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

{ Amira Ghenai, Daniel J. Lizotte } University of Waterloo, Department of Computer Science

Abstract

In the era of big data and social networks, user-generated reviews are becoming essential and valuable resources for product information. In this paper, we first explore the most relevant features that make a review 'useful', 'funny' or 'cool' in Yelp site using various feature selection techniques. We, then, apply different supervised machine learning techniques and evaluate the classification accuracy of each approach. Finally, by testing the performance of the classification approach, we reach a 95% accuracy to recommend the most 'useful', 'funny' and 'cool' reviews in Yelp.

Objectives

In this work, we address the available growing number of reviews challenge and how to predict the most meaningful ones by the following:

- Perform different feature extraction techniques on features related to users, businesses and reviews in Yelp site.
- Design a classification-based approach on Yelp user's reviews to identify whether they are helpful with a degree of confidence.
- Build review recommendations to users based on the classifier results.

Techniques & Methods

The following **feature selection** techniques were used to complete the feature selection task:

- Information Gain (IG)

$$\text{IG(D, c, f)} = \text{Entropy}(D, c) - \sum_{v \in \text{values(f)}} \frac{|D_v|}{D} \text{Entropy}(D_v, c) \quad (1)$$

- Greedy Backward Elimination (BE)
- Recursive SVM

The following **classification** methods were used:

- Naïve Bayes Classifier

$$C_{nb}(E) = \arg\max_{c} p(c) \prod_{i=1}^{n} p(a_i|c) \quad (2)$$

- Random Forest Classifier

The confidence score, which is the posterior probability in Naïve Bayes is , in Random forest, the distribution of training instances classified by the rule or leaf node (class distribution).

Feature Selection

To consider the relative importance of individual features, we picked the top 9 features as follows:

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

Table 2: Feature Selection on "Shopping" reviews

Greedy BE and IG ranked user features as the most important ones. On the other hand, Recursive SVM considered features related to the review readability and business features as significant features in addition to user features. All feature selection techniques agreed that user's reviewing behaviour is a strong predictor of the review type.

Classification Results

- Greedy BE was very sensitive to the number of subset features.
- The best accuracy was achieved using 8 features for the Greedy BE technique and IG techniques.
- Naïve Bayes classifier: the best accuracy was achieved using IG using 3 features with a value of 0.9453.
- Naïve Bayes classifier gets worse as the feature subset size increases which is not surprising as we are incorporating irrelevant features that badly effects the classification task.

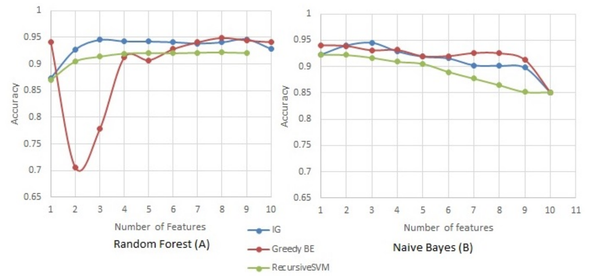

Figure 1: Classification by feature subset

Table 1: Classification by All/Subset of features

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

- The lowest accuracy is achieved when using all features with Naïve Bayes classification approach.
- When only relevant features are used, Naïve Bayes achieved a higher accuracy (94.54%).
- The best classification based approach is Random Forest with the 8^{th} Greedy BE features subset with an accuracy value of 94.87%.
- For recommendation, we consider how frequently the system manages to select 'useful', 'funny' or 'cool' reviews. Using this approach, 95.35% of the reviews were correctly labeled compared to only 35% of a random approach for labeling. Our approach achieved much more improved results compared to randomly selecting the most "useful", "cool" and "funny" reviews.

Conclusion

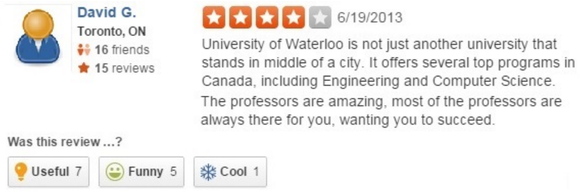

Figure 2: Yelp Review

- User features (user average helpfulness votes..) and structural features (the number of words, complex words and sentences..) 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 based approach was more robust to the presence of noisy features, while Naïve Bayes achieved best accuracy when only considering top ranked features

References

[1] Matthieu Cord and Pádraig Cunningham. *Machine learning techniques for multimedia: case studies on organization and retrieval*. Springer, 2008.

[2] Harry Zhang and Jiang Su. Naïve bayesian classifiers for ranking. In *Machine Learning: ECML 2004*, pages 501–512. Springer, 2004.

Future Research

Further analysis is required to solve the "Rare Class Classification" problem some classes in the dataset were suffering from either by doing 'Undersampling' or by 'Oversampling' or by using non-state-of-the-art classification approaches modified to deal with such dataset issues.

Additionally, another approach to classify reviews would be using the words in the review text to perform automatic feature extraction such as n-grams instead of using engineered features. Only experiments can prove which type of features is the best for this type of applications.

Contact Information

Email aghenai@uwaterloo.ca
Phone +1 (519) 722 1055
Address 200 University Ave W, Waterloo, ON N2L 3G1 Canada