Research Paper: Identifying Purchase Intent from Social Posts

Eighth International AAAI Conference on Weblogs and Social Media

The approach is to label the tweets text as having Purchase Intention and Not having Purchase intention. They collected 30,000 tweets from Quora and Yahoo using their own web crawler. After removing duplicate tweets they reserve 15,000 tweets for training data and remaining for testing. They defined definition of Purchase Intention as object that is having action word like (buy, eat) associated with it. Then they annoted the data by hiring workers that read their tweets and label them as purchase intention and non purchase intention tweet. Each tweet was read by 5 people and final class was decided by maximum voting. They traverse each tweet and find verb which they categorize as "Purchase Action" words list and list of "Non Purchase Action" by " pointwise mutual information score ". For feature extraction of post they find verb and then find to which category this verb belong using "WordNet". They then find "Purchase Object" and "non Purchase Object" using " pointwise mutual information score ". To extract the Purchase Object words in post they collected noun phrase and then using "Free base" they categorize each noun phrase as Purchase Object and non Purchase Object. Then we find relationship between features if dependency parser, if Consumable Object have Purchase Action word then it is given value as 1 otherwise 0. Then they find "Purchase Supportive word dependency" to find knowledge about Purchase Object and it's corresponding Action word which is found using dependency parser. Since their data had imbalace class they used TF-IDF to label. They SVM classifier for training and test the data with annoted tweets and then find accuracy using ROC curve to find accuracy of work. They achieved 89% accuracy using above method.