Big Data at AT&T Regional Case Competition

We focus on Yelp, Google Map reviews, Twitter and Facebook 4 media platforms. There are 3 steps in our team plan.

1. Collect customer reviews from different social media platform using extract data tool, and generate a source data file.

Social media platform: Yelp, Google reviews

Extra data tool: Octoparse

Process:

1. Run loop, extract data (store name, address and zip code) from branch stores
2. for each store, run loop, extract each customer review data (star rating if any, review content, time etc.)
3. generate into a CSV file

Social media platform: Twitter

Extra data tool: Twitter Developers

Process:

1. Via API get reviews which contains place ID
2. Retrieve geographic location using place ID
3. Depending on geographic location, assign reviews to branch stores

Social media platform: Facebook

Extra data tool: Facebook Developers

Process:

Use API get customer comments and ratings from each branch store’s page

1. Group these data in this CSV file into 3 clusters: onTe cluster is Positive, one is Negative, the other is Invalid. Add one more attribute called Attitude into each record, value selected in (Positive, Negative, Invalid)

Set the grouping rule:

If a review data contains star rating,

If stars > 3, Attitude = Positive

If stars < 3, Attitude = Negative

else

Set a keyword aggregation a contains “good”, “like”, “nice”, “awesome”, “satisfied”, “happy” etc.

Set a keyword aggregation b contains “bad”, “dislike”, “disguise”,” angry”, “suck” etc.

if a review contains any elements in A, etc, but not contains elements in B

Attitude = Positive

if a review contains B elements, but not A

Attitude = Negative

If a review contains both A elements and B elements,

Count the times of A elements happened (set as A) and times of B (set as B) elements happened, then compare this two number

If A > B, Attitude = Positive

If B > A, Attitude = Negative

If A = B, Attitude = Invalid

If review content is void or do not contain key words,

go to Invalid

1. Cluster records with same address as branch stores.

To evaluate the performance of a store, assign Performance Score value, which range from [-2,2], to each record.

If a record has 5-star-ratings (like Yelp, Google reviews, Facebook),

set Performance value = 2 when review has 5 stars

set Performance value = 1 when 4 stars

set Performance value = 0 when 3 stars

set Performance value = -1 when 2 stars

set Performance value = -2 when 1 stars

else (data from twitter)

if Attitude = Positive,

set Performance value = 1

if Attitude = Negative

set Performance value = -1

Then calculate average of Performance value of each store.

The performance value is a measure to evaluate a store’s performance.