# Probabilistic Item Social Reputation Model for Recommender System

Lifan Guo
College of Information Science and Technology, Drexel University
Philadelphia, PA19104, USA
lg367@drexel.edu

Jia Huang
College of Information Science and Technology, Drexel University
Philadelphia, PA19104, USA
Jia.huang@drexel.edu

Yuan Ling, Xiaohua Hu
College of Information Science and Technology, Drexel University
Philadelphia, PA19104, USA
{yl638, xiaohua.hu}@drexel.edu

## ABSTRACT
Recommender system has been quite popular in today's commercial and entertainment business. With strong support of the recommender, a customer spends less time in searching for products that he/she desires. However, the final decision in selecting the one from available options sometimes is time-consuming?. Enhancing current recommender system by influencing customers' decision is more important to improve conversion rate in internet marking. Under such background, we proposed a new concept, namely, Item Social Reputation (ISR), which may influence customers' buying decisions with social knowledge. We present a probabilistic model to generate ISR. At last, we compared our algorithms with other influential models in sentimental analysis domain, namely, JST, AUSM, and found that our model outperformed them.

## Categories and Subject Descriptors
I.2.7 [**ARTIFICIAL INTELLIGENCE**]: Natural Language Processing – *text analysis, language parsing and understanding*; B.3.3 [**INFORMATION STORAGE AND RETRIEVAL**]: Information Storage – *Information filtering, clustering*; B.2.4

## General Terms
Algorithm, Sentimental Analysis

## Keywords
Recommender System, Algorithm

## 1. INTRODUCTION
Recommender system has been quite popular in today's commercial and entertainment business. With strong support of recommenders, a customer spends less time in searching for products that he/she desires. However, the final decision in selecting the one from available options sometimes is time-consumingConsidering an online shopping scenario, influencing customers' decisions on buying their products is even more important in internet marketing since it is directly linked to conversion rate. Marketing research has shown that consumer make decisions for several reasons [3, 13]. Knowing the factors which contribute to the buying decision is the key for internet marketing. Generally speaking, when customers buy an item in real life, they will consider the price, the appearance of the product, and others' experience of using that product. Mimicking human shopping behavior in real life, the factors in online shopping also come from two resources: a) metadata, which comes from products themselves, such as price, and weights; b) reviews, which comes from users' experience, such as like "the bag has high quality", "the bag is perfect as a gift"; a) is naturally used in online shopping, while b) cannot easily be utilized due to technical difficulties in natural language understanding. However, factor b) could be very important since: firstly they convey a general social reputation of products, like whether this bag is good or bad; secondly they contain latent information other than metadata, which could be better to match customers' preference. (E.g. a bag has a review: "I bought this bag as a Christmas gift for my husband. And I found it perfect". Then the recommendation system would label the bag as "gift", indicating that this bag is bought as gift by others. A new customer who wants to buy a bag as a gift finds a category "gift" in our system. Then, it will enhance the chance of purchase of this bag for him.)

In today's recommender system, however, the user's feedback on items is somewhat superficially processed; for example, online retailers have used reviews in different way: Many sites, such as Amazon, represent users' sentiments over star ratings. But this approach obviously lacks the factors why the products are given that rating. Furthermore, there is no further explanation of why one aspect was rated high or low. [14]. Besides, other retailers select sentences from high rated reviews as recommendation reasons or let others vote reviews that they think helpful. But it is still impossible for new customers to obtain a whole picture of what reasons people votes. Last but not least important, reviews may contain different sentiments about aspect. To be selected as buying reasons to new customers, aspects need be paired with sentimental value. It is reasonable that the system would recommend positive aspects as reasons to new customers to persuade they make decision. In other words, aspects need to be linked to sentiments.

It would be very helpful if an item has a well-established reputation which can be used by the new customer as a reference. It is under such motivation that we invent the concept of Item Social Reputation (ISR), which is representing top K salient positive aspects extracted from customers' reviews on a specific item. (It is reasonable to list salient positive aspects other than negative to customers for marking considerations. )In order to make sure the fairness, the reviews are collected from all over the web instead of from a single store or website. Each aspect of ISR contains a list of terms that has close semantic meanings to that aspect. ISR enhanced recommender system is able to analyze the reasons driving previous customers to buy an item from online reviews repository. ISR was extracted to help provide a better match of user's preference. Furthermore, ISR could be visualized as features adding on final recommendation results, providing facilities for customers to find their preference. Therefore, it achieves a much better performance in supporting the customer to achieve their goal in terms of improving conversation rate.

The rest of the paper is organized as follows. First, section 2 describes the related work. Next, section 3 introduces the probabilistic ISR model. In section4, we describe the experimental evaluation. Finally, the experimental results are presented in section 5.

## 2. RELATED WORK

The problem at hand is how to automatically discovering top K salient reasons (e.g. Capacity) with positive interpretive supports (e.g. it has enough spaces for credits cards). Those reasons present ISR aforementioned in our framework.

To the best of our knowledge, few works have addressed this type of problems. The most similar current research is focused on reviews mining, trying to extract aspects and sentiments together from reviews. There are two research lines being defined: First, frequently occurring noun phrased (NP), which presents aspects, are discovered for the purpose of using reviews [2, 4, 5]. However NP detection method depends on pre-defined rules in the system, thus they lack generality in cross domains. it is also very time consuming. Second, topic modeling, as a suitable method and a graphical model, has drawn a lot of attention for this task. Specifically, Latent Dirichlet Allocation (LDA) model is recently extended to address this problem. LDA model can be treated as a way to decompose high dimension matrix, with a semantic explanation of results. However, applying LDA model directly to this problem is not desirable since it assumes that 1) documents are independent 2) words are independently and identically distributed (i.i.d). [14] 3) it needs extension to integrate sentiments information with to aspects. Therefore, the challenges are two folds. First, 1) how can we extend LDA model to integrate sentiments information 2) how can we break the i.i.d. assumption.

Similar works like Topic Sentiment Mixture (TSM) model [12], Multi-Aspect Sentiment (MAS) model [16], Joint Sentiment/Topic (JST) model [10], and Aspect and Sentiment Unification Model (ASUM) have extended LDA model to extract aspects from reviews. But TSM and MAS separate aspects and sentiments, without explanation of how much a word is related to certain aspects and sentiments. These two models failed to link aspects and sentiments together, thus they are not suitable for our task. On the other hand, JST model and ASUM model indeed integrated aspects and sentiments into a single language model. In other words, they fixed the first challenge of extending LDA model. However, they ignore the second challenge, which needs to be addressed in practice.

From the observation of data, we found that such challenge could be addressed by considering words in syntactic level. For example, “I wish the canvas was heavier so it will hold up longer, but it was a great deal so if I only get one year out of it that is fine.” Obviously, the first half of this sentence expresses the aspect related to the material, while the second half is related to price. Moreover, “but” is a very important term since it plays a role in representing aspect shifting from material to price and sentiment shifting from negative to positive. Therefore, we believe that words need to be viewed together until we meet transition words or phrases, like “but”, or punctuation characters, like period. Besides, from this sentence, we found that some words contribute much meaning of the text, like “hold”. Therefore, terms weighting have to be incorporated since terms are not independently distributed in a topic [7]. Thirdly, the quality and trustworthiness of the extracted terms and the sentiments matter. As a matter of fact, such information is shown in the votings of reviews. Unfortunately, previous methods ignore such information and often produce Low quality topics for users Topics that mix unrelated or loosely related concepts substantially reduce users’ confidence in the utility of automated systems. [10]

Our work falls into the same area of using probabilistic topic modeling techniques for aspects and sentiments identification to address problems aforementioned. Aspect and Sentiment Aggregation Model with Term Weighting Schemes (ASAMTWS) is innovatively proposed with incorporating pre-existing knowledge in the form of automatically extracted constrains to achieve better modeling results. Our basic hypothesis is: 1) instead of treating terms equally, terms are weighting by their frequencies and sentiment, which provides better semantic meaning of topics; 2) instead of splitting a document into sentences, we use split documents into smaller chunks, which are finer to convey coherent meanings. The reason why we choose chunk instead of sentence and word as a unit is that one sentence is likely to convey more aspects while word level processing mix all the topics together. 3) instead of treating reviews equally, the ASAMTWS model considers the importance of reviews by extra voting information. Quality is especially important when topic modeling methods are applied to online review domain. In practice, however, we found that it is not desirable and feasible to present all aspects extracted from ASAMTWS model to users since 1) some potential buying reasons dominate the reviews of a specific item from dataset; 2) unsupervised model produce some poor quality topics, which should be detected and deleted automatically. Therefore, it is necessary to handle these problems. Unfortunately, we found few work focusing on this issue. The closest research is multi document summarization reference. However, we have a different target problem. Inspired by some work [4.6], we innovatively present a model: Diversity in Ranking High Quality Aspect (DRHQA) model that help to generate K top high quality salient reasons from the results of ASAMTWS model aforementioned.

## 3. PROBABLISTIC ITEM SOCIAL REPUTATION MODEL

In this section, we introduce probabilistic item social reputation model. This model adapts ASAMTWS model, to generate aspect and sentimental matrix. Then we use another innovative model DRHQA model to find top salient k positive aspect from the matrix aforementioned. We use bag as the domain in our example.

### 3.1 GENERAL FORMALIZATION

Let $p = \{p_1, p, \dots, p_m\}$ be a set of products which comes from “bag” domain. For each product $p_i$ there is a set of reviews $r = \{r_1, r_2, \dots, r_d\}$. For each review $r_i$ there are a set of chunks $c = \{c_1, c_2, \dots, c_l\}$ , and a non-negative value of the other’s voting information o. For each pair of two consecutive chunks, it has constrains including three possible conditions {must-link, cannot-link, no-link}. For each chunk $c_i$, there is a set of words $w = \{w_1, w_2, \dots w_n\}$. For the domain of bag, our job is to identify $K * S$ latent aspect and corresponding sentiment groups from $M * D$ reviews. Each group is presented by $N$ words ranking by how likely they will appear in it. Then, each product has a vector v with the length of $K * S$. This vector presents that how likely that product has the aspect and corresponding sentiment. Based on a vector $v$, top $K$ aspects are selected as reasons to new customers based on three criteria: 1) Top $K$ aspects have positive sentimental value; 2) Words ranked in each aspect have optimal semantic coherence. 3) Top $K$ aspects balance centrality and diversity.

### 3.2 CRAWLING & PRE-PROCESSING

In the pre-processing step: POS tools [15] are employed to generate chunks and constrains from pre-existing knowledge. The input are reviews on the website, the output are chunks and constrains. Crawling: reviews are unstructured data across the websites, we use crawler to crawl semi-structure reviews from public websites. Chunks and Constrains:

Step1: We use POS tools to do sentence splitting. [15]

Step2 Given a sentence: if it does not contain any transition words or phrases defined in [17], the sentence is considered as a chunk. Else, goes to step 3.

Step3. The whole sentence will be split by the transition words or phrases into two chunks, if any chunk has the transition words, goes to step 2.

Repeat step2-step3 until all chunks do not have any transition words or phrases, then, goes to step 4.

Step 4: Must-link, cannot link, and no-link: If there is the transition word or phrase between two consecutive chunks, either must-link or cannot link would be added: If the transition word or phrase belong to opposition/limitation/contradiction category, then a can-not link is built, else, a must-link is built. If there is no must-link or cannot link that can be built, then there is no-link between these two chunks.

Each word will be assigned a POS value by using [15].

Online reviews, after pre-processing, are treated as input to ASAMTWS model.

## 3.3 GENERATING POSITIVE ASPECTS FROM ASAMTWS MODEL

ASAMTWS illustrates the generative process of a review as follows: The reviewer decides to write review of a bag with a distribution of sentiments, say, 60% satisfied and 40% unsatisfied. Then, he decides to express the rational by expresses a distribution of aspects, say, 20% service, 60% color, 20% quality. Then he decides to write reviews to express which he feels that sentiments. Then if the review is useful to the others, it will get positive voting.

The generation process is presented as follows:

1) For every pair of sentiment s and aspect z, draw a word distribution $\emptyset_{ts}$ ~ Drichilet ($\beta_s$)

For each review r,

(a) Draw r's sentiment distribution $\pi_r$ ~ Drichilet ($\gamma$).
(b) For each sentiment s, draw an aspect distribution$\theta_{rs}$~Dirichilet ($\alpha$) based on the sentiments dictionary.
(c) For each chunk
   - Choose a sentiment j ~ Multinomial ($\pi_r$) based on other chunks that has constrains.
   - Given sentiment j, choose an aspect k ~ Multinomial ($\theta_{rs}$) based on other chunks that has constrains.
   - Generate words w ~ Multinomial ($\emptyset_{ts}$) based on word frequency in the dataset and reviews' voting o.

The graphical representation of ASAMTWS is shown in figure 3 and the notation used in ASAMTWS is in table 1.

**Table 1. Meanings of the notation**

| | | | |
|---|---|---|---|
| $R$ | the number of reviews | $\pi$ | multinomial distribution over sentiments |
| $C$ | the number of chunks | $\theta$ | multinomial distribution over aspects |
| $N$ | The number of words | $s$ | sentiments |
| $V$ | The vocabulary | $z$ | aspect |
| $S$ | The number of sentiments | $\emptyset$ | Multinomial distribution over aspects |
| $T$ | The number of aspects | $w$ | word |
| $\gamma$ | Dirichlet prior of $\pi$ | $s_{-i}$ | Sentiments for all chunks except i |
| $\alpha$ | Dirichlet prior of $\theta$ | $z_{-i}$ | Aspects for all chunks except i |
| $\beta$ | Dirichlet prior of $\emptyset$ | $\boldsymbol{w}$ | All words in reviews |
| $C_{rj}^{RS}$ | The number of chunks that assigned sentiment j in review r | Dictionary | Sentiment Dictionary[14] |
| $C_{djk}^{RST}$ | The number of chunks that assigned sentiment j and aspect k | Constrains | Must-link and cannot –link from [16] |
| $M_{jkw}^{STW}$ | The weighting of words that assigned sentiment j and aspect k | $m_{lw}$ | Number of words in chunk l |

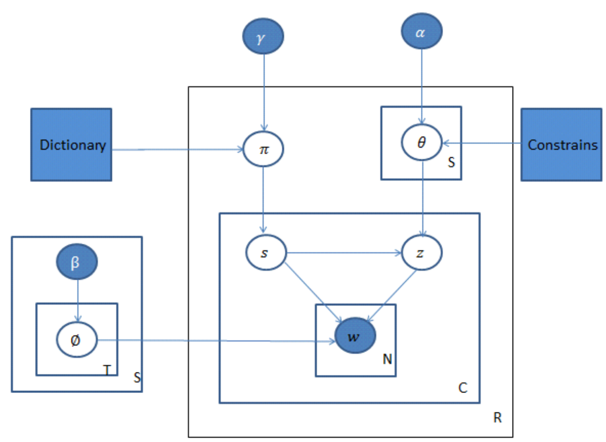


**Figure 1. ASAMTWS Model (Nodes are random variables, edges are dependencies, and plates are replications. Only shaded nodes are observable. )**

## 3.4 INFERENCE PROCESS

The latent variables in Fig.1 are inferred by Gibbs Sampling in [7]. At each transition step of the Markov chain, the sentiment and aspect of the lth chunk are chosen according to the conditional probability:

$$P(s_i = j, z_i = k | s_{-i}, z_{-i}, w) q(s_i = j) q(z_i k)$$

$$\frac{C_{rj}^{RS}+\gamma_j}{C_{rj(.)}^{RS}+\gamma_{j(.)}} \frac{C_{djk}^{RST}+\alpha_k}{C_{djk(.)}^{RST}+\alpha_{k(.)}} \frac{\Gamma(\sum_{w=1}^{W} M_{jkw}^{STW}+\beta_{jw})}{\Gamma\left(\sum_{w=1}^{W}(M_{jkw}^{STW}+\beta_{jw})+m_l\right)} \prod_{w=1}^{W} \frac{\Gamma(M_{jkw}^{STW}+\beta_{jw}+m_{lw})}{\Gamma\left(M_{jkw}^{STW}+\beta_{jw}\right)} \quad (1)$$

The approximate probability of sentiment j in review r is:

$$\pi_{rj} = \frac{C_{rj}^{RS}+\gamma_j}{C_{rj(.)}^{RS}+\gamma_{j(.)}} \quad (2)$$

The approximate probability of aspect k for sentiment j in review d is:

$$\theta_{rjk} = \frac{C_{djk}^{RST}+\alpha_k}{C_{djk(.)}^{DST}+\alpha_{k(.)}} \quad (3)$$

The approximate probability of word w is senti-aspect k-j is

$$\phi_{jkw} = \frac{M_{jkw}^{STW}+\beta_{jw}}{\sum_{w=1}^{W}(M_{jkw}^{STW}+\beta_{jw})} \quad (4)$$

The Eq. 1 is an intuitive result: The middle two terms, $\frac{C_{rj}^{RS}+\gamma_j}{C_{rj(.)}^{RS}+\gamma_{j(.)}}\frac{C_{djk}^{RST}+\alpha_k}{C_{djk(.)}^{DST}+\alpha_{k(.)}}$, indicate the importance of chunk in sentiment j and aspect k. The last two terms indicate the importance of sentiment j and aspect k in review d. $q(s_i = j)$ and $q(z_i = k)$play the role of intervention from pre-existing knowledge of constrains. $M_{jkw}^{STW}$ plays the role of weighting terms based on frequency and reviews' quality. More specifically:

$q(z_i = k)$: Chunk's topic depends on constrains. To calculate the probability of the lth chunk topic, for candidate topick, if the must-link chunk has high probability in k, we use $q(z_i = k)$: to enhance words probability in k in lth chunk. If the cannot link chunk has high probability in k, we use $q(z_i = k)$ to decrease words probability in k in lth chunk. If there is no chunk has links to the current chunk, $q(z_i = k) = 1$.

Formally, if there are must-link chunks:

$$q(z_i = k) = \text{Normalized} \sum_{\text{must-link chunk's topic}} \frac{(1 + \text{Max}(q(z_n = k)))}{\text{The number of chunks must} - \text{linked to lth chunk}}$$

(5)

If there are cannot chunks:

$$q(z_i = k) = \text{Normalized} \sum_{\text{must-link chunk's topic}} \frac{\left(1 - \text{Max}\left(q(z_n = k)\right)\right)}{\text{The number of chunks must} - \text{linked to lth chunk}}$$

(6)

Else:

$$q(z_i = k) = 1 \quad (7)$$

$q(s_i = j)$: Chunk's sentiment depends on two facets: 1) sentiment lexicon would assign the opinioned words with sentiment value as a prior knowledge$p(w_i)$, the sentiment distribution; 2) current chunk's must-link and cannot link chunks' sentiment would have impact on the current chunk. It can be described as follows: (Opinioned words are J.J. words defined in POS)

$$q(s_i = j) = \frac{p(w_i+\varepsilon)q\,(s_j=k)}{\text{Normalization Value}} \quad (8)$$

ε indicates a dump value that controls the influence of dictionary. $q\left(s_j = k\right)$ indicates the impact from linked chunks' sentiments, which has similar formula with $q(z_i = k)$.

$M_{jkw}^{STW}$: weighting terms based on frequency and reviews' quality. Formally,

$$M_{jkw}^{STW} = -\log_2 \frac{\text{Number of the word in review j}}{\text{Total number of the word in the dataset}} \left(\frac{\text{postive voting in review j}}{\text{Total voting of the review j}} + 1\right) C_{jkw}^{STW} \quad (9)$$

$C_{jkw}^{STW}$ indicates the number of words that are assigned sentiment j and aspect k. The first item is similar to PMI, which has solid basis in information theory and has shown to work well in LSI . It is possible this value of a term is negative, like background words (e.g. 'bag', 'purse'). When this happens we assign the weighting of that term to 0. The second item is used to leverage the importance of the reviews. More positive voting a review draw, more weightings added to the words in the review.

It is obvious that those constrains can be reduced in some way. These constrains enhanced the popular extensions of other topic modeling methods, with the ability to consider different scenario and be easily extended into different context. The reasons why these constrains could help original topic modeling can be summarized as: 1) ASAMTWS model changes original unsupervised topic modeling to semi-supervised; 2) ASAMTWS model explores and utilizes shallow semantic meaning of documents to break original topic modeling which has i.i.d. assumption for both terms and document; 3) ASAMTWS model innovatively incorporate social information, voting information of reviews, into tackling aspects and sentiments identification problem.

## 3.5 SELECT K ASPECTS by QRHQA MODEL

Recall that our mission is to discover top K positive salient buying reasons. From ASAMTWS model, however, we get a term-aspect matrix; with the length of row is the size of vocabulary of words in the dataset and the length of column is k aspects in the dataset. It is a challenge problem to select k. If k is selected too small, then topics would mix together; otherwise, then it takes more efforts for human to find which topics have higher quality in terms of coherent and consistent. In our problem, we want to find K high quality aspects that have positive sentiments as reasons from $k * s$ aspects.

Inspired by GRASSHOPPER [18], we design QRHQA model for our tasks, the input of which is matrix postive aspect × terms with probability of that aspect For the term-aspect matrix generated from ASAMTWS model, we consider using eigenvector centrality in the matrix could represent the centrality we want. But eigenvector does not consider diversity upfront and top ranked topic is going to dominate. Thus, each ranked topic is be transformed into absorbing state as the ranking process continues. Effectively, the importance of similar unranked nodes in the graph would be lowered by the absorbing nodes. Therefore, each ranked node is novel and thus it encourages diversity of the extracted set.

Similar to GRASSHOPPER algorithm [18], the major difference of QRHQA Model is:

1) We calculate aspect similarity by using eq.10 after pre-processing (Chose positive aspect and chose top words in each aspect). Since each column represents the word distribution of a certain aspect, KL-divergence is better to estimate the similarity of two aspects:

$$\text{sim}\left(V_i^{(t_a)}, V_j^{(t_b)}\right) = 10^{\text{IR}\left(V_i^{(t_a)}, V_j^{(t_b)}\right)} \quad (10)$$

$$\text{IR}\left(V_i^{(t_a)}, V_j^{(t_b)}\right) = \text{KL}(V_i^{(t_a)} || \frac{V_i^{(t_a)}+V_j^{(t_b)}}{2}) + \text{KL}(V_j^{(t_a)} || \frac{V_j^{(t_b)}+V_i^{(t_a)}}{2}) \quad (11)$$

Where,

$$\text{KL}(p||q) = \sum_i p_i \log\frac{p_i}{q_i} \quad (12)$$

The reason why we used this instead of KL divergence is that there is no problem of infinite value since $p_i$ and $q_i \neq 0$. And S $\left(V_i^{(t_a)}, V_j^{(t_b)}\right)$ is symmetric.

2) We consider the quality of aspects to define better ones as reasons. The quality of aspect is calculated as follows:

We use the same method in [10], which has been approved to be close to human judge, to define topic quality as

$$C(t; V^{(t)}) = \sum_{m=2}^{M} \sum_{f=1}^{m-1} \log \frac{D\left(v_m^{(t)}, v_f^{(t)}\right)+1}{D(v_f^t)} \quad (13)$$

Where $D(v)$ denotes the review frequency of word type v (i.e., the number of reviews with least one token of type v) and $D(v, v')$ be co-review frequency of word types v and $v'$. $V^{(t)} = (v_1^t, \dots, v_M^t)$ is a list of the M most probable words in topic t.

3) After a qualified aspect is selected, we decrease the quality of aspects similar to selected ones by:

$$C\left(t; V^{(t_j)}\right) = (1 - \omega S\left(V_i^{(t_j)}, V_j^{(t_j)}\right))C\left(t; V^{(t_j)}\right) \quad (14)$$

The QRHQA model is defined as follows:

Input:

Matrix W (postive aspect terms with probability of that aspect) , aspect quality matrix q, dumping value ב,ω, quality threshold ρ

1, Create the initial Markov chain P from W, q, ב.

2, Repeat compute $P'$s stationary distribution $\pi$, and pick the first item $g_1 = \operatorname{argmax} \pi_i$. If $C(g_1) > \rho$ break and add $g_1$ to results;

3, Repeat until no more items are needed to be ranked:

(a) Turn ranked items into absorbing states

(b) Update all aspect's quality based on Eq.

(c) Compute the expect number of visits v for all remaining items. Pick the next item $g_{next} = \operatorname{argmax} v$.

Calculate $C(g_{next})$. If $C(g_{next})$> ρ , add $g_{next}$ to results.

# 4. EXPERIMENTS & EVALUATION

4.1 Datasets and Experimental Setting.

We did experiments on two datasets. Dataset a) is bag reviews crawled from amazon. This dataset contains 30000 reviews. Each review has title, star, time, votes and contents. We consider reviews, which gave 4 stars or 5 stars to a bag, as positive; reviews, which gave 3 stars as neutral; reviews, which gave 1 star or 2 stars as negative. We used ISR model to select top K positive salient aspects from reviews. We also did experiments on public dataset b), which is restaurant review collected by [8.]. This review dataset is manfully annotated and covers 150 restaurants with about 3400 sentences, which are parsed on the aspect level. Each sentence is assigned one or more aspects and sentiment label. The six aspects labels are food, staff, price, ambience, anecdote and miscellaneous, and the three sentiment labels are positive, negative. For this dataset, we set topic number limit as 6 since we already know aspects number. We set α, β, ϒ as 0.1 T as 100, in the ASAMTWS model in the experiment. Also we set ρ as 0.5 in the QRHQA model for both datasets.

4.2 Evaluation Method

For dataset a), we automatically generate top K salient positive aspects. For dataset 2), we compare each of the six general aspects given by [17] on all of the two kinds of sentiments. We also use precision as the measure of our evaluation. The baseline models include ASUM, JST. The number of aspects needs to be pre-specified to be consistent with the gold standard for those models. Since all these baselines are based on LDA. And we test how well different algorithm can predict sentiments of reviews. We set the number to be 6 according to the gold standard. ASUM detects only two aspects by default. We also tune the parameter for the sentiment number of JST to 2 too.

4.3 Experimental Results

4.3.1 Dataset 1)

ISR model generated 12 different high quality aspects from 30000 reviews. For recommender purpose, we only need to present positive aspect as ISR to customers. Some of them are presented in the table.

**Table 2 ISR generated from dataset 1)**

| | |
|---|---|
| aspect1(compartment) | Strap, handle, shoulder, pocket, compliments |
| aspect2(size) | Insider, roomy, storage, insider, big |
| aspect3(style) | Clutch, handbag, hobos, backpack, Cross body |
| aspect4(love) | Sweet, great, cheap, wonderful, great |
| aspect5(gift) | Christmas, present, birthday, mother, husband |
| aspect6(brand) | Lauren, Rachel, jeansport, hobo |
| aspect7(quality) | Sturdy, material, durable, functional, solid |
| aspect8(convenience) | Pocket, zippered, protect, carried |
| aspect9(color) | Blue ,red ,brown, purple |
| aspect10(texture) | Leather, soft ,perfect ,feels, touch |
| aspect11(service) | Shipping, warranty, delivery, arrived, fast |
| aspect12(price) | Supervised, beat, money, pleased, deal |

We manually pick up top terms from each aspect we detected.From the table2, we found that with the consideration of topic quality and terms frequency, ISR model can generate high quality topics with self-explanation. Besides, it overcomes another drawback of unsupervised topic models by automatically deciding the number of topics.

Dataset 2)

**Table 3 Comparison with ASUM and JST**

| aspect | sentiment | ISR model | ASUM | JST |
|---|---|---|---|---|
| Food | positive | **0.77** | 0.61 | 0.447 |
| | negative | **0.44** | 0.39 | 0.23 |
| Staff | positive | **0.63** | 0.46 | 0.2 |
| | negative | **0.46** | 0.33 | 0.171 |
| Price | positive | **0.71** | 0.28 | 0.22 |
| | negative | **0.213** | 0.144 | 0.21 |

| | | | | |
|---|---|---|---|---|
| Ambience | positive | **0.44** | 0.41 | 0.3 |
| | negative | **0.21** | 0.11 | 0.13 |
| Anecdote | positive | **0.35** | 0.23 | 0.19 |
| | negative | **0.144** | 0.11 | 0.14 |
| Miscellaneous | positive | **0.45** | 0.36 | 0.23 |
| | negative | **0.68** | 0.49 | 0.25 |

As you can see, our model outperforms the others, demonstrating our models' efficiency to label sentiments on reviews. It improves results since 1) with consideration syntactic level of short sentences; we can better understand reviews' sentiments; 2) with consideration of term frequency and topic quality, topic modeling could automatically rank meaningful terms in each aspect, which gives a better way to express sentiments.

## 5. CONCLUSION AND FUTURE WORK

This paper proposed a new concept, namely Item Social Reputation, and the hybrid model to generate ISR. Our major contributions are three aspects: 1) to our best knowledge, it is the first work to consider syntax in sentence level in topical modeling. 2) After incorporated syntactic information, term's weights and existing dictionary, our model managed to achieve a better result in sentimental analysis in our experiments, compared with other the state of the art methods. 3) More importantly, this is the first work to create a new concept, namely Item Social Reputation, aiming enhances current recommender system.In the future work, we want to see how ISR results from this step can be used to improve recommender system results through experiments

## 6. REFERENCES


[1] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." the Journal of machine Learning research 3 (2003): 993-1022.

[2] Brody, Samuel, and Noemie Elhadad. "An unsupervised aspect-sentiment model for online reviews." Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2010.

[3] Chen, Pei-Yu, Samita Dhanasobhon, and Michael Smith. "All reviews are not created equal: The disaggregate impact of reviews and reviewers at amazon. com." Com (May 2008) (2006).

[4] Hu, Minqing, and Bing Liu. "Mining and summarizing customer reviews."Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2004.

[5] Esuli, Andrea, and Fabrizio Sebastiani. "Sentiwordnet: A publicly available lexical resource for opinion mining." Proceedings of LREC. Vol. 6. 2006.

[6] Feczko, Matthew, et al. "SentiSummary: Sentiment Summarization for User Product Reviews."

[7] Griffiths, Thomas. "Gibbs sampling in the generative model of latent dirichlet allocation." Unpublished note. http://citeseerx. ist. psu. edu/viewdoc/summary(2002).

[8] Ganu, G., N. Elhadad, and A. Marian, Beyond the Stars: Improving Rating Predictions using Revew Text Cotent. Twelfth International Workshop on the Web and Databases, Providence, Rhode, Island, USA, 2009

[9] Jo, Yohan, and Alice H. Oh. "Aspect and sentiment unification model for online review analysis." Proceedings of the fourth ACM international conference on Web search and data mining. ACM, 2011.

[10] Lin, Chenghua, and Yulan He. "Joint sentiment/topic model for sentiment analysis." Proceedings of the 18th ACM conference on Information and knowledge management. ACM, 2009.

[11] Mimno, David, et al. "Optimizing semantic coherence in topic models."Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2011.

[12] Mei, Qiaozhu, et al. "Topic sentiment mixture: modeling facets and opinions in weblogs." Proceedings of the 16th international conference on World Wide Web. ACM, 2007.

[13] Mudambi, Susan M., and David Schuff. "What makes a helpful online review? A study of customer reviews on Amazon. com." Mis Quarterly 34.1 (2010): 185-200.

[14] Sauper, Christina, and Regina Barzilay. "Automatic Aggregation by Joint Modeling of Aspects and Values." Journal of Artificial Intelligence Research 46 (2013): 89-127.

[15] Toutanova, Kristina, and Christopher D. Manning. "Enriching the knowledge sources used in a maximum entropy part-of-speech tagger." Proceedings of the 2000 Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics-Volume 13. Association for Computational Linguistics, 2000.

[16] Titov, Ivan, and Ryan McDonald. "A joint model of text and aspect ratings for sentiment summarization." Urbana 51 (2008): 61801.

[17] Transition words and phrases list,<http://www.smart-words.org/transition-words.html>. 2013

[18] Wilson, Andrew T., and Peter A. Chew. "Term weighting schemes for latent dirichlet allocation." Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2010.

[19] Zhu, Xiaojin, et al. "Improving diversity in ranking using absorbing random walks." Proceedings