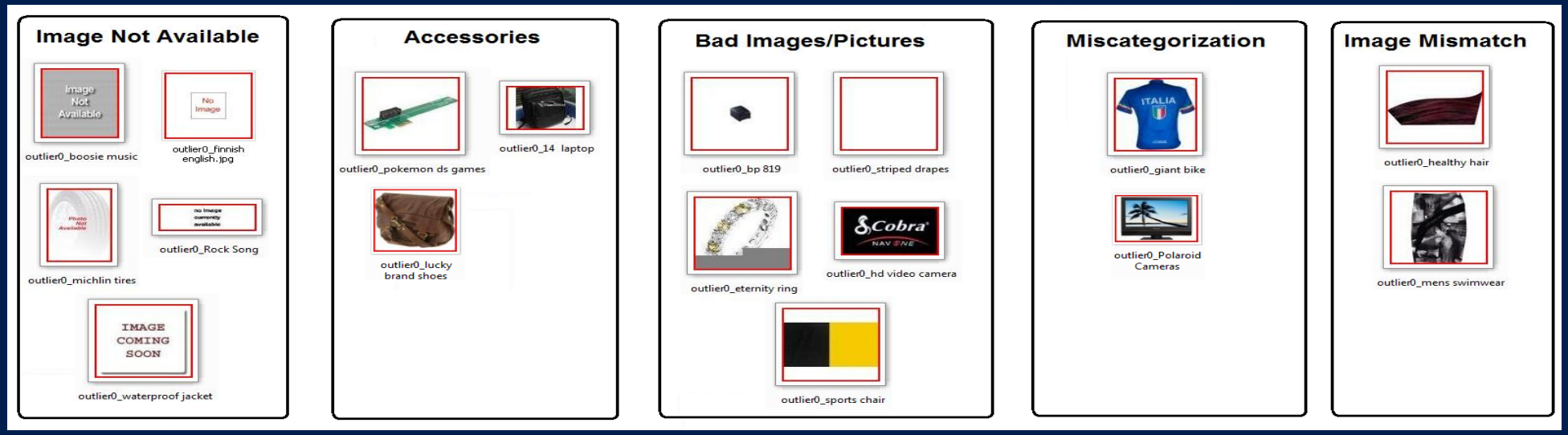
Lemma 2: **if the images are all significantly different, no outliers will be detected:**

Proof**:**

- *Std = N * Avg (N >> 0)*
- *Δ = Std/Avg ➔ Δ = N * Avg / Avg ➔ Δ = N*
- *$T = (2^{\Delta} + SEED)*Std + Avg$ ➔ $T = (2^{N} + SEED)*Avg + Avg$*
- *Assuming SEED = 0: $T = 2^{N}*Avg + Avg$ ➔ $T = Avg * (2^{N} + 1)$*
- *Outliers = {Image X | Penalty(X) >= $Avg * (2^{N} + 1)$}*
- *For N sufficiently large: Outliers = {Image X | Penalty(X) >= ∞}*
- *Outliers = {}*
- ***Therefore, no outliers will be detected!***

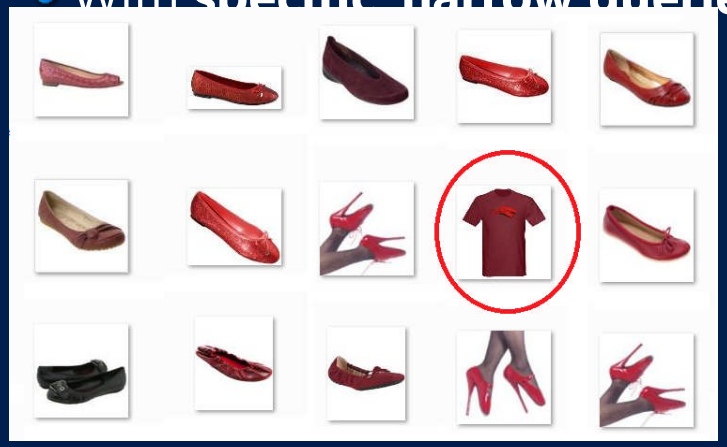
Head-To-Head Comparison

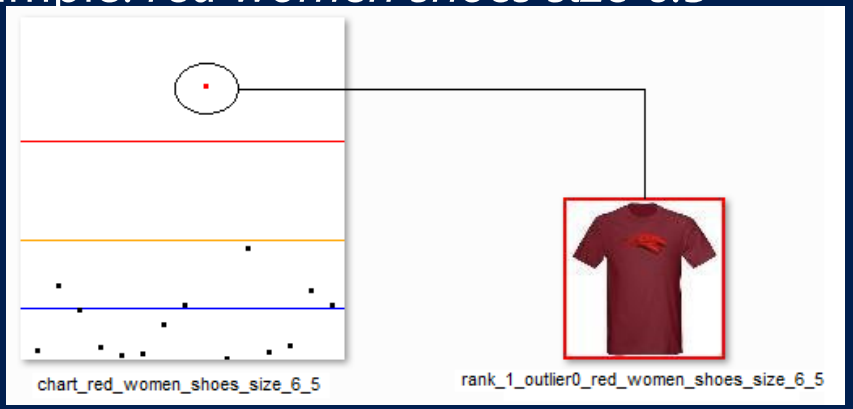
- **For top 1,000 queries (~20,000 images):**
  - Bing Shopping: 661 outliers
  - Google Shopping: 560 outliers
- **Types of issues found:**



Important aspects about the algorithm

- **Precision: ~60% in a labeled set of 600 images (20 groups of 15)**
- **Not all outliers are bugs! Actually majority might not be**
- **Works best:**
  - **Combined** with other algorithms/heuristics
  - With **specific, narrow queries**. Example: *red women shoes size 6.5*



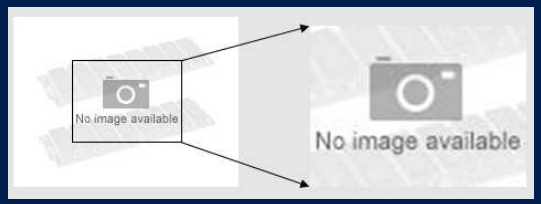


Case Study 2:
Finding "image-not-available" images

It starts with insights!

- OCR - Optical Character Recognition
  - It has limitations in terms of the size and orientation of the text

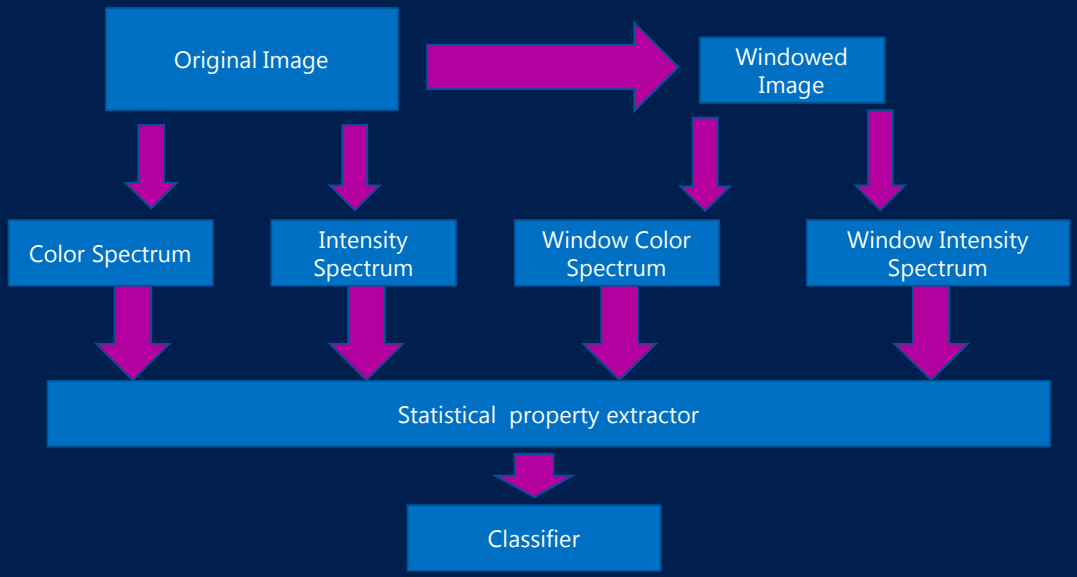
- Artificial Neural Network
  - Adaptive Classification System
  - Based on the human brain
  - Key is to determine the proper features + training data



- Two important insights:
  - *Center of the image was of key importance*
  - Color (value) and Intensity (brightness) were interesting features

Artificial Neural Network Solution

- Classifier: supervised perceptron (*Rosenblatt 1962*)
- Datasets: 2,012 NoImages, 2,000 good images
- Features:



Artificial Neural Network Solution

- Evaluation method:
  - **4-fold Cross Validation (CV)**:
    - Images were divided into 4 equal sized subsets
    - 3 of the subsets were used for training
    - The remaining one for testing/validation.
    - This procedure was repeated 4 times
    - Precision & recall obtained were averaged, which gives an aggregate result of testing all 2,012 NoImages.
- Precision: 88.8%
- Recall: 86.3%

In-practice evaluation

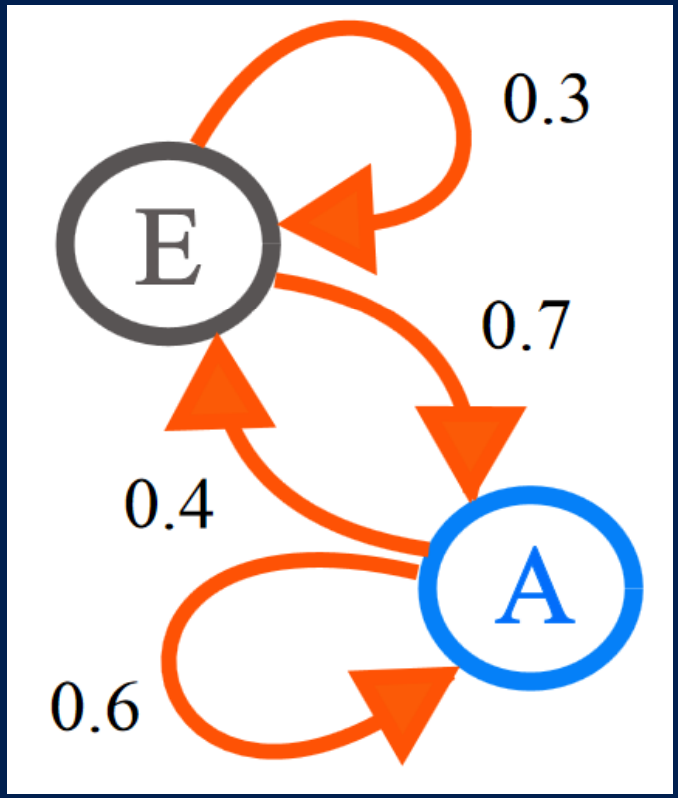
- Bing and Google Comparison

Shopping Site (~2010)	Total Image	Precision	*Estimated* Recall
Bing	138136	55.4%	76.1%
Google	74123	62.0%	94.7%

- Combine with other techniques (heuristics, text, other Machine Learning algorithms)
- ***1.5M images were blocked as a result of this project!***

Case Study 3:

Creating a model of production traffic based on Markov Chains

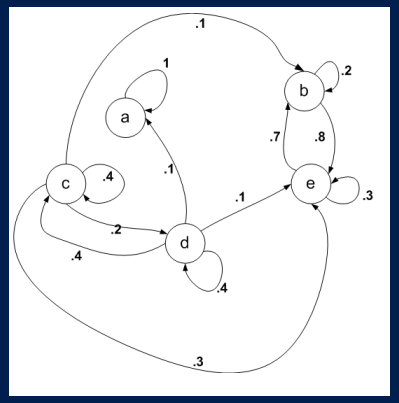


The Problem

- *"Some issues only happen in production"*
- Reproduce the same pattern of traffic observed in a live site environment in order to perform more accurate tests.
- Stress/Performance/Load tests become more accurate when following the same traffic patterns as in Production.
  - ***How can we extract such patterns?***

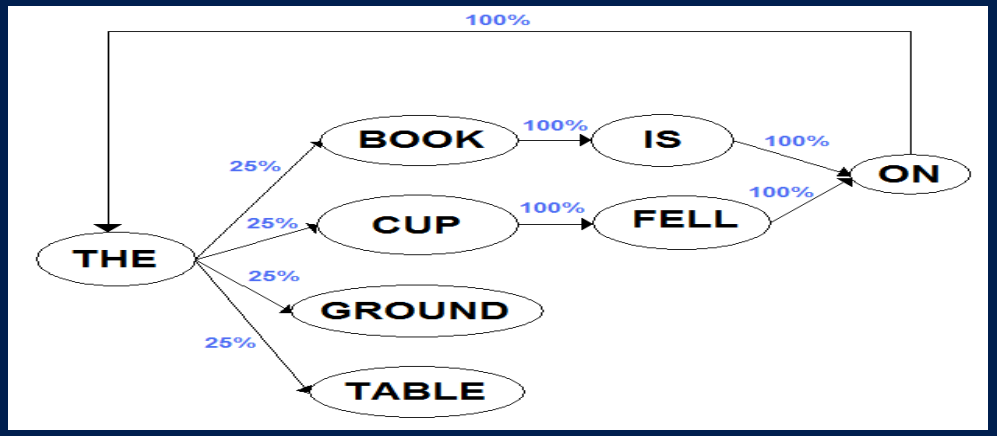
What are Markov Chains?

- Formally:
  - *"A discrete-time stochastic process, where the conditional probability of the system state in the future depends only on the current state, and not on past states"*
- Informally:
  - A **Finite State Machine** where transitions from one state to another one are based on **probabilities** and are only dependent on the **current state**



What are Markov Chains?

- Example: automatic text generation
  - Step 1: creating the Markov Chains from training data:
    - "The book is on the table"
    - "The cup fell on the ground"



- Step 2: generation of results by traversing the chains:
  - *"The book is on the ground"* (P = 25%)
  - *"The cup fell on the book is on the cup"* (P = 1.5625%)

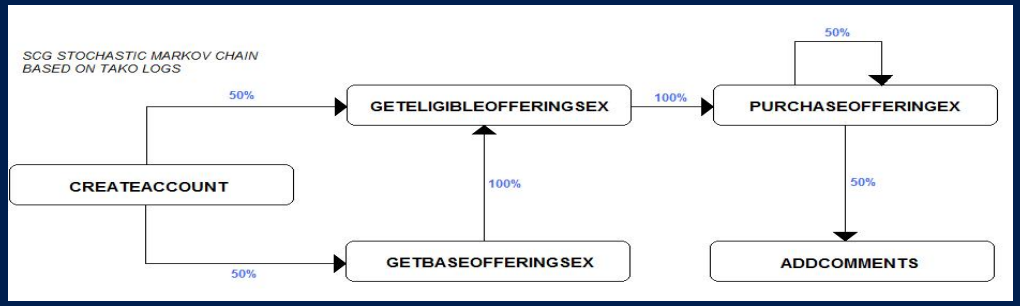
Using Markov Chains

- Example:
  - A sample log from the Microsoft Billing System:

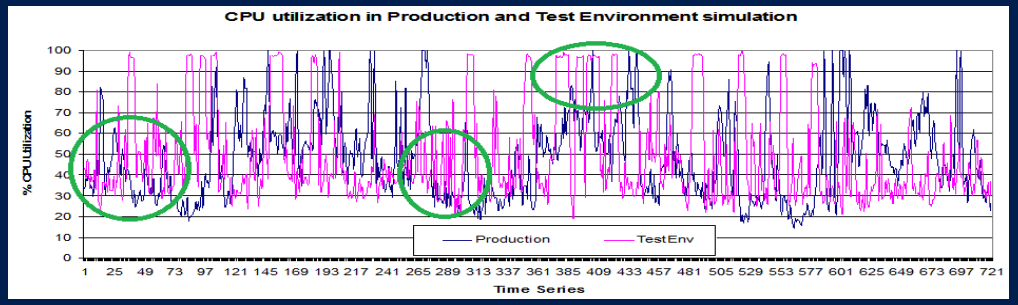
API	SessionID	TimeStamp
CreateAccount	1	12/6 10:15:31am
CreateAccount	13	12/6 10:17:01am
GetEligibleOfferingsEx	1	12/6 10:16:31am
PurchaseOfferingEx	1	12/6 10:16:36am
PurchaseOfferingEx	1	12/6 10:16:38am
GetBaseOfferingsEx	13	12/6 10:17:50am
GetEligibleOfferingsEx	13	12/6 10:17:58am
PurchaseOfferingEx	13	12/6 10:18:28am
AddComments	1	12/6 10:16:45am
AddComments	13	12/6 10:18:34am

Using Markov Chains

- Step 1: *knowledge retriever*



- Step 2: *knowledge executer*



For Free! Insight into how APIs are being used by customers!