WIKIPEDIA-BASED KERNELS FOR DIALOGUE TOPIC TRACKING

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
Dialogue topic tracking aims to segment on-going dialogues into topically coherent sub-dialogues and predict the topic category for each next segment. This paper proposes a kernel method for dialogue topic tracking to utilize various types of information obtained from Wikipedia. The experimental results show that our proposed approach can significantly improve the performances of the task in mixed-initiative human-human dialogues.

Index Terms— Spoken Dialogue Systems, Dialogue Topic Tracking, Kernel Methods, Wikipedia

1. INTRODUCTION
During the past few years, spoken dialogue systems have come into the spotlight as a next-generation user interface, since they are based on the most natural way for human-human communications. However, the practical use of these systems is still limited. One of the reasons is that the majority of previous work on spoken dialogue technologies has focused on dealing with a single target task. Although some approaches for handling multi-domain or multi-task dialogues have been proposed [1, 2, 3], these studies have aimed at choosing the most probable system among the candidates, each of which is independently built for an individual task from others. Since human communications in real-world situations consist of a series of multiple topics holding conversational coherence, spoken dialogue systems have to be able to track the topic sequence considering dialogue contexts for more natural conversations.

Some researchers [4, 5, 6] attempted to solve this dialogue topic identification task by considering it as a text categorization problem for the recognized utterances in a 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 only with its contents themselves. However, the dialogue 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 dialogue management purposes. Thus, the effectiveness of these text categorization approaches can be limited only for the user-initiative cases when users tend to mention the topic-related expressions explicitly in their utterances. Furthermore, given that it is impossible to refer to following conversations after each turn in an ongoing dialogue, the dialogue topic should be identified in an online manner only with the already mentioned utterances that are available at the current turn. That is substantially different from the traditional topic detection tasks considering both forward and backward features all over the given text document.

The other direction of dialogue topic tracking approaches made use of external knowledge sources including domain models [7], heuristics [8], and agendas [9, 10]. These knowledge-based methods have the advantage for tracking the topics of system-initiative dialogues, because dialogue flows can be controlled by the system based on given resources. However, this aspect can cause the limited flexibility to handle the user’s responses which are contradictory to the flow suggested based on the resources. Moreover, these approaches necessarily face cost problems for building a sufficient amount of resources to cover broad states of complex dialogues, because these resources should be manually prepared by human experts for each specific domain.

In this paper, we propose a kernel method utilizing various types of information obtained from Wikipedia for dialogue topic tracking. Since huge amount of contents have been already created by collaborative efforts and are freely available from Wikipedia, our proposed approach aims to improve performances of topic tracking in mixed-initiative dialogues without significant costs for building resources.

2. DIALOGUE TOPIC TRACKING
Dialogue topic tracking is divided into two sequential sub-tasks: segmenting a dialogue session into topically coherent sub-dialogues and identifying the next topic category at each time of topic transition, each of which can be conceptually considered to be a classification problem. The first classifier for dialogue segmentation determines whether a topic transi-
Fig. 1. Examples of dialogue topic tracking on Singapore tour guide dialogues

The fundamental goal of kernel methods is to map the data into a higher dimensional feature space with the ability to improve the classification performances. According to the previous work on Wikipedia-based semantic kernels for text classification [14], an extended feature space is defined by concatenating the concept space with the previous term vector space as follows:

\[ \phi'(x) = (\alpha_1, \alpha_2, \cdots, \alpha_{|W|}, \beta_1, \beta_2, \cdots, \beta_{|D|}) \in \mathbb{R}^{|W| + |D|}, \]

where \( \beta_i \) is the semantic relatedness between the input \( x \) and the concept in the \( i \)-th Wikipedia article and \( |D| \) is the number of articles in the Wikipedia collection. The value for \( \beta_i \) is computed with the cosine similarity between term vectors as follows:

\[ \beta_i = \frac{\text{sim}(x, d_i) = \cos(\theta) = \frac{\phi(x) \cdot d_i}{||\phi(x)|||d_i||}}, \]

where \( d_i \) is the term vector composed from the \( i \)-th Wikipedia article in the collection.

Then, each extended vector can be transformed into a new space with \( \phi(x) = \phi'(x)S \), where \( S \) is the transformation matrix defined as follows:

\[ S_{ij} = \begin{cases} 1, & \text{if } i = j \\ s(d_{i-|W|}, d_{j-|W|}), & \text{else if } i > |W| \text{ and } j > |W| \\ 0, & \text{otherwise} \end{cases}, \]

where \( s(d_i, d_j) \) is the relatedness between \( d_i \) and \( d_j \). As illustrated in Figure 2, each value in the concept space is updated by the matrix multiplication as \( \tilde{\beta}_i = \sum_{j=1}^{|D|} s(d_i, d_j) \cdot \beta_j \), which utilizes the relationships between a given concept and all the others in the collection.

This aspect enables the topic tracking models to consider various types of domain knowledge that encodes the relatedness among the given concepts. In this work, each value \( s(d_i, d_j) \) is computed based on the following five types of information derived from Wikipedia: category relatedness, category overlap score, contents similarity, co-occurrence frequency, and geographical closeness.

Every concept in the Wikipedia belongs to one or more categories. Since all categories are organized in hierarchical...
The relatedness between two concepts can be computed based on which both concepts belong. This can be computed by Jacquet’s coefficient as follows:

\[ s_1(d_i, d_j) = \frac{2 \cdot \text{depth}(\text{lcs}(d_i, d_j))}{\text{depth}(d_i) + \text{depth}(d_j)}, \]

where \( \text{depth}(d) \) is the length of the path from the root node to \( d \) and \( \text{lcs}(d_i, d_j) \) is the least common subsumer of the two articles in the category hierarchy.

Another way to obtain the semantic relatedness between two concepts is based on the ratio of common categories to articles in the category hierarchy. The cosine similarity between term vectors extracted from the body texts of the corresponding articles can be considered to indicate the relatedness between two concepts as follows:

\[ s_3(d_i, d_j) = \cos(\theta) = \frac{\phi(d_i) \cdot \phi(d_j)}{||\phi(d_i)|| ||\phi(d_j)||}, \]

where \( \phi(d) \) is the term vector obtained from the body texts of \( d \).

In addition to the above-mentioned values representing semantic relationships, the discourse relatedness can also be obtained from Wikipedia. We assume that the more frequently the mentions about two concepts co-occurred in the Wikipedia articles, the more similar aspects both concepts take in dialogue flows related to them. This co-occurrence frequency is computed by normalized pointwise mutual information as follows:

\[ s_4(d_i, d_j) = \frac{\text{pmi}(d_i, d_j)}{-\log(n(d_i, d_j))} = \frac{\log \left( \frac{n(d_i, d_j)}{n(d_i) \cdot n(d_j)} \right)}{-\log(n(d_i, d_j))}, \]

where \( n(d) \) is the total number of the hyperlinks appeared in the entire Wikipedia collection and \( n(d_i, d_j) \) is the number of the cases that both \( d_i \) and \( d_j \) occur in a same paragraph.

The other type of information is defined especially for the domains which require to deal with the locations of some places, such as tour guide domain. Each Wikipedia concept related to a spatial entity provides its geographic coordinate values. For each pair of spatial concepts, the geographical closeness can be computed by normalizing the geographical distance between them as follows:

\[ s_5(d_i, d_j) = \frac{1.0 - e^{\delta(d_i, d_j) - \theta}}{1.0 + e^{\delta(d_i, d_j) - \theta}}, \]

where \( \delta(d_i, d_j) \) is the geographical distance between \( d_i \) and \( d_j \) and \( \theta \) is a threshold value to adjust the normalized score.

The values of \( s_1, s_2, \) and \( s_5 \) are in the range of \([0,1]\), while \( s_3 \) and \( s_4 \) have the value in the range of \([-1,1]\). For all types of relatedness, the larger the value, the more related two concepts are. The final score \( s(d_i, d_j) \) is computed by linear combinations of these scores as follows:

\[ s(d_i, d_j) = \sum \lambda_k \cdot s_k(d_i, d_j), \]

where \( \sum \lambda_k = 1 \). In this work, we set the same values for all \( \lambda \) weights.

Figure 3 shows the computed scores among four different concepts each of which on a place in Singapore. The pair of ‘Clarke Quay’ and ‘Boat Quay’ has a higher overall score than other pairs, because these two concepts belong to several common categories, contain similar contents in their body texts, co-occurred frequently all over the collection, and located close to each other. Thus, if a feature vector in the original space shows that the input state is relevant to only
one of this concept pair with a high score, the other concept also gain more weight in the transformed feature space.

4. EVALUATION

To demonstrate the effectiveness of our proposed kernel method for dialogue topic tracking, we performed experiments on the Singapore tour guide dialogues which consists of 35 dialogue sessions collected from real human-human mixed initiative conversations related to Singapore between guides and tourists. All the recorded dialogues with the total length of 21 hours were manually transcribed, then these transcribed dialogues with 19,651 utterances were manually segmented into 1,642 topically-meaningful segments and annotated with 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, each instance for segmentation was created for each turn of tourists with the binary label indicating topic transition. Then, the topic prediction instances labeled with next topic categories were also prepared for the tourists’ turns at segment boundaries. The term vector of a given instance was generated from the utterances in current user turn, previous system turn, and history turns within the window sizes \( h = 10 \) for segmentation and \( h = 2 \) for topic prediction. The values for its concept space were computed based on 3,155 articles related to Singapore collected from Wikipedia database dump as of February 2013. Totally, 8,318 and 1,607 instances were used for training the models for user-turn-level segmentation and dialogue-segment-level topic prediction tasks, respectively.

We trained the SVM models for both subtasks using LibSVM [16] with the following seven different kernels. BOW is a baseline model using linear Kernel with bag-of-words. WK0 uses the expanded vector \( \phi'(x) \) without transformation. The other models WK1, \cdots, WK5 are based on our proposed Wikipedia-based kernel method with incrementally combined relatedness scores. The score for each WK\( \nu \) is defined as \( s(d_i,d_j) = \sum_{k=1}^{n} (s_k(d_i,d_j)) / n \). The multi-label classifications for topic prediction were performed with one-against-all strategy. All the evaluations were done in five-fold cross validation to the manual annotations with the metrics of precision, recall, and F-measure for segmentation and accuracy for topic prediction.

Table 1 compares the performances of the seven models in topic segmentation, prediction, and cascaded process of two subtasks. The results indicate that our proposed kernel methods based on Wikipedia achieved significant performances improvements for all the tasks. Especially, WK5, the kernel incorporating all the defined scores, outputted the best results compared to the other models. It obtained higher segmentation performances than the bag-of-words model by 3.16 in F-measures; and it also outperformed the baseline in topic prediction by 28.16% in accuracy. These improved models in both separate subtasks finally led to better results in the overall topic prediction process performed by cascading of two subtasks. The difference in cascaded performances between the baseline and the final model was 8.04% in accuracy.

However, these improved performances still do not seem to be high enough for practical uses in dialogue systems. To investigate the reason for these poor performances, we analyzed the errors on the cascaded process with the final model. The distributions of errors in Table 2 show that 71.4% of errors resulted from segmentation and 60.0% of errors occurred for the topic shifts initiated by systems; and as many as 42.1% of the errors belonged to the intersection of these two categories. It suggests that the detection of system-initiative topic transitions is critical for dialogue topic tracking.

5. CONCLUSIONS

This paper presented a Wikipedia-based kernel method for dialogue topic tracking. This approach aimed to incorporate various types of information obtained from Wikipedia into the models. Experimental results show that the proposed Wikipedia kernels helped to improve both segmentation and prediction performances in mixed-initiative dialogues than the baseline model.

However, we expect that our methods can be further improved in future work. First, we can consider other ways of determining the parameters which were manually assigned in this work. If we discover much more optimized parameters, they can raise the performances. The other direction of our future work is to derive more various types of knowledge from Wikipedia for dialogue topic tracking. We plan to investigate
what kinds of additional features in Wikipedia can contribute to improve the performances especially in segmentation of system-initiative topic transitions.

6. REFERENCES


