Credible Review Detection with Limited Information using Consistency Features

Subhabrata Mukherjee, Sourav Dutta, Gerhard Weikum

Max Planck Institute for Informatik, Germany

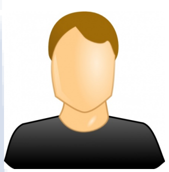
ECML-PKDD, 2016
Italy

Outline

- Motivation and Prior Work
- Consistency Analysis
- Parameter Learning
- Experiments
- Conclusions

Motivation

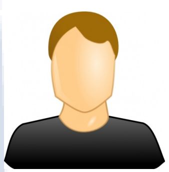
My $200 Gucci sunglasses were stolen out of my bag on the 16th. This was such a disappointment, as we liked the hotel and were having a great time in Chicago. Our room was really nice, with a great view. The hotel charged us $25 to check in early. [Rating: 3.5]



I have never been inside James. I have never checked in, and never visited the bar. Yet, it is one of my favorite hotels in Chicago. James has dog friendly-area. My dog loves it there ! [Rating: 5]

Motivation

My $200 Gucci sunglasses were stolen out of my bag on the 16th. This was such a disappointment, as we liked the hotel and were having a great time in Chicago. Our room was really nice, with a great view. The hotel charged us $25 to check in early. [Rating: 3.5]



I have never been insid ... I have never checked in, and never visited ... Yet, it is one of my favorite hotels in Chica ... es has dog friendly-area. My dog loves it the [Rating: 5]

Non Credible Review

Prior Work

- *Linguistic:* Distributional features (e.g., N-grams, sentiment etc.)
  - Issues: Performs poorly on real-world noisy data
- *Activity*: Extensive user activity history in community
  - Community features like friends, social graph, upvotes,

    Spam activity from location, IP address, device, temporal burst etc.
  - Issues:
    - Not available for "long tail" items or newcomers in community
    - Transferability due to domain dependence
    - Poor performance in domains with sparse labeled training data

However, no interpretation is provided for classification decision

Outline

- Motivation and Prior Work
- Consistency Analysis
- Parameter Learning
- Experiments
- Conclusions

Latent Facet Model

- John: "Hilton Chicago offers free wi-fi"
- Mary: "Internet is charged in a 200 dollar hotel !"

RQ: How do we spot inconsistencies between these reviews?

Latent Facet Model

- John: “Hilton Chicago offers free wi-fi”
- Mary: “Internet is charged in a 200 dollar hotel !”

RQ: How do we spot inconsistencies between these reviews?

- Objective 1: Understand “wifi” and “internet” are similar concepts
- Objective 2: Understand “free wifi” depicts positive sentiment, and “internet charged” depicts negative sentiment about similar facets

Latent Facet Model

- Assume we learn a tensor $\Phi_{k,l}(w)$ --- depicting probability of word 'w' belonging to facet 'k' with sentiment label 'l'

- We can use this to compute divergence
  - KL($\Phi_{k,l}$("free wi-fi") || $\Phi_{k,l}$("internet charged"))

as a measure of inconsistency between these facet descriptions

Prior Works: Learning Φ

- Prior work on Joint Sentiment Topic Model (Lin et al., CIKM 2009) learn Φ using a generative process based on Latent Dirichlet Allocation.

- Recent works learn more sophisticated models incorporating local dependencies (Li et al., AAAI 2010), aspects (Lu et al., ICDMW 2011), coherence (Lakkaraju et al., SDM 2013), user-preferences (Mukherjee et al., SDM 2014), and user-experience (Mukherjee et al.: ICDM 2015, KDD 2016).

- Due to the limited information constraint, we use the most basic model (Lin et al,. CIKM 2009).

Consistency Features (1/4)

Review

DO NOT BUY THIS. I used turbo tax since 2003, it never let me down until now. I can't file because Turbo Tax doesn't have software updates from the IRS "because of Hurricane Katrina". [Rating: 1]

Obj: Does this review discuss *relevant* item facets?

Consistency Features (1/4)

Review

DO NOT BUY THIS. I used turbo tax since 2003, it never let me down until now. I can't file because Turbo Tax doesn't have software updates from the IRS "because of Hurricane Katrina". [Rating: 1]

Obj: Does this review discuss *relevant* item facets?

- Learn important facet-sentiment dimensions for any item. E.g. "ease of filing" and "tax refund " for Turbo Tax are more important than "Hurricane Katrina".
- Given each review r_i on an item 'i' with words {w}, create a feature vector (dimension: K x L):

$$\Phi'_{k,l}(r_i) = f(\Phi_{k,l}(w))$$

  - Weight of the dimensions learned during training

Consistency Features (1/4)

Review

DO NOT BUY THIS. I used turbo tax since 2003, it never let me down until now. I can't file because Turbo Tax doesn't have software updates from the IRS "because of Hurricane Katrina". [Rating: 1]

Obj: Does this review discuss *relevant* item facets?

Learn important facet-sentiment dimensions for any item.

E.g. … e more

impo…

[Verdict]: Not Credible

[Interpretation]: Review focuses on irrelevant facets

Given each review r_i on an item 'i' with words {w}, create a feature vector (dimension: K x L):

$$\Phi'_{k,l}(r_i) = f(\Phi_{k,l}(w))$$

– Weight of the dimensions learned during training

Consistency Features (2/4)

Review

Internet is charged in a 300 dollar hotel! [Rating: 3]

Obj: Do majority customers conform to this opinion?

- Aggregate facet-sentiment distributions over all reviews from all users on an item to create the item description vector:

$$\Phi''_{k,l}(i) = f\,(\Phi'_{k,l}(r_i))$$

- Compute divergence between facet-sentiment distribution of review r_i on item 'i' with item description (unary feature):

$$JSD(\Phi''(i) \,||\, \Phi'(r_i))$$

Consistency Features (2/4)

Review

Internet is charged in a 300 dollar hotel! [Rating: 3]

Obj: Do majority customers conform to this opinion?

Aggregate facet-sentiment distributions over all reviews from all users on an item to create the item description vector:

[Verdict]: Not Credible

[Interpretation]: Review diverges from community description of the item's facets

Comp tribution of review r_i on item 'i' with the item description (unary feature):

$$JSD(\Phi''(i) \,||\, \Phi'(r_i))$$

Consistency Features (3/4)

Review

I have never been inside James. Never checked in. Never visited bar. Yet, one of my favorite hotels in Chicago. James has dog friendly area, my dog loves it there. [Rating: 5]

Obj: Does rating conform with the review description?

- Infer review rating from given description:

$$\Pi_l = f(\Phi'_{k,l}(r_i))$$

- Compute (absolute) deviation between user-assigned rating and inferred rating (feature vector of dimension: L)

Consistency Features (3/4)

Review

I have never been inside James. Never checked in. Never visited bar. Yet, one of my favorite hotels in Chicago. James has dog friendly area, my dog loves it there. [Rating: 5]

Obj: Does rating conform with the review description?

Infer

[Verdict]: Not Credible

[Interpretation]: Review description does not conform with rating assigned to the item

Compute (absolute) deviation between user-assigned rating and inferred rating (feature vector of dimension: L)

Consistency Features (4/4)

Yelp Spam Filter

Dan's apartment was beautiful and a great downtown location...	(3/14/2012) [Rating: 5]
I highly recommend working with Dan and NSRA...	(3/14/2012) [Rating: 5]
Dan is super friendly, demonstrating that he was confident...	(3/14/2012) [Rating: 5]
my condo listing with no activity, Dan really stepped in...	(4/18/2012) [Rating: 5]

- Burstiness of review r_i at time t_i relative to all other reviews $\{r_j\}$ at timepoints $\{t_j\}$ on an item ("unary" feature):

$$\sum_{j,j\neq i} \frac{1}{1+e^{t_i - t_j}}$$

- Additionally, capture extreme ratings (feature vector of dimension: L) as sensationalization indicative

Outline

- Motivation and Prior Work
- Consistency Analysis
- Parameter Learning
- Experiments
- Conclusions

Learning Parameters

- Classification: Incorporate consistency features in a classifier to learn weights of the (latent) dimensions
  - Train on review credibility labels (e.g. spam or not)
  - In this work, we use Support Vector Machines
  - Incorporate additional features like n-grams, limited behavioral etc. to boost performance

- Ranking: Learning to rank to find weights of consistency features
  - Train on item rankings (e.g., #sales volume of items in Amazon)
  - In this work, we use Ranking SVM

Domain Transfer

- Many domains do not have review credibility labels, or item meta-data for training classifiers
  - Train on labeled data in one domain, and transfer model to another

- Issues: (for details refer to paper)
  - Domain semantics changes for latent facet model. E.g. from Yelp (restaurants) to Amazon (consumer goods)
  - Label Imbalance

Outline

Motivation and Prior Work

Consistency Analysis

Parameter Learning

- Experiments

Conclusions

Experiments: Datasets

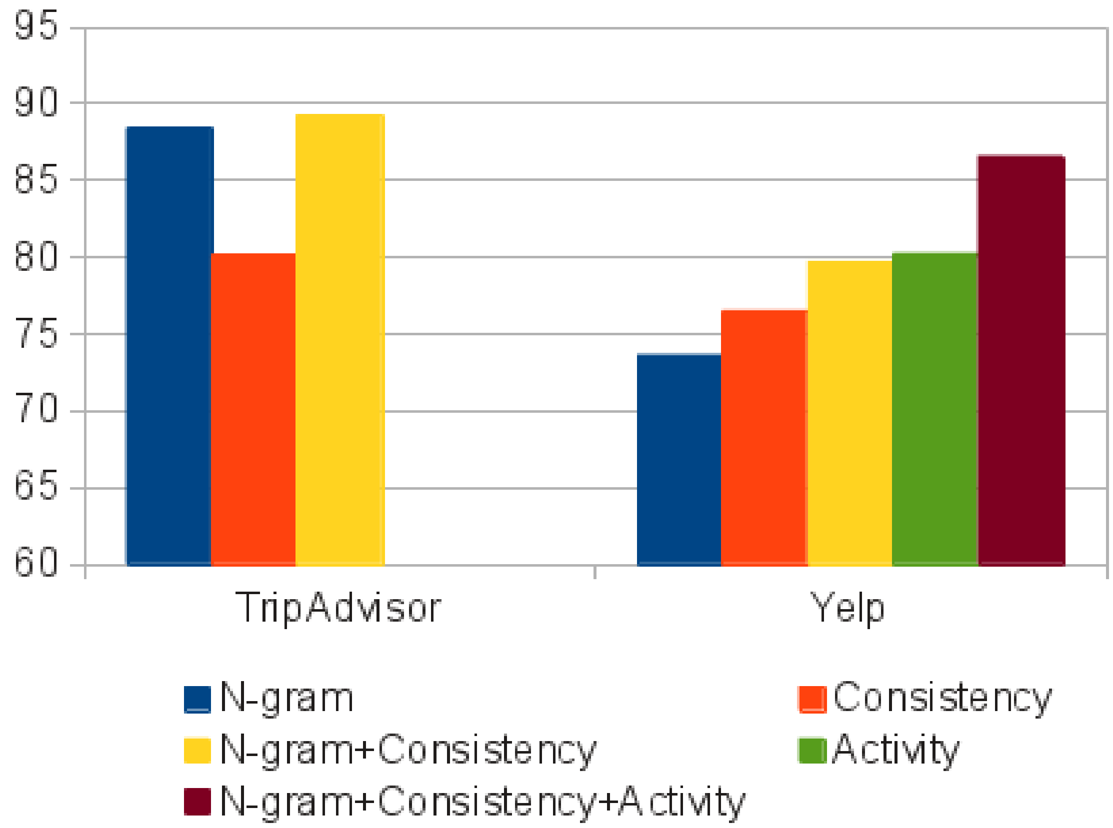
Dataset	Non-Credible Reviews	Credible Reviews	Items	Users
TripAdvisor	800	800	20	-
Yelp	5169	37,500	273	24,769
Yelp*	5169	5169	151	7898

Domain	#Users	#Reviews
Amazon		
Consumer Electronics	94,664	1,21,234
Software	21,825	26,767
Sports	656	695

Credibility Classification: Accuracy

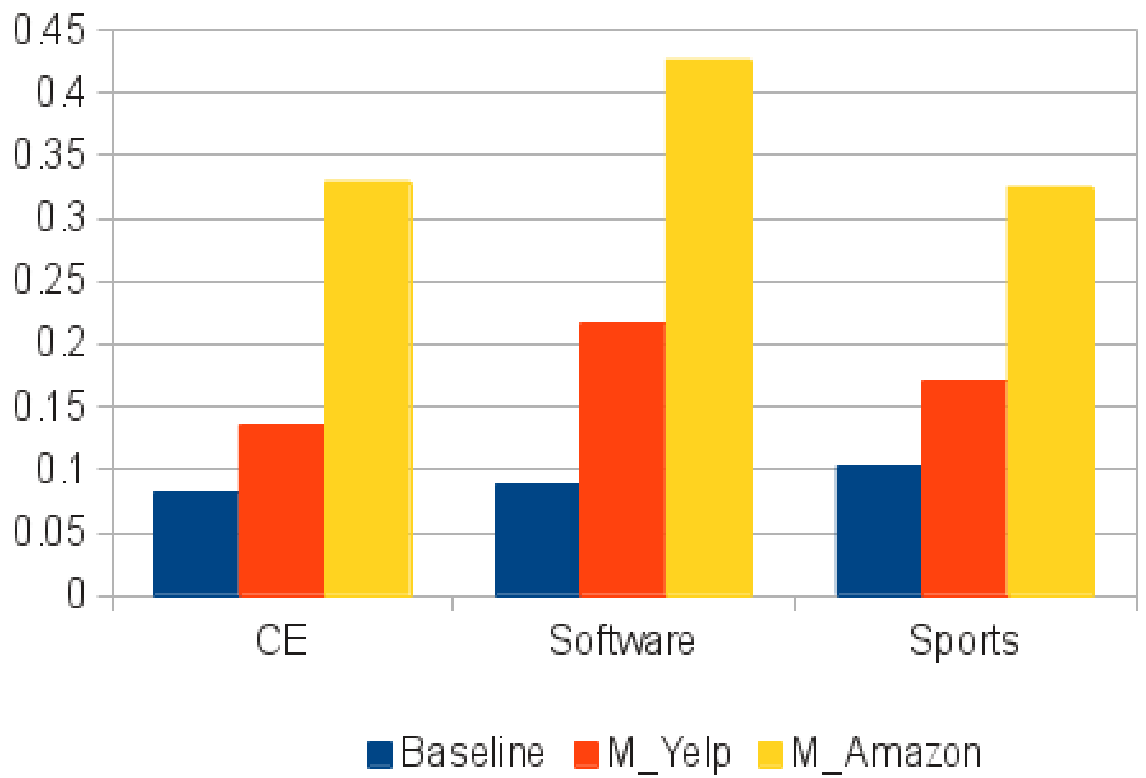
Negative Training Instances:

TripAdvisor: Amazon Mechanical Turk,   Yelp: Spam Filter



Credibility Ranking: Kendall-Tau

M_{Yelp}: Trained on Yelp and tested on Amazon with hyper-parameter tuning
M_{Amazon}: Trained and tested on Amazon using Ranking SVM
Training: Reference ranking based on #sales volume of items in Amazon



Conclusions

- We propose an interpretable model for credibility analysis with limited information:
  - Catering to "long-tail" users and items
  - Provide domain adaptation (cross-domain model transfer)
  - Avoid meta-data aggregation over time
- Provides interpretable (in)consistency evidence
  - Explain to end-user why a review should be "not recommended"

Inconsistency Features	**Yelp** Review & [Rating]	**Amazon** Review & [Rating]
user review – rating *(promotion/demotion)*:	never been inside James. never checked in. never visited bar. yet, one of my favorite hotels in Chicago. James has dog friendly area. my dog loves it there. [5]	Excellant product-alarm zone, technical support is almost non-existent because of this i will look to another product. this is unacceptible. [4]
user review – facet description *(irrelevant)*:	you will learn that they are actually EVANGELICAL CHRISTIANS worki... fa...	DO NOT BUY THIS. I used turbo tax ... never let me down un- ... because Turbo Tax ... dates from the IRS ... rina”. [1]
user review – item descrip... *(deviation fro... community)*:		...is a joke! All it ...ich is not writ- ...ve any sample of ... it appeals. [1]
extreme user rating:	... up. [5]	...era takes pictures. [1]
temporal bursts [5]:	Dan's apartment was bea... great downtown location... (3/14/2012) [5] I highly recommend working with Dan and NSRA... (3/14/2012) [5] Dan is super friendly, demonstrating that he was confident... (3/14/2012) [5] my condo listing with no activity, Dan really stepped in... (4/18/2012) [5]	

QUESTIONS ?

Credibility Classification: Accuracy

Models	Features	TripAdvisor	Yelp*
Deep Learning	Doc2Vec	69.56	64.84
	Doc2Vec + ARI + Sentiment	76.62	65.01
Activity & Rating	Activity+Rating	-	74.68
	Activity+Rating+Elite+Check-in	-	79.43
Language	Unigram + Bigram	88.37	73.63
	Consistency	80.12	76.5
Behavioral	Activity Model^{-}	-	80.24
	Activity Model^{+}	-	86.35
Aggregated	N-gram + Consistency	**89.25**	79.72
	N-gram + Activity^{-}	-	82.84
	N-gram + Activity^{+}	-	88.44
	N-gram + Consistency + Activity^{-}	-	86.58
	N-gram + Consistency + Activity^{+}	-	**91.09**
	M_{Yelp}	-	89.87