Detecting General Opinions from Customer Surveys

Evgeny A. Stepanov & Giuseppe Riccardi

Signals & Interactive Systems Lab
University of Trento, Italy
{stepanov,riccardi}@disi.unitn.it

SENTIRE ICDM 2011-12-11

Outline

Signals & Interactive Systems

Online Review vs. Customer Survey

- have user ranking & free-text comment
- rank different aspects (some review web sites)

Online Review vs. Customer Survey

- have user ranking & free-text comment
- rank different aspects (some review web sites)

- initiative: survey – collected vs. review – provided
- author's intentions: deliver an opinion vs. deliver problem report, suggestion, opinion, or all together

Signals & Interactive Systems

Online Review vs. Customer Survey

- have user ranking & free-text comment
- rank different aspects (some review web sites)

- initiative: survey – collected vs. review – provided
- author's intentions: deliver an opinion vs. deliver problem report, suggestion, opinion, or all together

Italian: ATTENZIONE NON DI MIA COMPETENZA - COMPILATO PER CHIUSURA NECESSARIA

English: *WARNING NOT IN MY COMPETENCE - FILLED IN TO BE ABLE TO CLOSE*

Signals & Interactive Systems

Motivations

- **Problem**: Pairing of free-text comments with user ratings
  - User rating might not reflect the true "rating" of the text
    - additional comments
    - comments on only negative aspects
- **Effect**: Inaccurate models for classification
- **Solution**: Identify *general* opinions

Signals & Interactive Systems

Motivations

- **Problem**: Pairing of free-text comments with user ratings
  - User rating might not reflect the true "rating" of the text
    - additional comments
    - comments on only negative aspects
- **Effect**: Inaccurate models for classification
- **Solution**: Identify *general* opinions

General Opinion

A comment (or its part) that targets an entity as a whole rather than one of its parts or attributes.

Signals & Interactive Systems

Examples

General Negative Opinion (user raiting 4):
Il sistema **product$_i$** va bene per la clientela Consumer e Business di piccole dimensioni. E' del tutto inadeguato per la clientela Corporate.
The system **product$_i$** *is good for small size Consumer and Business customers. It is totally inadequate for Corporate customers.*

Performance Problem (user rating 3):
Il problema maggiore e' sicuramente la lentezza del sistema in generale nell' effettuare transazioni.
The biggest problem, certainly, is the general slowness of the system in making transactions.

Questions to Answer

- Are user provided ratings useful to identify documents bearing general opinions?
- Are user ratings effective in predicting opinion polarity of their free-text comments?

Methodology

- Supervised machine learning for general opinion detection and polarity classification
- Balanced training set vs. Real-life distribution training
    - Random under-sampling
    - 5-fold cross-validation & Test Set
- User ratings as a feature for
    - General Opinion Detection
    - Polarity Classification
- Classifiers
    - SVM
    - BoosTexter
- (Paired) Two-Tail T-Test; $\alpha = 0.05$

Signals & Interactive Systems

Telecom Italia Data Set

- Telecom Italia Customer Care Surveys (3 years)
- Services provided by telecom operators
- Language: Italian
- Questionnaire: 27 directed questions with either multiple choice or 1-10 rating answers
  - User information related to the use of the product (3)
  - Functionality (e.g. quality of visualization, etc) (11)
  - Performance (e.g. speed) (2)
  - Quality of Tech Support (7)
  - Training and documentation (3)
  - *Global Satisfaction Score* (1)
- Free-text comment
  - noisy textual data
  - ungrammatical sentences
  - user generated lexical and orthographic variations

SiS LAB Signals & Interactive Systems

User Comment Annotation

Manual Annotation Categories (15)

- 11 labels for different product aspects users were reporting problems about;
- 1 label for the suggestions;
- 2 labels for general opinions: positive and negative;
- 1 label for everything else, i.e. others;

Classification Categories (4); Automatic grouping

- Problems
- Suggestions
- General Opinions
- Other

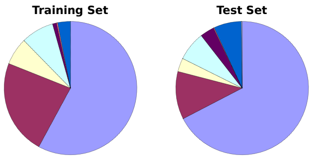
Training and Test Sets

- Test Set – chronologically the latest survey
- Training Set – the rest of data

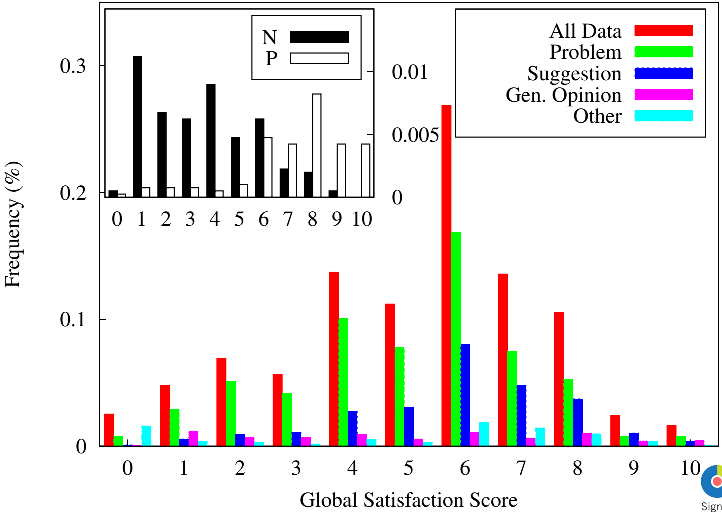
- 5-fold cross-validation on Training set

Distribution of Categories

Category	Training		Testing	
Problems	2324	57.9%	389	67.4%
Suggestions	926	23.1%	67	11.6%
Gen. Opinions	266	6.6%	19	3.3%
Other	321	8.0%	41	7.1%
Problem & Gen. Opinion	40	1.0%	20	3.5%
Suggestion & Gen. Opinion	10	0.3%	1	0.2%
Problem & Suggestion	125	3.1%	39	6.8%
Problem & Other	0	0.0%	1	0.2%
Total	4012	100.0%	577	100.0%



Global Satisfaction Score Distribution



General Opinion Detection

- 4-way classification: Problems, Suggestions, General Opinions, Others
- 2-way classification: General Opinions vs. the rest

- Balanced vs. Real distribution Training

Features

- Global Satisfaction Score
- unigrams

Real Distribution Training: Binary (Av. F1)

Classifier		CV5	Test Set
GSS	SVM	0.000	0.000
	BoosTexter	0.000	0.000
Unigram	SVM	0.055	0.000
	BoosTexter	0.288	0.224
Unigram + GSS	SVM	0.000	0.000
	BoosTexter	0.307	0.233

- GSS-only – 0 recall
- GSS effect on unigram classifier is not significant

Signals & Interactive Systems

Real Distribution Training: 4-way (Av. F1)

Classifier		CV5	Test Set
GSS	BoosTexter	0.000	0.000
Unigram	BoosTexter	0.320	0.317
Unigram + GSS	BoosTexter	0.341	0.309

- GSS-only – 0 recall
- Positive GSS effect on unigrams is not significant in CV5
- Effect is not seen on a Test Set

Signals & Interactive Systems

Real. Distribution Training: Summary

- Binary vs. 4-way 5-fold cross-validation performance difference is not significant (Two-Tail T-Test; $\alpha = 0.05$)
- GSS has no significant effect on performance

Balanced Training: Binary (Av. F1)

Classifier		CV5	Test Set
GSS	SVM	0.538	0.355
	BoosTexter	0.631	0.194
Unigram	SVM	0.671	0.212
	BoosTexter	0.691	0.168
Unigram + GSS	SVM	0.689	0.187
	BoosTexter	0.701	0.175

- Data Size: $2 * 250$
- Non-zero GSS-only performance – altered distribution

Balanced Training: 4-way (Av. F1)

Classifier		CV5	Test Set
GSS	SVM-MV	0.444	0.114
	BoosTexter	0.422	0.182
Unigram	SVM-MV	0.796	0.171
	BoosTexter	0.544	0.225
Unigram + GSS	SVM-MV	0.765	0.154
	BoosTexter	0.581	0.235

- Data Size: $4 * 250$
- SVM-MV: Majority Voting
- BoosTexter vs. SVM-MV: different GSS distributions

GO Detection: Summary of the Results

- GSS is not effective for General Opinion Detection
- High performances of classifiers trained on balanced data disappear when tested on data with natural distribution
- Training with natural distribution of labels (CV5) is predictive of the Test Set results
- Multiclass training yields the closest results to the Test Set performance

Polarity Classification

Binary classification into positive and negative general opinions

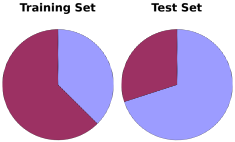
- Balanced vs. Real distribution training
- Balancing w.r.t. rating scale: 0-10 vs. 1-4 & 7-10

Features

- Global Satisfaction Score
- unigrams

General Opinion-Only Data Set

Polarity	Training		Test	
Positive	118	37.3%	28	70.0%
Negative	197	62.7%	12	30.0%
Mixed	1	0.3%	0	0.0%
Total	316	100.0%	40	100.0%



Real Distribution Training (Av. Accuracy)

Classifier		CV5	Test Set
GSS	SVM	0.835	0.805
	BoosTexter	0.813	0.745
Unigram	SVM	0.703	0.525
	BoosTexter	0.758	0.595
Unigram + GSS	SVM	0.842	0.795
	BoosTexter	0.819	0.745

- CV5: GSS > Unigram
- CV5: GSS + Unigram > GSS; not significant
- Data Size is small!

Signals & Interactive Systems

Balanced Distr. Training: 1-10 (Av. Acc.)

Classifier		CV5	Test Set
GSS	SVM	0.791	0.805
	BoosTexter	0.778	0.780
Unigram	SVM	0.644	0.565
	BoosTexter	0.713	0.590
Unigram + GSS	SVM	0.804	0.820
	BoosTexter	0.804	0.775

- Data Size: $2 * 115$

Balanced Distr. Training: 1-4 & 7-10 (Av. Acc.)

Classifier		CV5	Test Set
GSS	SVM	0.875	0.825
	BoosTexter	0.875	0.825
Unigram	SVM	0.588	0.530
	BoosTexter	0.688	0.510
Unigram + GSS	SVM	0.875	0.830
	BoosTexter	0.838	0.765

- Data Size: $2 * 80$

Polarity Classification: Summary of the Results

- GSS is very effective
- GSS-only: removing the "ambiguous" middle part has a positive effect
- Random Under-sampling has a negative effect on unigram-only and CV5
- Unigram + GSS > GSS on the balanced data with "ambiguous" middle range documents

- GSS only classifier trained on balanced data with removed "ambiguous" range can predict document polarity sufficiently well.
- Unsupervised polarity classification with thresholding is reasonable for general opinions.

Conclusions

General Opinion Detection

- User ratings are not useful for detecting generality of an opinion
- Balancing (random under-sampling) the training data has a negative effect when evaluated on unbalanced test set

Polarity Classification

- User ratings are effective for polarity classification of *general* opinions

Questions?