A variational Bayesian model for user intent detection

Conference Paper *in* Proceedings - ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing · May 2014

DOI: 10.1109/ICASSP.2014.6854367

5 authors, including:

Dilek Hakkani-Tur
Microsoft

Celikyilmaz Asli

Larry P. Heck
Institute of Electrical and Electronics Engineers

Gokhan Tur
Microsoft

A VARIATIONAL BAYESIAN MODEL FOR USER INTENT DETECTION

Yangfeng Ji[*1] *Dilek Hakkani-Tür*[2] *Asli Celikyilmaz*[2] *Larry Heck*[2] *Gokhan Tur*[2]

[1] School of Interactive Computing, Georgia Institute of Technology
[2] Microsoft

ABSTRACT

Intent detectors in state-of-the-art spoken language understanding systems are often trained with a small number of manually annotated examples collected from the application domain. Search query logs provide a large number of unlabeled queries that would be beneficial to improve such supervised classification. Furthermore, the contents of user queries as well as the clicked URLs provide information about user's intent. In this paper, we propose a variational Bayesian approach for modeling latent intents of user queries and clicked URLs when available. We use this model to enhance supervised intent classification of user queries from conversational interactions. Experiments were run with large volumes of search queries and show significant improvements over state-of-the-art systems.

***Index Terms*—** spoken language understanding, query click logs, variational inference, graphical models, intent classification.

1. INTRODUCTION

In task-oriented dialog systems, the aim of the spoken language understanding (SLU) models is to automatically capture and tag the semantic frames that include user intents and related concepts given their utterances [1]. State-of-the-art SLU models are trained with examples collected for each domain and task, and manually annotated according to a semantic schema, often designed for each one. Previous work mainly viewed SLU intent detection as an utterance classification task, and investigated various classification methods [2, 3, 4, 5, 6, 7, 8].

Intents of spoken queries to a dialog system may be classified as *informational*, *navigational*, and *transactional* in a similar way to the taxonomy for web search [9]. User utterances with navigational intents aim to navigate in the dialog, such as "*go back to the previous results*", and can usually be shared across ontologies of similar dialog system applications. Utterances with informational intents aim to seek an answer to questions, such as "*find the movies of a certain genre and director*". Answers to such queries are likely to be included in knowledge repositories, such as *Freebase*[1]. The ontology of the user intents for informational queries can be formed based on the semantic web ontologies [10, 11, 12, 13], such as the ontology of *Freebase* or *schema.org*[2]. Furthermore, the populated knowledge in the graph can be used to mine examples that include surface forms of entities and their relations in natural language [14, 15]. In our previous work [15], for each relation type in the semantic graph, we leveraged the complete set of entities that are connected to each other with a specific relation, and mined surface form patterns from the web that realize such relations in natural language sentences. These patterns are then used to detect relations in new user utterances to a spoken dialog system.

Utterances with transactional intents aim to perform an operation, such as "*play a movie*", or "*reserve a table at a restaurant*". The ontology of user intents for such queries are usually defined by dialog system designers and developers, and are mainly driven by the capabilities of the back-end applications. For Internet search queries, they can also be mined from search queries [16].

Earlier work used the URLs that users click on from amongst the search results to identify the intent of the user from their search query [17, 18]. In our previous work [12], we proposed a click intent model that uses search query click logs to discover new user intents in addition to the ones that appear in the semantic graphs. Although these models are efficient in capturing new user intents, they model the entities and context words together, whereas these may play a different role for intent detection. In this paper, we focus on entities and context words in a given utterance along with the clicked URLs and propose a novel click intent model to improve supervised intent detection. Our contributions are: (i) joint modeling of entities, entity types and context words in user utterances as aspects of utterances, (ii) incorporating the entire URL and its components rather than just the base URLs and (iii) using large amounts of search query click logs in the new click intent model for improving supervised intent classification.

In the following sections, we first provide examples motivating our new model, and then present the new model and its variations. In Section 4, we describe the inference for the new model. Then, we present our analysis of the discovered latent clusters, followed by the experimental results for intent detection and conclude.

2. MOTIVATION

We illustrate our motivation and main characteristics of the queries and associated URLs with two sample utterances from web search:

(1) "*find movies by James Cameron*"
(2) "*James Cameron*"

Both of these queries are related to searching some information about the director *James Cameron*. A significant difference between the two examples with respect to intent understanding is that the user's intent is clear in example (1), whereas the intent of example (2) is not as clear as the first one. This calls for additional information to better understand the user's intent. Specifically, if the clicked URL points to, for example, the Wikipedia page of *James Cameron*, it is likely that the user is looking for his biographical information. Similarly, if the URL points to a page within the imdb.com domain, then the user's intent is most likely similar to example (1). Furthermore, in order to obtain a better understanding about users' intents, we can investigate queries by segmenting them into parts with respect to the function of each part on intent understanding. Using the first utterance as an example, we first identify its intent

[*] The first author performed the work while at Microsoft.
[1] `www.freebase.com`
[2] `http://www.schema.org`

as "*find movie*". Note that, if we replace "*James Cameron*" with another movie director's name, the intent of example (1) would not change. Therefore, we propose that queries comprise of two types of word sequences: the context words and the entity words, both of which influence the way we model the intent understanding.

Following earlier work [16], we assume in our models that each search query has a single intent. In addition, in our statistical model, for efficiency reasons, we opt out the ordering of the context words as well as the entities in a given query and use bag-of-context-words and bag-of-entities without losing too much information.

3. MODEL SPECIFICATION

We formalize the intuitions described in the previous section into a Bayesian model that given the user's intent considers entities and context words from queries and the clicked URLs independently. By performing variational inference in this model, we recover a probability distribution of intents for a query. The plate notation of our graphical model is shown in Figure 1 and the symbols are summarized in Table 1.

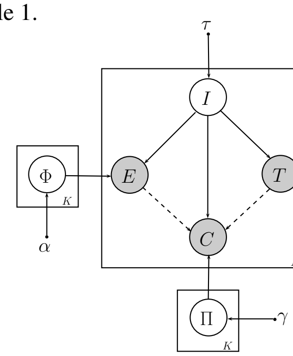

Fig. 1. Plate notation of our intent detection model. The model with the dependencies shown by dashed lines is considered as the full model, which is discussed in section 3.4.

Symbol	Description
I	Intent distribution
E	Entities in a query
T	Context words in a query
C	Clicked URL components
Φ_k, Ψ_k, and Π_k	Entity, context words and clicked URL components distribution of the kth intent
α, β, γ, and τ	Hyper-parameters of entity, context words, URL components and intent distributions

Table 1. Notation used in this paper.

As with all generative models, we interpret the model in a generative way: users start with an intent in mind, before doing search on the information they inquire. This intent guides the way for generating their query and the following clicks. Keeping this in mind, we formulate our generative story and specify the distributions of the related random variables, starting with the abstractions in the next section.

3.1. Distributions of Intent, Entities and Context words

Following the characteristics discussed in section 2, each query has only one intent. Therefore, for query d, the intent distribution is a multinomial $I_d \sim \mathrm{Mult}(\mathrm{n}, \tau)$, where $n_I = 1$. Furthermore, a query Q can be split into two parts: entities, E, and context words, T. With the bag-of-words assumption, both E and T are defined as multinomial distributions. Given a query with intent k, its entities are sampled from $E \sim \mathrm{Mult}(\mathrm{n_e}, \Phi_\mathrm{k})$, where n_e is the number of entities, and the context words are sampled from $T \sim \mathrm{Mult}(\mathrm{n_t}, \Psi_\mathrm{k})$, where n_t is the number of context words.

An important property of our model is that we do not consider the entity names as observed values, which would otherwise yield an inefficient bag-of-words model. This is simply because if we use the entity names directly, the vocabulary size would be very large. In addition, in the model design the numeric representation of a query will be a high-dimensional sparse vector. Therefore, rather than the entity names, we use the category types of the entity names. Specifically, we introduce a preprocessing step — before modeling the distributions of a query, by replacing the entities with their entity categories. For example, in the running example (1), we replace "*James Cameron*" with its entity category, "*director*". We construct the vocabulary based on the entity categories. There are several advantages of using entity categories: first, it can obviously reduce the vocabulary size, which will save our computational resources. Secondly, for a given query, an entity category can be more important than the value of the entity in understanding the intent of the user. This processing step will strengthen the occurrence statistics among different queries, which will eventually lead to a more confident estimation of intent distribution.

3.2. Distributions of URL Components

We tokenize each URL and represent them as pseudo-documents considering bag-of-URL-tokens assumption. For tokenization, we split each URL into tokens with respect to slashes ("/"). Our motivation in representing URLs as bag-of-URL-tokens is that we find that some tokens of URLs are actually meaningful lexical units that reflect valuable information about the associated web page. Furthermore, these terms may be used with with the same purpose across different sites. Take the following URL from the query log dataset for example:

www.imdb.com/title/tt1518740/review

After tokenization, we end up with four tokens: *www.imdb.com, title, tt1518740, review*. Each token corresponds to an aspect of its associated web page, i.e., "*www.imdb.com*" refers to the IMDB website,"*review*" indicates that this web page is a review page about the movie with ID "*tt1518740*". Thus, we represent URL tokens as random variables of our graphical model. We construct a vocabulary for the tokens extracted from the observed URLs in our data. Each URL is a random variable with a multinomial distribution $\mathrm{Mult}(\mathrm{n_c}, \Pi_\mathrm{k})$ with Π_k if it is sampled from the k-th intent.

3.3. Formal Generative Story

We define our generative intent detection model as follows:

1. For each intent k, draw distributions of entity categories Φ_k, context words Ψ_k and URL components Π_k,
2. For each query d draw the intent I_d from $\mathrm{Mult}(1, \tau)$, entities E_d from $\mathrm{Mult}(\mathrm{n_e}, \Phi_{\mathrm{I_d}})$, context words T_d from $\mathrm{Mult}(\mathrm{n_t}, \Psi_{\mathrm{I_d}})$.
3. For each query d a click pertaining to a certain URL with components C_d are drawn from $\mathrm{Mult}(\mathrm{n_c}, \Pi_{\mathrm{I_d}})$[3]

3.4. Investigation of URL Components in Click-Intent Model

One possible extension of our model is to introduce a dependency between a query and its associated URL. Specifically, we can extend our current model by introducing a dependency link between the entity categories and the URL components, as well as the context words and the URL components (demonstrated in Figure 1 with the dashed lines). Unfortunately, these dependencies introduce an additional cost while estimating the random variable Π. With the

[3]Similar to LDA, n_e, n_t and n_c are also drawn from Poisson distributions with certain parameters. For simplicity, we ignore this part.

introduction of the dependencies, the total number of parameters associated for Π would be $O(K \cdot |E| \cdot |T| \cdot |C|)$, which would require large amounts of data for training. Therefore, in this paper, we did not include those dependencies and left it as future work.

3.5. Missing Values

Our corpus comprises of different types of queries with missing information, that introduce additional challenges to our graphical model design. On one hand, we have intrinsically short queries from web search query logs. They either lack context or entity words. Consider the query (2) of the running example in Section 2. It is a single entity query with no context words. On the other hand, we have translated spoken language queries from a conversational dataset which lack the click information. Henceforth, we introduce the missing value component into our graphical model by way of prior information to compute the expected values of the random variables with missing values. Specifically, following the non-informative priors[19], we impute some values without adding any biased information to specify the prior distribution.

4. INFERENCE

As in LDA, exact inference in this model is computationally intractable [20]. We derive a variational main-field inference, approximating the distribution over intents, entity categories, context words and URL components with one fully factorized variational distribution $q(I, \Phi, \Psi, \Pi)$

$$q(I, \Phi, \Psi, \Pi; \lambda, \mu, \nu, \rho) = \prod_{d=1}^{D} q(I_d; \lambda_d) \prod_{k=1}^{K} q(\Phi_k; \mu_k) \prod_{k=1}^{K} q(\Psi_k; \nu_k) \prod_{k=1}^{K} q(\Pi_k; \rho_k)$$

where $q(I_d; \lambda_d)$ is the multinomial distribution $\mathrm{Mult}(1, \lambda_d)$, $q(\Phi_k; \mu_k)$, $q(\Psi_k; \nu_k)$ and $q(\Pi_k; \rho_k)$ are Dirichlet distributions. $\{\lambda_d\}$, $\{\mu_k\}$, $\{\nu_k\}$ and $\{\rho_k\}$ are variational parameters. As we can observe from Eq.(1), there is no dependence between any two random variables in q.

Given a variational distribution with specified variational parameters, we can measure the difference between the posterior distribution $p(I, \Phi, \Psi, \Pi|E, T, C; \alpha, \beta, \gamma, \tau)$ and the variational distribution $q(I, \Phi, \Psi, \Pi; \lambda, \mu, \nu, \rho)$ using KL divergence, which leads to a measurement called variational bound — by maximizing the variational bound, we will find the best approximation based on the full factorization assumption.

Since all the variational parameters are coupled, we have to update all the variational parameters iteratively. Due to the property of conjugate family and the fully factorization, we can find the closed form for all the update equations.

4.1. Intent Distribution

For a given query Q, the variational distribution of I_d is given by a multinomial distribution with parameter λ_d, where the probability for $I_d = k$ is:

$$\lambda_{dk} \propto \exp\Big\{ \sum_i e_{di}(\psi(\mu_{ki}) - \psi(\mu_{k0})) + \sum_r t_{dr}(\psi(\nu_{kr}) - \psi(\nu_{k0})) + \sum_j c_{dj}(\psi(\rho_{kj}) - \psi(\rho_{k0})) + \log \tau_k \Big\} \quad (1)$$

where $\nu_{k0} = \sum_{i=1} \mu_{ki}$, $\nu_{k0} = \sum_{r=1} \nu_{kr}$, $\rho_{k0} = \sum_{j=1} \rho_{kj}$, $\psi(\cdot)$ is the digamma function and also the logarithmic derivative of the gamma function.

4.2. Distributions of Entities and Context Words

For a given intent k, we also need to update the variational parameters for the distribution of entity categories, context words, and URL components.

The probability of i-th category in intent k is

$$\mu_{ki} = \alpha_i + \sum_{d=1}^{D} e_{di}\lambda_{dk}. \quad (2)$$

and the probability of r-th context word in intent k is

$$\nu_{kr} = \beta_r + \sum_{d=1}^{D} t_{dr}\lambda_{dk}. \quad (3)$$

Similarly, the probability of j-th URL component in intent k is

$$\rho_k = \gamma_j + \sum_{d=1}^{D} c_{dj}\lambda_{dk} \quad (4)$$

Due to the property of conjugate distributions, there is a closed form for every update equation as in Eq.(1) - Eq.(4). In the inference stage, we can use the updating equations directly without employing any optimization routine explicitly. The updating process is summarized in Algorithm 1.

Algorithm 1 Variational inference algorithm for our model.

```
while not converged do
    for d = 1, ..., D do
        for k = 1, ..., K do Update λ_dk using equation 1
    for k = 1, ..., K do
        for i = 1, ..., |E| do Update μ_ki using equation 2
        for r = 1, ..., |T| do Update ν_kr using equation 3
        for j = 1, ..., |C| do Update ρ_kj using equation 4
end while
```

5. EXPERIMENTAL RESULTS

We evaluate our model on both query logs data and conversational data sets. The purpose of the evaluation is to verify whether our model can discover any interesting intents and help us to improve the performance of conventional understanding.

5.1. Data and Experimental Setting

We use two different kinds of datasets: (1) query logs with clicked URLs that were sampled from Bing search engine logs using 13 base URL names (i.e., `www.{imdb,netflix,fandango}.com`); (2) conversational dataset, from the same domain, that was collected using crowd sourcing, and manually cleaned and annotated by domain-expert human annotators. The basic information of the two datasets are described in Table 2.

The experiments include two parts. The first part is the qualitative evaluation of our model on the query logs data to demonstrate that our model can discover interesting and meaningful intents. The second part of the experiment is a quantitative evaluation of our model. We apply the model on the conversational data together with the query logs data. As discussed in section 3.5, in our model, the conversational dataset is used as query logs data with missing values on C. In this case, the distributions of URL components estimated from the query logs data are used as potential click information (prior information) for the conversation dataset.

Search queries	Number of query logs	80,512
	Number of entity categories	27
	Number of context words	9720
	Number of URL components	1531
Conversational queries	Number of training examples	3575
	Number of dev examples	1200
	Number of test examples	1200

Table 2. Data description.

As mentioned in section 2, we assume that each query can be split into two parts: entities and context words. Because the entity extraction (*or* labeling) is a research topic on its own [21], in this paper, we only use a simple entity extraction method [22]. First, we construct a list of entities using Freebase. The list includes 21 types of entities that include movie names, actors, directors, and genre. Each entry in this list consists of one entity and its entity category. Then, we refine this entity list by removing some highly ambiguous entities. For example, entity "*UP*" refers to a movie name, but it also could be just a single word. For entity extraction, we use the longest word match between the query and the entity list. In the experiments, we noticed that an accurate entity extraction method will definitely boost the performance of intent detection. We leave this as a future work.

5.2. Analysis of New Intent Detection

Table 3 shows the top entity, context words and URL components estimated by our model in the first part of experiments. In this experiment, we only used the query log data with the number of intent $K = 40$. We show examples for two randomly picked intents. There are several different ways to define the intent name for both examples. One can notice the correlations between the top terms of each intent. For instance, the term *starring* from the top entity categories, *bio, actor* from the top context words and *person, biography.html* from the top URL components indicate that this intent is probably searching for the biographical information about a movie star. Similar correlations are observed in the second example. The terms *film/name, film/genre* from the top entity category, *office, box, new* from the top context words and *news* indicate that this intent is most likely related to finding information about new movies.

Top entity categories	Top context words	Top URL components
film/staring, film/name film/production companies	movies, bio, person played, actor, did	movies.yahoo.com person, contributor biograph.html
film/name, film/genre film/production companies	movies, yahoo, office, box , new	movies.yahoo.com news, movie imdb.com

Table 3. Two randomly selected samples of entity categories, context words and URL components as shown. Only instances with high confidence from each category are shown.

5.3. Intent Detection

We further test our model for conversational understanding with a quantitative evaluation. The baseline for the evaluation only includes unigram features — each utterance is represented as a set of unigram features with the bag-of-words assumption. Then, we perform an incremental evaluation on our model. On top of the unigram features, we also use the intent distribution estimated from our model as additional features FEAT_1. Since, the conversational dataset does not include any clicked URL information, we simply employ the model with missing values, where the missing values were replaced

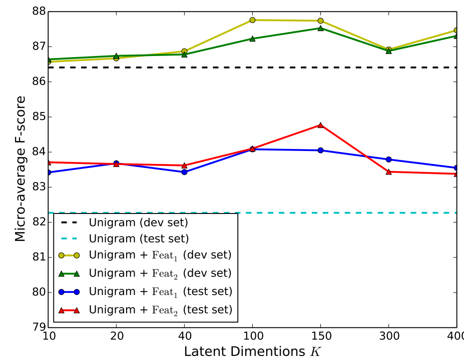

Fig. 2. Experimental results of intent detection on conversational data with different numbers of intents and different feature sets. FEAT_1 includes the output of our model trained on the conversational data, FEAT_2 includes features from FEAT_1 and also the output of our model trained on the query log data.

by a non-informative prior distribution. Moreover, we also train the model with query log dataset and infer the intent distributions on the conversational dataset. By adding the distributions into FEAT_1, we have an augmented feature set FEAT_2. Then, we can further use FEAT_2 with unigrams for intent detection.

The intent detection task in this experiment can be reduced into a multi-class classification problem with 22 categories, such as finding a movie and finding showtimes . Note that, the conversational dataset includes multiple intents and each utterance in this dataset may also have more than one intent. Therefore, the problem here is a multi-class, multi-label classification problem. We build one multi-class multi-label classifier from a binary linear SVM classifier [23] with the "one vs. the rest" rule. The performance is measured with micro-averaged f-score.

We test the performance with different numbers of intent clusters $K \in \{10, 20, 40, 100, 150, 300, 400\}$. All other parameters were chosen to optimize f-score on the development set of conversational data.

The results on the test set of conversational data are summarized in Figure 2. Our model strongly outperforms the unigram baseline (F=82.3%). The additional features FEAT_2 lead to a further improvement when $K \in \{10, 20, 40, 100, 150\}$. The best performance on the development set happens at $K = 100$ for both types of features. As shown in Figure 2, it is also the best number of intents for the test set. Overall, on the test set, we get an F-measure of 84.1% and 84.8% for FEAT_1 and FEAT_2, respectively, which are both better than the baseline. We also tested our model with bigram features, and obtained an F-measure of 83.7%. Adding FEAT_1 features to the bigram baseline resulted in similar performance with the unigram features and FEAT_1, the performance on the test set, and resulted in 84.1% F-measure.

6. CONCLUSIONS

We described a variational Bayesian approach for modeling latent intents of user queries and URLs that they clicked on when available. Our new models separates the contribution of context words and entities, and tokenizes URL components, and models URLs as a bag of URL-tokens. We used this model to enhance supervised intent classification of user queries from conversational interactions. Our experimental results demonstrated the effectiveness of this approach for supervised intent classification, showing further improvements when a large number of unlabeled search queries are used.

7. REFERENCES

[1] Gokhan Tur and Renato De Mori, Eds., *Spoken Language Understanding: Systems for Extracting Semantic Information from Speech*, John Wiley and Sons, New York, NY, 2011.

[2] A. L. Gorin, G. Riccardi, and J. H. Wright, "How May I Help You? ," *Speech Communication*, vol. 23, pp. 113–127, 1997.

[3] J. Chu-Carroll and B. Carpenter, "Vector-Based Natural Language Call Routing," *Computational Linguistics*, vol. 25, no. 3, pp. 361–388, 1999.

[4] P. Natarajan, R. Prasad, B. Suhm, and D. McCarthy, "Speech Enabled Natural Language Call Routing: BBN Call Director," in *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, Denver, CO, September 2002.

[5] P. Haffner, G. Tur, and J. Wright, "Optimizing SVMs for Complex Call Classification," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Hong Kong, April 2003.

[6] C. Chelba, M. Mahajan, and A. Acero, "Speech Utterance Classification," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Hong Kong, May 2003.

[7] Stephen Cox, "Discriminative Techniques in Call Routing," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Hong Kong, April 2003.

[8] Ryuichiro Higashinaka, Katsuhito Sudoh, and Mikio Nakano, "Incorporating discourse features into confidence scoring of intention recognition results in spoken dialogue systems," *Speech Communication*, vol. 48, no. 3, pp. 417–436, 2006.

[9] Andrei Broder, "A Taxonomy of Web Search," in *ACM SIGIR Forum*, 2002, pp. 3–10.

[10] Larry Heck and Dilek Hakkani-Tür, "Exploiting the Semantic Web for Unsupervised Spoken Language Understanding," in *Proceedings of the IEEE Workshop on Spoken Language Technology (SLT 2012)*, Miami, Florida, 2012.

[11] Gokhan Tur, Minwoo Jeong, Ye-Yi Wang, Dilek Hakkani-Tür, and Larry Heck, "Exploiting semantic web for unsupervised statistical natural language semantic parsing," in *Proceedings of Interspeech 2012*, Portland, Oregon, 2012.

[12] Dilek Hakkani-Tür, Asli Celikyilmaz, Larry Heck, and Gokhan Tur, "A Weakly-Supervised Approach for Discovering New User Intents from Search Query Logs," in *Proceedings of Interspeech 2013*, Lyon, France, 2013.

[13] D. Hakkani-Tür L. Heck and G. Tur, "Leveraging knowledge graphs for web-scale unsupervised semantic parsing," in *Proceedings of Interspeech, International Speech Communication Association*, 2013.

[14] M. Mintz, S. Bills, R. Snow, and D. Jurafsky, "Distant supervision for relation extraction without labeled data," in *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL*, 2009.

[15] D. Hakkani-Tür, L. Heck, and G. Tur, "Using a Knowledge Graph and Query Click Logs for U nsupervised Learning of Relation Detection," in *Proceedings of the ICASSP*, 2013.

[16] Thomas Lin, Patrick Pantel, Michael Gamon, Anitha Kannan, and Ariel Fuxman, "Active Objects: Actions for Entity-Centric Search," in *Proceedings of World Wide Web Conference (WWW-12)*, Lyon, France, 2012, pp. 589–598.

[17] Xiao Li, Ye-Yi Wang, and Alex Acero, "Learning query intent from regularized click graphs," in *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2008, pp. 339–346.

[18] Filip Radlinski, Martin Szummer, and Nick Craswell, "Inferring query intent from reformulations and clicks," in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 1171–1172.

[19] Andrew Gelman, John Carlin, Hal Stern, and Donald Rubin, *Bayesian Data Analysis*, Chapman & Hall/CRC, 2004.

[20] D.M. Blei, A.Y. Ng, and M.I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research (JMLR)*, vol. 3, pp. 993–1022, 2003.

[21] J. Polifroni and F. Mairesse, "Using Latent Topic Features for Named Entity Extraction in Search Queries," in *Proceedings of Interspeech 2011*, 2011.

[22] J. Berant, A. Chou, R. Frostig, and P. Liang, "Semantic parsing on Freebase from question-answer pairs," in *Empirical Methods in Natural Language Processing (EMNLP)*, 2013.

[23] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin, "LIBLINEAR: A Library for Large Linear Classification.," *Journal of Machine Learning Research (JMLR)*, vol. 9, pp. 1871–1874, 2008.