Diana Amador

ATTRIBUTE SELECTION REPORT

Purpose: prioritize and speed up sentiment analysis of the iPhone and the Galaxy over the other handsets in the short list.

PROCESS

These procedures were applied to both Iphone and Samsung Galaxy matrixes.

1.- Assuming users will only compare sentiment between Iphone and Samsung Galaxy, we removed all the other brands attributes except those related to the operating systems (IOS and Google android). This step reduced the number of attributes from 59 to 29.

2.Data was converted from numeric to nominal in order to apply to Attribute Selection algorithms:

3.Visualization: it was possible to observe at this point the correlation between an attribute and the graded sentiment. Attributes with the most Class-0 instances were less correlated to high sentiment values. The opposite could also be observed. Therefore, even tough we have eliminated 30 attributes there was room for improvement.

4.Chi Squared Attribute Evaluation and InfoGain Attribute Evaluation algorithms were applied to the 29-Attribute data sets for Iphone and Samsung galaxy.

5.Outputs from Chi-Squared and InfoGain were compared and the 14 attributes that showed the lowest correlation to iphonesentiment for Iphone matrix and galaxysentiment for galaxy matrix were visualized to confirm their correlation index. These attributes were ranked low by both algorithms, and visualization confirmed the low correlation between them and the graded sentiment index.

6.The 14 attributes that ranked the lowest by either Chi-Squared and InfoGain evaluations were removed, and the reduced data sets were re-evaluated. No significant difference was found in the output. This confirmed that we could in fact reduce the numbers of attributes and preserve those strongly correlated to the graded sentiment.

7. We plan to continue the machine learning task using these 15-Attribute simplified data sets (Iphone and Galaxy) for the training section. Please find final attribute selection below.

IPHONE ATTRIBUTE SELECTION

== Attribute Selection on all input data ===

Search Method:
Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 15 iphonesentiment):
Information Gain Ranking Filter

Ranked attributes:
0.3151 1 iphone
0.1209 9 iphonedisunc
0.1111 2 samsunggalaxy
0.0949 6 iphonecamunc
0.0874 7 iphonedispos
0.0691 3 ios
0.0629 4 googleandroid
0.0536 10 iphoneperneg
0.0393 12 iphoneperunc
0.0308 8 samsungdisneg
0.0298 13 googleperpos
0.0283 14 googleperneg
0.0283 5 samsungcamneg
0.0281 11 samsungperneg

Selected attributes: 1,9,2,6,7,3,4,10,12,8,13,14,5,11 : 14

== Attribute Selection on all input data ===

Search Method:
Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 15 iphonesentiment):
Chi-squared Ranking Filter

Ranked attributes:
8111.158 1 iphone
3598.539 9 iphonedisunc
2737.756 6 iphonecamunc
2448.559 2 samsunggalaxy
2209.81 7 iphonedispos
2151.944 3 ios

1786.958 4 googleandroid
1415.525 10 iphoneperneg
1156.56 12 iphoneperunc
 891.47 13 googleperpos
 812.653 8 samsungdisneg
 774.611 14 googleperneg
 749.758 11 samsungperneg
 740.781 5 samsungcamneg

Selected attributes: 1,9,6,2,7,3,4,10,12,13,8,14,11,5 : 14

SAMSUNG GALAXY ATTRIBUTE SELECTION

=== Attribute Selection on all input data ===

Search Method:
	Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 15 galaxysentiment):
	Chi-squared Ranking Filter

Ranked attributes:
9151.081 1 iphone
3585.7989 11 iphonedisunc
3546.4066 10 iphonedisneg
3016.6599 7 iphonecamneg
2813.3931 2 samsunggalaxy
2748.3696 8 iphonecamunc
2260.3764 9 iphonedispos
2183.4723 3 ios
2025.0886 5 iphonecampos
1910.6792 4 googleandroid
1461.1692 13 iphoneperneg
1203.5415 12 iphoneperpos
1167.8427 14 iphoneperunc
1159.0254 6 samsungcampos

=== Attribute Selection on all input data ===

Search Method:
	Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 15 galaxysentiment):

Information Gain Ranking Filter

Ranked attributes:
0.3417 1 iphone
0.1206 10 iphonedisneg
0.12 11 iphonedisunc
0.119 2 samsunggalaxy
0.1026 7 iphonecamneg
0.0955 8 iphonecamunc
0.0889 9 iphonedispos
0.0766 5 iphonecampos
0.0702 3 ios
0.0654 4 googleandroid
0.0557 13 iphoneperneg
0.0491 12 iphoneperpos
0.0411 6 samsungcampos
0.0405 14 iphoneperunc

Selected attributes: 1,10,11,2,7,8,9,5,3,4,13,12,6,14 : 14