Mobile Advertising and Marketing

Building a Big Data Analytics Service Framework

CIS 601

Presented By
Sagar Dahiwala
Agenda

1. Introduction
2. **Basic concept** about **big data analytics** for mobile advertising
3. Big data **analytics service system** and its **framework**
4. Framework related **design and implementation**
5. Case study
6. Conclusion and future work
1. Introduction

Mobile Advertising poses three distinctive opportunities or challenges to marketers and advertisers.

1. **Unprecedented adaptation**
   - Potential of mobile internet is yet to be realized

2. **Customer engagement challenges** due to huge mobile datasets
   - diverse demographics, personal preference, behavior, social presence, and location usage.

3. **Impact of mobility on digital marketing**
   - Challenges:
     - Types of mobile data, structured and unstructured,
     - privacy and context information associated with the data,
     - mining of advertising insights from the user data.
     - Mobile user’s expectation and marketers’ campaigns
2. Understanding big data analytics for mobile advertising

• Big data computing environment
  • Hadoop the first open source distributed computing environment, not useful for real-time analytics of dynamic information
  • Strom (by Twitter) – stream processing framework, continues computation
  • Spark
    • Scalable data analysis platform
    • In-memory computation; performance advantage to Hadoop’s cluster storage method
    • offers single data processing and iterative task for distributed datasets
2.1 Big data analytics techniques

- Association rule learning
- Data mining
- Cluster analysis
- Crowd sourcing
- Machine learning
- Text analytics
- Classification

- Data fusion
- Network analysis
- Optimization
- Prediction modeling
- Regression
- Special analysis
- Time series analysis
Marketing analytics and advertising recommendation

- **Intermediate** nodes represent **attribute of the data**
- **Leaf node** represents the **outcome of the data**
- **Branches** holds the **attribute value**

![Diagram of a decision tree example](Image)
Information gain

Information gain of an attribute = entropy – information of that attribute

The attribute with the highest information gain is the root node, and the next level nodes are the next high information gain attribute

Ref:
http://www.cs.cmu.edu/afs/cs.cmu.eduacademic/class/15381-s06/www/DTs.pdf
Decision Tree Algorithm (self-explainable)

Step 1: Calculate the information gain for all the attributes

Step 2: Select the root node from the attribute list that has more information gain.

Step 3: For each value of the root node

Step 4: Create a node for the attributes with next highest information gain.

Step 5: For each value of the nodes

Step 6: Create subset of training data for this node

Step 7: If all the values of class node are same, create a leaf node and stop

Step 8: Else go to step 5 and continue
2.2 K means clustering

Three types of Initial clustering

1. Dynamically chosen
   • choose first K items and assign to K cluster

2. Randomly chosen
   • randomly select the values and assign them to K clusters

3. Choosing from upper and lower boundaries
   • choose the values that are very distant from each other

Ref.
ML: Clustering, by Pooyan Fazli Assistant Professor Computer Science Cleveland State University
2.3 K-nearest neighbor

- When the k closest points are obtained, the unknown sample is then assigned to the most common class among those k-points.

- The closeness is measured using the distance between the two points.
The table defines some of the approaches to find distances between two points.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minkowsky</td>
<td>[(\sum_{i=1}^{m}</td>
</tr>
<tr>
<td>Manhattan</td>
<td>[\sum</td>
</tr>
<tr>
<td>Chebychev</td>
<td>[\max_{i=1}^m</td>
</tr>
<tr>
<td>Euclidean</td>
<td>[\left(\sum_{i=1}^{m}</td>
</tr>
<tr>
<td>Canberra</td>
<td>[\sum_{i=2}^{m} \frac{</td>
</tr>
</tbody>
</table>

The most used similarity/distance metric
- Euclidean distance followed by Manhattan.

The table defines some of the approaches to find distances between two points.
2.4 Recommendation System
3. Big data Ad recommendation services system architecture
3.1 System functions

1. Device location
2. Supporting (latitude, longitude)
3. User profile and interest
4. Ad publisher products item sets
5. Identifying relevant advertisements
6. Customer oriented requirements
3.2 Key technologies and solutions

1. Real-time analytics based on spark
2. GEO information integration with profile datasets
3. Decisions based Machine Learning (ML) Algorithms
4. Clustering
5. Similarity Analysis
6. Machine learning with high performance
3.2 Key technologies and solutions

1. Real-time analytics based on spark
2. GEO information integration with profile datasets
3. Decisions based Machine Learning (ML) Algorithms
4. Clustering
5. Similarity Analysis
6. Machine learning with high performance
Fig. 8. Personalized Recommendation for the End-user
Case study

Figure 9. Number of merchants bubble chart based zip
Figure 10. User reviews by different cities around Phoenix, AZ
Conclusion and future work

• Paper support location based ad recommendation system

• The project provides a decision based approached to handle various use case associated with pushing relevant ads towards the end users

• A lot of features can be added to the system to achieve high scalability supporting real-time processing and recommendation.

• plan to provide online modeling and training of datasets to enrich mobile user experience.
Reference

1. Building a Big Data Analytics Service Framework for mobile Advertising and Marketing

2. Information gain
   http://www.cs.cmu.edu/afs/cs.cmu.edu/academic/class/15381-s06/www/DTs.pdf

3. ML: Clustering
   by Pooyan Fazli Assistant Professor Computer Science Cleveland State University
Thank you