What makes a review useful, funny or cool on Yelp.com?

Amira Ghenai, Daniel J. Lizotte

MLDAS 2015

UNIVERSITY OF
WATERLOO

uwaterloo.ca

Presenter: Amira Ghenai

Objectives

- Perform different feature extraction techniques on features related to users, businesses and reviews in Yelp site.

- Design a classification-based approach on Yelp reviews to identify whether they are helpful with a degree of confidence.

- Build review recommendations to users based on the classifier results.

UNIVERSITY OF
WATERLOO

Methods and Techniques

❖ Feature Selection:

- Information Grain (IG)
- Greedy Backward Elimination
- Recursive SVM

❖ Classification:

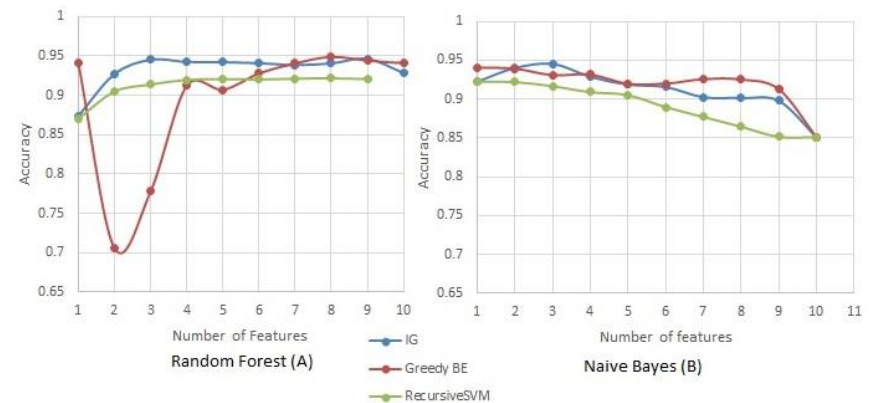- Naïve Bayes
- Random Forest

UNIVERSITY OF
WATERLOO

Results

Feature Selection

Features	IG	Recursive SVM	Greedy BE
1	R5	R1	U1
2	U1	R3	U13
3	U3	U11	U4
4	U4	B3	ST4
5	U5	U8	U6
6	U13	U4	ST5
7	U2	B5	U3
8	U8	U6	ST6
9	ST1	U3	U2

Classification

Random Forest (A)   Naïve Bayes (B)

IG   Greedy BE   RecursiveSVM

Classification by all features & best subset

	All features	Best Subset of features
Naïve Bayes	85.10%	94.54% (3 features)
Random Forest	92.83%	94.87% (8 features)

Summary

- User features and structural features proved to be most useful in terms of classification performance.

- Business features were less successful in the classification task. Such results give us an insight of what makes reviews 'useful', 'funny' or 'cool' in Yelp.com.

- Random Forest classification was more robust to the presence of noisy features, while Naïve Bayes achieved best accuracy when only considering top ranked features.

UNIVERSITY OF
WATERLOO