A Composite Kernel Approach for Dialog Topic Tracking with Structured Domain Knowledge from Wikipedia

Seokhwan Kim, Rafael E. Banchs, Haizhou Li
Human Language Technology Department
Institute for Infocomm Research
Singapore 138632
{kims,rembanchs,hli}@i2r.a-star.edu.sg

Abstract

Dialog topic tracking aims at analyzing and maintaining topic transitions in ongoing dialogs. This paper proposes a composite kernel approach for dialog topic tracking to utilize various types of domain knowledge obtained from Wikipedia. Two kernels are defined based on history sequences and context trees constructed based on the extracted features. The experimental results show that our composite kernel approach can significantly improve the performances of topic tracking in mixed-initiative human-human dialogs.

1 Introduction

Human communications in real world situations interlace multiple topics which are related to each other in conversational contexts. This fact suggests that a dialog system should be also capable of conducting multi-topic conversations with users to provide them a more natural interaction with the system. However, the majority of previous work on dialog interfaces has focused on dealing with only a single target task. Although some multi-task dialog systems have been proposed (Lin et al., 1999; Ikeda et al., 2008; Celikyilmaz et al., 2011), they have aimed at just choosing the most probable one for each input from the sub-systems, each of which is independently operated from others.

To analyze and maintain dialog topics from a more systematic perspective in a given dialog flow, some researchers (Nakata et al., 2002; Lagus and Kuusisto, 2002; Adams and Martell, 2008) have considered this dialog topic identification as a separate sub-problem of dialog management and attempted to solve it with text categorization approaches for the recognized utterances in a given turn. The major obstacle to the success of these approaches results from the differences between written texts and spoken utterances. In most text categorization tasks, the proper category for each textual unit can be assigned based only on its own content. However, the dialog topic at each turn can be determined not only by the user's intentions captured from the given utterances, but also by the system's decisions for dialog management purposes. Thus, the text categorization approaches can only be effective for the user-initiative cases when users tend to mention the topic-related expressions explicitly in their utterances.

The other direction of dialog topic tracking approaches made use of external knowledge sources including domain models (Roy and Subramaniam, 2006), heuristics (Young et al., 2007), and agendas (Bohus and Rudnicky, 2003; Lee et al., 2008). These knowledge-based methods have an advantage of dealing with system-initiative dialogs, because dialog flows can be controlled by the system based on given resources. However, this aspect can limit the flexibility to handle the user's responses which are contradictory to the system's suggestions. Moreover, these approaches face cost problems for building a sufficient amount of resources to cover broad states of complex dialogs, because these resources should be manually prepared by human experts for each specific domain.

In this paper, we propose a composite kernel to explore various types of information obtained from Wikipedia for mixed-initiative dialog topic tracking without significant costs for building resources. Composite kernels have been successfully applied to improve the performances in other NLP problems (Zhao and Grishman, 2005; Zhang et al., 2006) by integrating multiple individual kernels, which aim to overcome the errors occurring at one level by information from other levels. Our composite kernel consists of a history sequence and a domain context tree kernels, both of which are composed based on similar textual units in Wikipedia articles to a given dialog context.

Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers), pages 19–23,
Baltimore, Maryland, USA, June 23-25 2014.

t	Speaker	Utterance	Topic Transition
0	Guide	How can I help you?	NONE→NONE
1	Tourist	Can you recommend some good places to visit in Singapore?	NONE→ATTR
	Guide	Well if you like to visit an icon of Singapore, Merlion park will be a nice place to visit.	
2	Tourist	Merlion is a symbol for Singapore, right?	ATTR→ATTR
	Guide	Yes, we use that to symbolise Singapore.	
3	Tourist	Okay.	ATTR→ATTR
	Guide	The lion head symbolised the founding of the island and the fish body just symbolised the humble fishing village.	
4	Tourist	How can I get there from Orchard Road?	ATTR→TRSP
	Guide	You can take the north-south line train from Orchard Road and stop at Raffles Place station.	
5	Tourist	Is this walking distance from the station to the destination?	TRSP→TRSP
	Guide	Yes, it'll take only ten minutes on foot.	
6	Tourist	Alright.	TRSP→FOOD
	Guide	Well, you can also enjoy some seafoods at the riverside near the place.	
7	Tourist	What food do you have any recommendations to try there?	FOOD→FOOD
	Guide	If you like spicy foods, you must try chilli crab which is one of our favourite dishes here in Singapore.	
8	Tourist	Great! I'll try that.	FOOD→FOOD

Figure 1: Examples of dialog topic tracking on Singapore tour guide dialogs

2 Dialog Topic Tracking

Dialog topic tracking can be considered as a classification problem to detect topic transitions. The most probable pair of topics at just before and after each turn is predicted by the following classifier: $f(x_t) = (y_{t-1}, y_t)$, where x_t contains the input features obtained at a turn t, $y_t \in C$, and C is a closed set of topic categories. If a topic transition occurs at t, y_t should be different from y_{t-1}. Otherwise, both y_t and y_{t-1} have the same value.

Figure 1 shows an example of dialog topic tracking in a given dialog fragment on Singapore tour guide domain between a tourist and a guide. This conversation is divided into three segments, since f detects three topic transitions at t_1, t_4 and t_6. Then, a topic sequence of 'Attraction', 'Transportation', and 'Food' is obtained from the results.

3 Wikipedia-based Composite Kernel for Dialog Topic Tracking

The classifier f can be built on the training examples annotated with topic labels using supervised machine learning techniques. Although some fundamental features extracted from the utterances mentioned at a given turn or in a certain number of previous turns can be used for training the model, this information obtained solely from an ongoing dialog is not sufficient to identify not only user-initiative, but also system-initiative topic transitions.

To overcome this limitation, we propose to leverage on Wikipedia as an external knowledge source that can be obtained without significant effort toward building resources for topic tracking. Recently, some researchers (Wilcock, 2012; Breuing et al., 2011) have shown the feasibility of using Wikipedia knowledge to build dialog systems. While each of these studies mainly focuses only on a single type of information including category relatedness or hyperlink connectedness, this work aims at incorporating various knowledge obtained from Wikipedia into the model using a composite kernel method.

Our composite kernel consists of two different kernels: a history sequence kernel and a domain context tree kernel. Both represent the current dialog context at a given turn with a set of relevant Wikipedia paragraphs which are selected based on the cosine similarity between the term vectors of the recently mentioned utterances and each paragraph in the Wikipedia collection as follows:

$$\text{sim}\,(x, p_i) = \frac{\phi(x) \cdot \phi(p_i)}{|\phi(x)||\phi(p_i)|},$$

where x is the input, p_i is the i-th paragraph in the Wikipedia collection, $\phi(p_i)$ is the term vector extracted from p_i. The term vector for the input x, $\phi(x)$, is computed by accumulating the weights in the previous turns as follows:

$$\phi(x) = \left(\alpha_1, \alpha_2, \cdots, \alpha_{|W|}\right) \in R^{|W|},$$

where $\alpha_i = \sum_{j=0}^{h} \left(\lambda^j \cdot tfidf(w_i, u_{(t-j)})\right)$, u_t is the utterance mentioned in a turn t, $tfidf(w_i, u_t)$ is the product of term frequency of a word w_i in u_t and inverse document frequency of w_i, λ is a decay factor for giving more importance to more recent turns, $|W|$ is the size of word dictionary, and h is the number of previous turns considered as dialog history features.

After computing this relatedness between the current dialog context and every paragraph in the Wikipedia collection, two kernel structures are constructed using the information obtained from the highly-ranked paragraphs in the Wikipedia.

3.1 History Sequence Kernel

The first structure to be constructed for our composite kernel is a sequence of the most similar paragraph IDs of each turn from the beginning of the session to the current turn. Formally, the sequence S at a given turn t is defined as:

$$S = (s_0, \cdots, s_t),$$

where $s_j = \text{argmax}_i\,(\text{sim}\,(x_j, p_i))$.

Since our hypothesis is that the more similar the dialog histories of the two inputs are, the more similar aspects of topic transtions occur for them, we propose a sub-sequence kernel (Lodhi et al., 2002) to map the data into a new feature space defined based on the similarity of each pair of history sequences as follows:

$$K_s(S_1, S_2) = \sum_{u \in \mathcal{A}^n} \sum_{\mathbf{i}: u = S_1[\mathbf{i}]} \sum_{\mathbf{j}: u = S_2[\mathbf{j}]} \lambda^{l(\mathbf{i})+l(\mathbf{j})},$$

where $\mathcal{A}$ is a finite set of paragraph IDs, S is a finite sequence of paragraph IDs, u is a subsequence of S, $S[\mathbf{j}]$ is the subsequence with the i-th characters $\forall i \in \mathbf{j}$, $l(\mathbf{i})$ is the length of the subsequence, and $\lambda \in (0, 1)$ is a decay factor.

3.2 Domain Context Tree Kernel

The other kernel incorporates more various types of domain knowledge obtained from Wikipedia into the feature space. In this method, each instance is encoded in a tree structure constructed following the rules in Figure 2. The root node of a tree has few children, each of which is a subtree rooted at each paragraph node in:

$$\mathcal{P}_t = \{p_i | \text{sim}\,(x_t, p_i) > \theta\},$$

where θ is a threshold value to select the relevant paragraphs. Each subtree consists of a set of features from a given paragraph in the Wikipedia collection in a hierarchical structure. Figure 3 shows an example of a constructed tree.

Since this constructed tree structure represents semantic, discourse, and structural information extracted from the similar Wikipedia paragraphs to each given instance, we can explore these more enriched features to build the topic tracking model using a subset tree kernel (Collins and Duffy, 2002) which computes the similarity between each pair of trees in the feature space as follows:

$$K_t(T_1, T_2) = \sum_{n_1 \in N_{T_1}} \sum_{n_2 \in N_{T_2}} \triangle\,(n_1, n_2)\,,$$

where N_T is the set of T's nodes, $\triangle\,(n_1, n_2) = \sum_i I_i\,(n_i) \cdot I_i\,(n_2)$, and $I_i(n)$ is a function that is 1 iff the i-th tree fragment occurs with root at node n and 0 otherwise.

3.3 Kernel Composition

In this work, a composite kernel is defined by combining the individual kernels including history sequence and domain context tree kernels, as well as

```
<TREE>:=(ROOT <PAR>...<PAR>)
 <PAR>:=(PAR_ID <PARENTS>
              <PREV_PAR><NEXT_PAR><LINKS>)
  <PARENTS>:=('PARENTS' <ART><SEC>)
   <ART>:=(ART_ID <ART_NAME><CAT_LIST>)
    <ART_NAME>:=('ART_NAME' ART_NAME)
    <CAT_LIST>:=('CAT' <CAT>...<CAT>)
     <CAT>:=(CAT_ID *)
   <SEC>:=(SEC_ID <SEC_NAME><PARENT_SEC>
                  <PREV_SEC><NEXT_SEC>)
    <SEC_NAME>:=('SEC_NAME' SEC_NAME)
    <PARENT_SEC>:=('PRN_SEC', PRN_SEC_ID)
    <PREV_SEC>:=('PREV_SEC', PREV_SEC_NAME)
    <NEXT_SEC>:=('NEXT_SEC', NEXT_SEC_NAME)
  <PREV_PAR>:=('PREV_PAR', PREV_PAR_ID)
  <NEXT_PAR>:=('NEXT_PAR', NEXT_PAR_ID)
  <LINKS>:=('LINKS' <LINK>...<LINK>)
   <LINK>:=(LINK_NAME *)
```

Figure 2: Rules for constructing a domain context tree from Wikipedia: PAR, ART, SEC, and CAT are acronyms for paragraph, article, section, and category, respectively

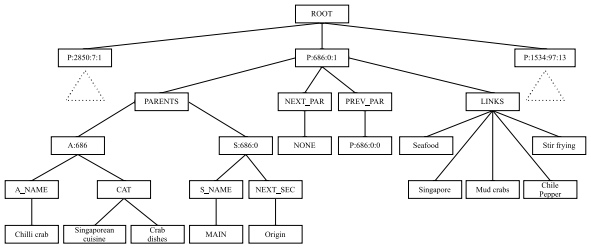

Figure 3: An example of domain context tree

the linear kernel between the vectors representing fundamental features extracted from the utterances themselves and the results of linguistic preprocessors. The composition is performed by linear combination as follows:

$$K(x_1, x_2) = \alpha \cdot K_l(V_1, V_2) + \beta \cdot K_s(S_1, S_2) + \gamma \cdot K_t(T_1, T_2),$$

where V_i, S_i, and T_i are the feature vector, history sequence, and domain context tree of x_i, respectively, K_l is the linear kernel computed by inner product of the vectors, α, β, and γ are coefficients for linear combination of three kernels, and $\alpha + \beta + \gamma = 1$.

4 Evaluation

To demonstrate the effectiveness of our proposed kernel method for dialog topic tracking, we performed experiments on the Singapore tour guide dialogs which consists of 35 dialog sessions collected from real human-human mixed initiative conversations related to Singapore between guides

and tourists. All the recorded dialogs with the total length of 21 hours were manually transcribed, then these transcribed dialogs with 19,651 utterances were manually annotated with the following nine topic categories: Opening, Closing, Itinerary, Accommodation, Attraction, Food, Transportation, Shopping, and Other.

Since we aim at developing the system which acts as a guide communicating with tourist users, an instance for both training and prediction of topic transition was created for each turn of tourists. The annotation of an instance is a pair of previous and current topics, and the actual number of labels occurred in the dataset is 65.

For each instance, the term vector was generated from the utterances in current user turn, previous system turn, and history turns within the window sizes $h = 10$. Then, the history sequence and tree context structures for our composite kernel were constructed based on 3,155 articles related to Singapore collected from Wikipedia database dump as of February 2013. For the linear kernel baseline, we used the following features: n-gram words, previous system actions, and current user acts which were manually annotated. Finally, 8,318 instances were used for training the model.

We trained the SVM models using SVM^{light} [1] (Joachims, 1999) with the following five different combinations of kernels: K_l only, K_l with $\mathcal{P}$ as features, $K_l + K_s$, $K_l + K_t$, and $K_l + K_s + K_t$. The threshold value θ for selecting $\mathcal{P}$ was 0.5, and the combinations of kernels were performed with the same α, β, or γ coefficient values for all sub-kernels. All the evaluations were done in five-fold cross validation to the manual annotations with two different metrics: one is accuracy of the predicted topic label for every turn, and the other is precision/recall/F-measure for each event of topic transition occurred either in the answer or the predicted result.

Table 1 compares the performances of the five combinations of kernels. When just the paragraph IDs were included as additional features, it failed to improve the performances from the baseline without any external features. However, our proposed kernels using history sequences and domain context trees achieved significant performances improvements for both evaluation metrics. While the history sequence kernel enhanced the coverage of the model to detect topic transitions,

[1] http://svmlight.joachims.org/

	Turn-level	Transition-level		
	Accuracy	P	R	F
K_l	62.45	42.77	24.77	31.37
$K_l + \mathcal{P}$	62.44	42.76	24.77	31.37
$K_l + K_s$	67.19	39.94	40.59	40.26
$K_l + K_t$	68.54	45.55	35.69	40.02
All	69.98	44.82	39.83	42.18

Table 1: Experimental Results

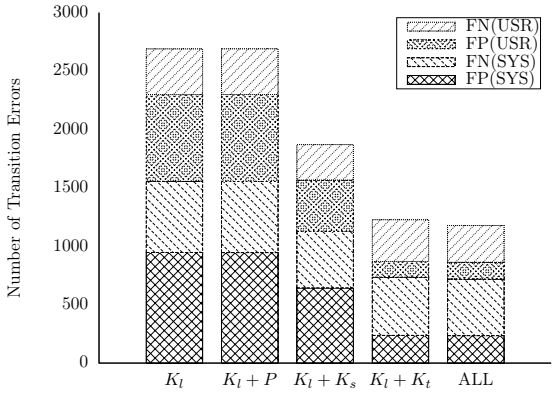

Figure 4: Error distibutions of topic transitions: FN and FP denotes false negative and false positive respectively. USR and SYS in the parentheses indicate the initiativity of the transitions.

the domain context tree kernel contributed to produce more precise outputs. Finally, the model combining all the kernels outperformed the baseline by 7.53% in turn-level accuracy and 10.81% in transition-level F-measure.

The error distributions in Figure 4 indicate that these performance improvements were achieved by resolving the errors not only on user-initiative topic transitions, but also on system-initiative cases, which implies the effectiveness of the structured knowledge from Wikipedia to track the topics in mixed-initiative dialogs.

5 Conclusions

This paper presented a composite kernel approach for dialog topic tracking. This approach aimed to represent various types of domain knowledge obtained from Wikipedia as two structures: history sequences and domain context trees; then incorporate them into the model with kernel methods. Experimental results show that the proposed approaches helped to improve the topic tracking performances in mixed-initiative human-human dialogs with respect to the baseline model.

References

P. H. Adams and C. H. Martell. 2008. Topic detection and extraction in chat. In *Proceedings of the 2008 IEEE International Conference on Semantic Computing*, pages 581–588.

D. Bohus and A. Rudnicky. 2003. Ravenclaw: dialog management using hierarchical task decomposition and an expectation agenda. In *Proceedings of the European Conference on Speech, Communication and Technology*, pages 597–600.

A. Breuing, U. Waltinger, and I. Wachsmuth. 2011. Harvesting wikipedia knowledge to identify topics in ongoing natural language dialogs. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 445–450.

A. Celikyilmaz, D. Hakkani-Tür, and G. Tür. 2011. Approximate inference for domain detection in spoken language understanding. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pages 713–716.

Michael Collins and Nigel Duffy. 2002. New ranking algorithms for parsing and tagging: Kernels over discrete structures, and the voted perceptron. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 263–270.

S. Ikeda, K. Komatani, T. Ogata, H. G. Okuno, and H. G. Okuno. 2008. Extensibility verification of robust domain selection against out-of-grammar utterances in multi-domain spoken dialogue system. In *Proceedings of the 9th INTERSPEECH*, pages 487–490.

T. Joachims. 1999. Making large-scale SVM learning practical. In B. Schölkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, chapter 11, pages 169–184. MIT Press, Cambridge, MA.

K. Lagus and J. Kuusisto. 2002. Topic identification in natural language dialogues using neural networks. In *Proceedings of the 3rd SIGdial workshop on Discourse and dialogue*, pages 95–102.

C. Lee, S. Jung, and G. G. Lee. 2008. Robust dialog management with n-best hypotheses using dialog examples and agenda. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 630–637.

B. Lin, H. Wang, and L. Lee. 1999. A distributed architecture for cooperative spoken dialogue agents with coherent dialogue state and history. In *Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*.

Huma Lodhi, Craig Saunders, John Shawe-Taylor, Nello Cristianini, and Chris Watkins. 2002. Text classification using string kernels. *The Journal of Machine Learning Research*, 2:419–444.

T. Nakata, S. Ando, and A. Okumura. 2002. Topic detection based on dialogue history. In *Proceedings of the 19th international conference on Computational linguistics (COLING)*, pages 1–7.

S. Roy and L. V. Subramaniam. 2006. Automatic generation of domain models for call centers from noisy transcriptions. In *Proceedings of COLING/ACL*, pages 737–744.

G. Wilcock. 2012. Wikitalk: a spoken wikipedia-based open-domain knowledge access system. In *Proceedings of the Workshop on Question Answering for Complex Domains*, page 5770.

S. Young, J. Schatzmann, K. Weilhammer, and H. Ye. 2007. The hidden information state approach to dialog management. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 149–152.

Min Zhang, Jie Zhang, Jian Su, and Guodong Zhou. 2006. A composite kernel to extract relations between entities with both flat and structured features. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 825–832.

Shubin Zhao and Ralph Grishman. 2005. Extracting relations with integrated information using kernel methods. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 419–426.