User Intent Identification from Online Discussions Using a Joint Aspect-Action Topic Model

Ghasem Heyrani-Nobari, Tat-Seng Chua
National University of Singapore
{ghasem,dcscts}@nus.edu.sg

Abstract

Online discussions are growing as a popular, effective and reliable source of information for users because of their liveliness, flexibility and up-to-date information. Online discussions are usually developed and advanced by groups of users with various backgrounds and intents. However because of the diversities in topics and issues discussed by the users, supervised methods are not able to accurately model such dynamic conditions. In this paper, we propose a novel unsupervised generative model to derive aspect-action pairs from online discussions. The proposed method simultaneously captures and models these two features with their relationships that exist in each thread. We assume that each user post is generated by a mixture of aspect and action topics. Therefore, we design a model that captures the latent factors that incorporates the aspect types and intended actions, which describe how users develop a topic in a discussion. In order to demonstrate the effectiveness of our approach, we empirically compare our model against the state of the art methods on a large-scale discussion dataset crawled from the apple discussion forum with over 3.3 million user posts from 340k discussion threads.

1 Introduction

Online discussions are usually developed by users with diverse backgrounds and intents. For example, in a discussion regarding to a movie, a group of users may try to discuss the movie ending, while others may focus on the story-line and plot-holes. Therefore, it is essential for other users to know about the underlying intents of an online discussion before joining them.

To identify latent topics among documents, *aspect* (such as ”iPhone”, ”Internet”, ”Wi-Fi”, etc.) terms are among the main features used in most studies for topic modeling. However, in online discussions, for one subject we may find several discussions with various intents. In this case, if we just utilize the aspect terms, we may have problem in clearly distinguishing between different intents among discussions, as they all share a similar set of aspects. Thus, we need another feature set that can help to identify the potential intents among the user posts. *Actions* are kind of activity or functioning of a group of aspects (Lin et al. 2012), which lead us to generate the *aspect-action* relationship model.

Table 1: Discrimination power of action topics between discussions with similar subject ”iPhone slow internet”.

Aspects	apple, iphone, wifi, network, safari, connection	
Actions	*General Discussion*	*Technical Discussion*
	try, tell, post, help, call, support	reboot, fix, upgrade, check, install

The relations between aspects and actions in a discussion can help us to identify complex connections and latent intents among user posts. Here we assume that each discussion has its own underlying intent. In this paper, we define ”*intent*” as a set of aspect/action topics that describe the key semantics of a discussion.

As we can see from Table 1, we have two discussions based on ”iPhone slow internet” subject. However it shows that for one set of aspects, we have different set of action topics between the two discussions with different user intents.

For accurate and comprehensive modeling of online discussions we need a unified framework that is applicable to this broad domain. In previous studies, the focus was more on topic modeling, taxonomy generation and content summarizations (Moghaddam 2012; Carman et al. 2010; Jain and Pennacchiotti 2011; Pantel and Fuxman 2011). However, in the context of online discussions, there are lots of users posts with various lengths and informal languages, all taking place in a dynamic environment within a threaded structure. Also within a lifespan of a discussion, users may shift between various sub-topics and discuss several topics in parallel all making use of an overlapping set of features. As a result of such behaviors, current approaches are not able to accurately model online discussions in real world scenarios.

To the best of our knowledge we are the first to propose the learning and modeling of aspect-action relations within online discussions. To overcome these limitations, we propose a novel model by jointly modeling the aspect and action topics inside a threaded discussion. Through the model we are able to find the group of highly connected topics, that

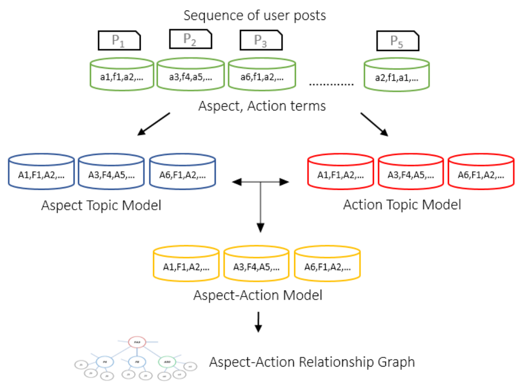

Figure 1: Framework for generating the Aspect-Action Relationship Graph from the sequence of user posts.

enable us to automatically identify the user's objectives and intents within a discussion. We carry out extensive evaluation of our joint model on a large scale discussion dataset crawled from apple discussion forum.

2 Problem Statement

For each discussion we have a sequence of user posts. For the modeling process, we first need to identify and extract aspect and action terms from these user posts. We utilize the POS Tagger and shallow parsing methods to derive the group of aspect and action terms. Next, by assuming that each user post is a document, we run two separate generative process and topic sampling: one is based on the aspect terms and the other on action terms. By generating the *aspect topic model* and *action topic model*, we replace the existing aspect and action terms with their associated topic ids. Then, we generate the third topic modeler on the sequence of user posts to derive the joint aspect-action topic model. Finally, by generating the relationship graph and calculating the coreness of the graph, we are able to obtain the ranked order of potential discussion intents. The overview of the framework is illustrated in Figure 1.

Here are the main definitions of our approach:

DISCUSSION: The action or process of a *group of users*, describing a *topic*, typically in order to reach a decision or to converge their knowledge. Here we assume that each discussion has its own underlying intent.

THREAD & POST: A *Thread* is created by a user with a given topic which continues with other users' posts and replies.

ASPECT / ACTION: *Aspect* refers to any features that has been discussed in user posts, usually in noun form, such as "wi-fi", "router", "network connection". *Action* refers to all kind of functionalities of aspects, usually in verb form, such as "connect", "shut down", "restore", etc.

ASPECT/ACTION PAIRS : It refers to a relation between one aspect and its associated action (e.g., "[slow] wifi", "[Update] iPhone5").

INTENT : Here, we define "intent" as a set of aspect-action pairs, that derived from the latent relationship of aspect and action topics within a sequence of user posts. For instance, for the subject "iPhone slow internet", we may have the following set of aspect and action topics that infer a possible discussion intent: {'wifi', 'router', 'fix', 'reboot', 'upgrade'}.

TOPICS & SUB-TOPICS: $T = t_1, t_2, t_3, ..., t_n$ is a set of topics that t_i indicates a topic and $r(t_i, t_j)$ denotes that t_j is a sub-topic of t_i. $R = r_1, r_2, ..., r_n$ is a set of sub-topic relations between the topics in set T.

3 Methodology

3.1 Extraction of Aspect/Action Terms

The first step of our generative process is to detect and extract relevant aspect and action terms from user posts. As we mentioned earlier, we assume that aspects are generally of type nouns and actions are of type verbs. There exists various methods for extracting terms and entities from documents. In this paper, as we need to handle user posts, with short and informal texts, we utilize the shallow parsing methods to analyze each user post to identify the constituents (noun groups, verbs, verb groups, etc.)

We accomplish this in two steps: a) we run a POS tagger(Toutanova et al. 2003) on user posts, to tag all the terms with their associated POS labels such as NN, CD, VB, etc. b) We capture the compound terms involving aspects and actions such as the 'iPad front camera' and 'tried resetting', we group these relevant POS tags in terms of NLP rules for compound nouns and compound verbs after removing the stop words (Punyakanok, Roth, and Yih 2008). Algorithm 1 shows the detail of extracting aspect and action terms from a sequence of user posts.

Algorithm 1 Extracting Aspect/Action Terms

Input: array of user posts in a discussion thread.
Output: array of aspect/action terms.
Method:

```
for all Posts p in Discussion Thread T do
  - Run the POS Tagger to get the POS label of each term.
  for all Terms t in Post p do
    - For Aspects check the following POS Tag combination
      rules: {"NN", "NNS", "CD", "JJ", "IN"}
    - For Actions check the following POS Tag combinations rules: {"VB", "VBP", "VBZ", "MD", "RB", "RP",
      "WRB", "IN"}.
    if rules match with the group g_{i-1} then
      add t_i to the group g_{i-1}
      else create empty group g_i and add t_i
    end if
    if g_i contains STOPWORD then
      Remove group g_i
    end if
  end for
end for
```

3.2 Aspect-Action Topic Model

In user discussions, we usually have a variety of posts with different properties and structures. Our approach consists of three topic models: (a) aspect topic model; (b) action topic model; and (c) joint aspect-action topic model. The first two models have their specific distributions over user posts. For the joint aspect-action topic model, topics can be sampled from the mixture of these two models. For example, in the

Table 2: Meaning of the notations

Symbol	Description
$F_{d,p,n}$, $A_{d,p,m}$	the n_{th}, m_{th} aspect, action term in user-post p of discussion thread d
$X_{N_{d,p}}$	generated aspect topics from #N aspects for post p
$V_{M_{d,p}}$	generated action topics from #M actions for post p
$Z_{p,d}$	sampled aspect-action relationship topic from posts in discussion thread d
α, β, γ	the parameters of Dirichlet distribution priors
θ^d	the distribution of aspect-action relationship in discussion threads
$\phi^{f,t}$	the distribution of aspect terms in user-posts
$\phi^{a,t}$	the distribution of action terms in user-posts.
X, V	set of aspect terms X and action terms V for each user post p

discussion based on "iPad Wi-Fi issue", we have ["wifi", "router" and "iOS"] topics from the aspect distribution, and ["connect", "upgrade", "restore"] topics from the action distribution. As a result of our joint model, we may have the following sets of topics: [iOS, wifi, restore], [wifi, connect] and [router, upgrade]. We note that the use of aspect-action relationship is crucial to modeling online discussions. If we were to model only aspects, we cannot differentiate between various types of discussions which can only be distinguished by the relations between action-aspect pairs in user posts. The as-ac relations will help our topic modeler to clearly learn the user intents toward topics.

The Latent Dirichlet Allocation (LDA) model is one of the most popular topic models based on the assumption that documents are mixtures of topics, where a topic is a probability distribution over words (Blei, Ng, and Jordan 2003). The LDA model is effectively a generative model from which a new document can be generated in a predefined probabilistic procedure.

In order to model the Aspect-Action relation, we propose a new joint aspect-action topic model on online discussions by adding an additional layers of latent variables between user-posts and topic layers. A graphical model of our approach is represented in Figure 2.

Assume that we have a discussion thread d with a collection of user posts as $\{p1, p2, .., p_n\}$, where each user post p of the discussion thread d contains a sequence of terms denoted by $T \in \{f, a | f \in aspects, a \in actions\}$. For the conditional probability $P(P_n|D_n)$, we have their joint probability as $P(P_n, D_n)$. The formal definition of the generative process is as follows:

- For each discussion thread, we draw the aspect-action relationship topics from $\theta^d \sim Dir(\gamma)$
- For each user post p, we draw two separate topics distributions, one from $\phi^{f,t} \sim Dir(\alpha)$ and the other from $\phi^{a,t} \sim Dir(\beta)$
- For each term i, in user-post p of discussion thread d:
  - choose aspect topic $X_{N_{d,p}} \sim \phi^{f,t}$ from the aspect sets.
  - choose action topic $V_{M_{d,p}} \sim \phi^{a,t}$ from the action sets.
  - choose an aspect-action relationship topic $Z_{p,d}$ from the aspect-action distribution θ^d_z

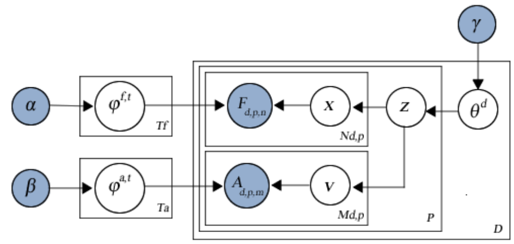

Figure 2: Aspect-Action Relationship Model

Figure 3, shows the hierarchical connections between aspect and action together with their associated topics in user-posts. The end nodes in this graph are the actual terms in user posts.

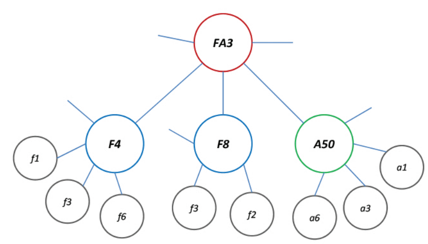

Figure 3: A graphical representation of aspect-action relationship model. $\{f_i, a_i\}$: aspect/action terms, $\{F_i, A_i\}$: aspect/action topics, and FA_i: joint aspect-action topics.

In order to obtain the distributions of $\theta^d, \phi^{f,t}$ and $\phi^{a,t}$, we first estimate the posterior distribution over $X_{Nd,p}$ and $V_{Md,p}$ and then over $Z_p = z$. We then calculate the conditional distribution:

$$P(Z_p = z, X_{Nd,p} = x, V_{Md,p} = v | z_{\neg}, x_{\neg}, v_{\neg}, A, F) \quad (1)$$

(a) We start by showing how the joint probability of aspect and action topics and their relationship $p(z, x, v, A, F)$ can be derived:

$$P(z, x, v, A, F) = P(A, F|x, v)P(z, x, v) = P(F|x)P(A|v)P(z, x, v) \quad (2)$$

In order to compute the third term, we factor it as: $P(z, x, v) = P(z|x, v)P(x)P(v)$, where for $P(z|x, v)$, we have:

$$P(z|x, v) = (\frac{\Gamma(R\gamma)}{\Gamma(\gamma)^R})^P \prod_x \prod_v \frac{\prod_z \Gamma(C_{xvz} + \gamma)}{\Gamma(C_{xv} + R\gamma)} \quad (3)$$

where P is the total number of user posts, R is the total number of aspect-action relationship topics, C_{xvz} is number of times aspect x and action v appeared in a relationship topic of z. C_{xv} is number of times aspect x and action v appeared in one post.

(b) We compute each of the factors individually to estimate the conditional probability for as-ac relationship topic as follows:

$$P(Z_p = z, X_{Nd,p} = x, V_{Md,p} = v | z_\neg, x_\neg, v_\neg, A, F) \propto \frac{C_x^{F,T_f} + \alpha}{C_x + F\alpha} \times \frac{C_v^{A,T_a} + \beta}{C_v + A\beta} \times \frac{C_{xvz}^{PR} + \gamma}{C_{xv} + R\gamma} \quad (4)$$

(c) We utilize Gibbs sampling to sequentially sample each variable of interest from the distribution over that variable given the current values of all other variables and the data. As a result we can easily generate the values of θ_z^d, $\phi_x^{f,t}$ and $\phi_v^{a,t}$:

$$\begin{aligned} \theta_z^d &\propto \frac{C_{xvz}^{PR} + \gamma}{\sum_{k=1}^{R} C_{xv}^{PR} + R\gamma} \\ \phi_x^{f,t} &\propto \frac{C_x^{F,T_f} + \alpha}{\sum_{k=1}^{F} C_x^{F,T_f} + F\alpha} \\ \phi_v^{a,t} &\propto \frac{C_v^{A,T_a} + \beta}{\sum_{k=1}^{A} C_v^{A,T_a} + A\beta} \end{aligned} \quad (5)$$

where θ_z^d is the distribution of aspect-action relationship in a discussion thread; $\phi_x^{f,t}$ is the distribution of aspects in user-posts; and $\phi_v^{a,t}$ is the distribution of actions in user-posts.

3.3 Aspect-Action Relationship Graph

After we have estimated the values of $\theta_z^d, \phi_x^{f,t}$ and $\phi_v^{a,t}$, we obtained the matrices of counts with dimensions $R \times P$, $F \times T_f$ and $A \times T_a$ respectively. By utilizing these values and the array of user-posts which are already ordered by their submission time, we can generate the relationship graph G(V,E) by using the following procedures:

Algorithm 2 Relationship Graph Generation

Input: $\theta_z^d[R \times P]$, $\phi_x^{f,t}[F \times T_f]$, $\phi_v^{a,t}[A \times T_a]$: Matrices; UP: array of user-posts ordered by time
Output: G(V,E): An undirected weighted complete graph where $V \in \{F_{d,p,n}, A_{d,p,m} | F \in aspects, A \in actions\}$ and E represent relation between two terms of weight 0;
Method:
  for all UP $\in$ Discussion Thread D **do**
    for all Aspect and Action topics in user-post p **do**
      - Increment weight of each edge $e(V_1, V_2)$ where $V_1 \in F_{d,p,n}$ and $V_2 \in A_{d,p,m}$
    end for
  end for
  - Sort E in decreasing order of weights

For each user-post p from the discussion thread d, we create a new link between all aspect and action topic nodes $F_{d,p,n}$ and $A_{d,p,m}$. However if the link already existed, we simply increase its weight. After traversing all the user-posts in thread d, we can sort all links by their respective weights. Full details of the procedure can be found in Algorithm 2. The ranking order of nodes and connections in graph G represents the impact and values of sub-topics for each discussion thread. $E = V \times V$ is the set of possible transitions between any two nodes in graph G. Figure 4 shows the graph of aspect-action relationship generated based on the "Wi-Fi & Network" topic.

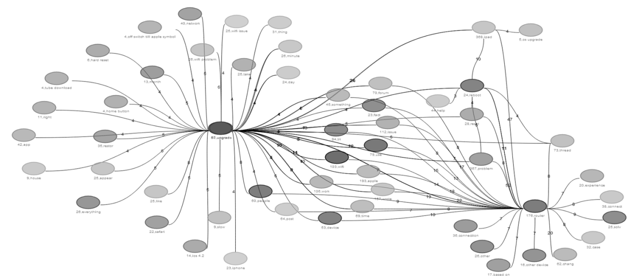

Figure 4: Graphical representation of aspect-action relationship, generated based on "Wi-Fi & Network" Topic.

3.4 Identification of Discussion Intents

In this paper, we represent the aspect-action relationship as a graph $G = (V, E)$, where V corresponds to a set of vertices $v \in V$ and $E \in V \times V$ refers to the set of edges linking the vertices.

K-core decomposition was first proposed by (Seidman 1983; Alvarez-Hamelin et al. 2005) indicated that when the number of nodes in a graph is highly increased, it is get very difficult to understand the unique relationships between edges and complex interconnections between attributes of their nodes. Formally, the K-core decomposition is defined as follows: A subgraph $H = (C, E|C)$ induced by the set $C \subseteq V$ is a k-core or a core of order K iff $\forall v \in C$: $degH(v) \geqslant K$, and H is the maximum subgraph with this property. The k-core of a graph G is the largest subgraph such that every node in the subgraph has at least k degrees. The k-core of a graph is determined by recursively pruning all vertices whose degrees are less than k, until a new subgraph whose vertices have at least k degrees is formed. A vertex v has a coreness c if it belongs to the c-core but not to $(c+1)$-core. We represent c_v as the coreness of vertex v. A shell S_c consists of all the vertices whose coreness is c. The maximum value c such that S_c is not empty is denoted by c_{max}. The $K - core$ is thus the union of all shells S_c with $c \geqslant k$.

Theoretically, the higher the coreness of a vertex, the more connected it is in the graph. By calculating the coreness c and shell S_c for the graph of aspect-action relations, we are able to discover highly relevant and connected topics that we call user intent. Each discussion may have several user intents with various support. Next we will discuss about our experiments and results.

4 Experiments and Evaluations

In order to qualitatively and quantitatively evaluating our model, we perform the following types of evaluation analysis. *Convergence Analysis:* We use perplexity value to measure the convergence of our approach as compared to LDA and MG-LDA models; we also conduct experiments with different settings such as the number of topics or iterations to evaluate the convergence of our model. *Accuracy Analysis:* We measure the accuracy of our method in terms of capturing user intent in a discussion, where intent is modeled as a set of as-ac pairs. Here we measure the accuracy of our method as compared to other baseline methods based on the number of correct aspect and action terms that we identified from the relationship graph.

4.1 Experimental Setup

We setup our dataset using the data crawled from Apple discussions[1]. For each discussion we have extracted the following properties: topic tags, number of replies and views, date/time and status of each user posts.

Table 3: Meta statistics of our datasets.

Datasets	#Threads	#Posts	#Main Topics
iPad	$\sim 127,000$	$\sim 1,400,000$	21
iPhone	$\sim 84,000$	$\sim 1,000,000$	26
iPod	$\sim 55,000$	$\sim 320,000$	16
Macbook	$\sim 73,000$	$\sim 600,000$	19
Total	$\sim 339,000$	$\sim 3,320,000$	82

For the experiments, We have selected four apple products as shown in table 3. They contain about 340,000 discussion threads with over three million user submissions. The examples of topics includes ”network”, ”application”, ”Battery” etc. On average, each thread has 5 posts. In apple forum, each discussion and post contain a set of topic tags; these tags are generated either by the system or by the users. For instance based on the discussion ”How do I make IOS 6 keep wifi on”, we obtain the set of tags such as ”ipod, wireless, wifi, router, connect, lock, sleep, ios6” and we use them as the key topics for this discussion. To evaluate the performance of our model, we compare our ranked topics with these set of tags as our ground truth. To do sampling from our relationship graph, we select the top ranking topics, such that they cover 70% of the topic instances labeled by our model. For our testing set, we categorize our dataset into six major topics as follows: ”Wi-Fi & Network”, ”Backup & Sync”, ”3G & Data”, ”Mail & Messages”, ”Accessories”, ”Camera & Photos”.

4.2 Joint Aspect-Action Evaluation

We use perplexity to measure the convergence of topics to measure the ability of our model acting as a generative model. Perplexity is an important indicator to demonstrate the generalization performance of a model which is widely used in the language modeling fields to evaluate the predictive power of a model. A lower perplexity value means that the topics are less surprising to the model and their user preference is similar with model's prediction. Given a test dataset D, the perplexity value can be calculated as follows:

$$Perplexity(D) = exp\{-\frac{\sum_{d=1}^{P} \log(p(w_d|D_{train}))}{\sum_{d=1}^{P} N_d}\} \quad (6)$$

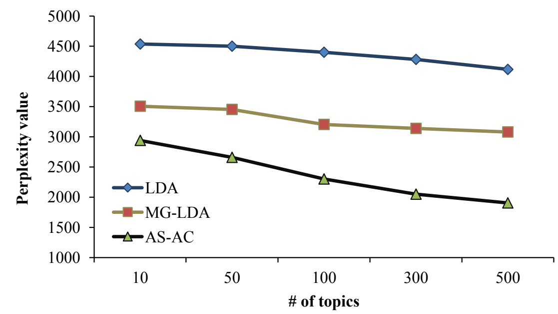

Figure 5: Perplexity values versus number of topics.

Figure 5 shows the result of the perplexity comparison of LDA, MG-LDA and our approach. It shows that AS-AC model has lower perplexity values as compared to the other two models. This is due to the ability of action terms that correctly identifies aspect terms for each topic in our sampling and generative procedure. The results also indicates that our model can achieve better perplexity performance with a smaller size of topics, as on average, we have about 4-6 topics in each thread. Our model has a relatively stable perplexity values even with a high number of topics, which demonstrate that it owns a better predicting ability for unseen user posts by incorporating as-ac relations into the process of extracting topics. This will lead to better perplexity performance when the number of topics is smaller than the other two topic models.

4.3 Topic Ranking Evaluation

To evaluate the accuracy of our joint aspect-action model and our generated relationship graph, we utilize the metric of precision $@n$, which gives the precision at different topic rank positions. As a result we are able to evaluate how well our learned relationship model performs in discovering the correct set of topics for each discussions thread.

Table 4 presents the results on different n values. The results indicate that our approach outperform the other two models, LDA and MGLDA. What is remarkable here is that our approach is able to achieve a much higher performance for the value of $n = 10$, which indicates that it is able to rank more accurate topics at higher ranked positions. Figure 6 shows the precision/recall curves for the top three topics in our dataset evaluated using three different settings: (a) applying as-ac relations, (b) only aspect's(as) relations and (c) only action(ac) relations. It shows that by applying the as-ac

[1]http://discussions.apple.com

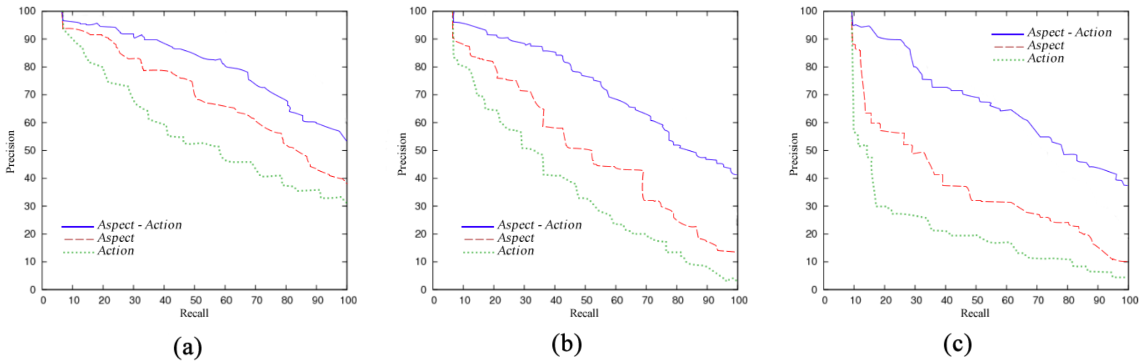

Figure 6: Precision/Recall curves for the top three topics: (a)Wi-Fi, (b)Backup/Sync, (c)Shipping/3G.

Table 4: Performance in terms of P@n for topic detection on LDA, MG-LDA and AS-AC

Topics	LDA			MG-LDA			AS-AC		
	P@10	P@50	P@100	P@10	P@50	P@100	P@10	P@50	P@100
Wi-Fi & Network	0.51	0.58	0.63	0.70	0.65	0.67	0.96	0.85	0.83
Backup & Sync	0.50	0.65	0.60	0.60	0.70	0.67	0.98	0.90	0.88
3G & Data	0.43	0.55	0.56	0.80	0.75	0.73	0.95	0.85	0.76
Mail & Messages	0.50	0.65	0.60	0.70	0.75	0.67	0.87	0.83	0.80
Accessories	0.68	0.70	0.63	0.70	0.70	0.63	0.93	0.85	0.78
Camera & Photos	0.55	0.65	0.50	0.60	0.80	0.76	0.86	0.84	0.81

relationship in online discussions, we can significantly improve both the topic detection and topic ranking of online discussions. As we can see, the precision of actions on topic ”(a) Wi-Fi” is at the highest, which shows the discrimination power of actions in this topic.

5 Related Work

Several unified models of topics, sentiment and users have been proposed recently. They have extended LDA (Blei, Ng, and Jordan 2003) to capture various inferences on sentiment trends, authorship and aspect ratings, all with topics generated from documents in user reviews and comments (Carenini and Murray 2012; Moghaddam 2012; Hassan, Qazvinian, and Radev 2010). However, studies on user discussions and forum threads, starts with considering threads as a collection of pages which moved toward more refined levels such as user posts and sentence level. We can categorize studies in this domain at various levels(Threads, Posts, Sentences). The preliminary studies is mostly based on the links and connection between different threads by considering hyperlinks and user replies (Xu and Ma 2006; Cong et al. 2008). However More recent studies are in user's post and sentence level modeling. In this category we have works on modeling thread structures to do indexing and thread ranking for finding discussion similarities and forum searching (Lee, Borodin, and Goldsmith 2008; Duan and Zhai 2011; Seo, Croft, and Smith 2009; Lafferty and Blei 2005; Singh 2012).

The use of Multi-Grain Latent Dirichlet Allocation model (Titov and McDonald 2008) has also been used to study the topic assignments at the sub-topic sentence level. However, this model is unable to handle a variety of user-post types, contiguity and ordering similarity, because sentences not in close proximity are only connected through a series of window frames. In contrast, our model is designed to capture both close and long-range topic dependencies in user posts.

6 Conclusions

The main contributions of this paper are two-fold. (a) We proposed a novel unsupervised topic modeler, that is able to identify primary topics with their dependencies from a sequence of user posts. In particular, we jointly model aspects with their associated actions to boost the precision of our generative process where actions play the role of defining the functionalities for a group of aspects. (b) We developed a model for the automated identification of user's objectives and intents within a discussion which can be used for extensive evaluation and comparisons between discussions. Also from the AS-AC graph, were able to calculate the semantic similarity between topics which can be used for entity identification. We have shown in our experiments that by considering as-ac relationship, we could achieve high accuracy in terms of finding the correct set of user intents and topics in online discussions.

References

Alvarez-Hamelin, J. I.; Dall'Asta, L.; Barrat, A.; and Vespignani, A. 2005. Large scale networks fingerprinting and visualization using the k-core decomposition. In *Advances in neural information processing systems*, 41–50.

Blei, D.; Ng, A.; and Jordan, M. 2003. Latent dirichlet allocation. volume 3, 993–1022. JMLR. org.

Carenini, G., and Murray, G. 2012. Methods for mining and summarizing text conversations. In *SIGIR12*, 1178. New York, New York, USA: ACM Press.

Carman, M. J.; Crestani, F.; Harvey, M.; and Baillie, M. 2010. Towards query log based personalization using topic models. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, 1849–1852. ACM.

Cong, G.; Wang, L.; Lin, C.; Song, Y.; and Sun, Y. 2008. Finding question-answer pairs from online forums. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, 467–474. ACM.

Duan, H., and Zhai, C. 2011. Exploiting thread structures to improve smoothing of language models for forum post retrieval. 350–361. Springer.

Hassan, A.; Qazvinian, V.; and Radev, D. 2010. What's with the attitude?: identifying sentences with attitude in online discussions. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 1245–1255. Association for Computational Linguistics.

Jain, A., and Pennacchiotti, M. 2011. Domain-independent entity extraction from web search query logs. In *Proceedings of the 20th international conference companion on World wide web*, 63–64. ACM.

Lafferty, J. D., and Blei, D. M. 2005. Correlated topic models. In *Advances in neural information processing systems*, 147–154.

Lee, H.; Borodin, A.; and Goldsmith, L. 2008. Extracting and ranking viral communities using seeds and content similarity. In *Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*, 139–148. ACM.

Lin, T.; Pantel, P.; Gamon, M.; Kannan, A.; and Fuxman, A. 2012. Active objects: actions for entity-centric search. In *Proceedings of the 21st international conference on World Wide Web*, 589–598. ACM.

Moghaddam, S. 2012. Aspect-based Opinion Mining from Product Reviews Categories and Subject Descriptors. In *SIGIR12*, 1184.

Pantel, P., and Fuxman, A. 2011. Jigs and lures: Associating web queries with structured entities. In *ACL*, 83–92.

Punyakanok, V.; Roth, D.; and Yih, W. 2008. The importance of syntactic parsing and inference in semantic role labeling. volume 34.

Seidman, S. B. 1983. Network structure and minimum degree. volume 5, 269–287. Elsevier.

Seo, J.; Croft, W.; and Smith, D. 2009. Online community search using thread structure. In *Proceedings of the 18th ACM conference on Information and knowledge management*, 1907–1910. ACM.

Singh, A. 2012. Retrieving Similar Discussion Forum Threads : A Structure Based Approach. In *SIGIR12*, 135–144.

Titov, I., and McDonald, R. 2008. Modeling online reviews with multi-grain topic models. In *Proceedings of the 17th international conference on World Wide Web*, 111–120. ACM.

Toutanova, K.; Klein, D.; Manning, C.; and Singer, Y. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, 173–180. Association for Computational Linguistics.

Xu, G., and Ma, W. 2006. Building implicit links from content for forum search. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, 300–307. ACM.