TRACKING DIALOG STATES USING AN AUTHOR-TOPIC BASED REPRESENTATION

Richard Dufour, Mohamed Morchid, Titouan Parcollet

LIA - University of Avignon (France)
`{firstname.lastname}@univ-avignon.fr`

ABSTRACT

Automatically translating textual documents from one language to another inevitably results in translation errors. In addition to language specificities, this automatic translation appears more difficult in the context of spoken dialogues since, for example, the language register is far from "clean speech". Speech analytics suffer from these translation errors. To tackle this difficulty, a solution consists in mapping translations into a space of hidden topics. In the classical topic-based representation obtained from a Latent Dirichlet Allocation (LDA), distribution of words into each topic is estimated automatically. Nonetheless, the targeted classes are ignored in the particular context of a classification task. In the DSTC5 main task, this targeted class information is crucial, the main objective being to track dialog states for sub-dialog segments. For this challenge, we propose to apply an original topic-based representation for each sub-dialogue based not only on the sub-dialogue content itself (words), but also on the dialogue state related to the sub-dialogue. This original representation is based on the Author-Topic (AT) model, previously successfully applied on a different classification task. Promising results confirmed the interest of such a method, the AT model reaching performance slightly better in terms of F-measure than baseline ones given by the task's organizers.

Index Terms— Author-Topic Model, Dialog State Tracking, Sub-dialog Level

1. INTRODUCTION

Spoken dialogues are a particular case of human/human interactions where automatic processing encounters many difficulties, especially due to the language register, far from a "clean" speech (ungrammaticality, disfluencies, word repetitions, specific vocabulary...). Many studies and applications have been proposed in recent years, such as theme identification in dialogues [1], dialogue strategy learning [2].... One of them concerns the tracking of dialogue states. The Dialog State Tracking Challenge (DSTC) was proposed in the idea

of quickly developing a set of solutions from worldwide research teams on a similar task and a similar dataset. In the fifth edition (DSTC5) [3], the main objective of the challenge is to track dialog states for sub-dialog segments. For each turn in a given sub-dialog, depending on the history of the dialogue prior to the turn, the proposed solution must provide an hypothesis of slot-value pairs (dialog state).

In this particular DSTC5, one of the major difficulty concerns the automatic translations of dialogues from Chinese to English language in the development and test sets (more information about the task and its particularities can be found in Section 4). This automatic process inevitably leads to translation errors that may negatively affect the systems using them directly.

To tackle these translations errors, an efficient way could be to map dialogues into a topic space abstracting the translation outputs. As a result, instead of directly considering the words to track dialogue states, it will be achieved in this topic space. This type of approaches demonstrated its interest in the past in many applications. Latent Dirichlet Allocation (LDA) [4] was then largely used in speech analytics applications [1]. During the LDA learning process, distribution of words into each topic is estimated automatically. Nonetheless, the class (*i.e.* the dialogue state or slot-value pair in the DSTC5 main task) associated to the sub-dialogue is not directly taken into account in the topic model. Indeed, the classes are usually only used to train a classifier at the end of the process. Finally, such a system considers separately the document content (*i.e.* words), to learn a topic model, and the labels (*i.e.* dialogue states) to train a classifier. We can however note that, in the considered application, sub-dialogues are labeled by a human annotator: a relation between the document content (words) and the document label (class) should then exist.

In the context of dialogue state tracking, this relation is crucial to efficiently label unseen (*i.e.* new) sub-dialogues. This model (LDA) needs to infer an unseen document into each topic space. The processing time during the inference process as well as the difficult choice of an efficient number of iterations, do not allow us to evaluate effectively and quickly the best dialogue state related to a sub-dialog level. In this challenge, we propose to use a topic model, called

Author-Topic (AT) model [5], that takes into consideration all information contained into a sub-dialogue: the content itself (*i.e.* words), the label (*i.e.* dialogue state) and the relation between the distribution of words into the dialogue and the label, considered as a latent relation. From this model, a vector representation in a continuous space is built for each dialogue. Then, a supervised classification approach, based on Support Vector Machines (SVM) [6], is applied. This approach was previously successfully applied to a similar speech analytics task, where the purpose was to identify the main theme appearing in a conversation automatically transcribed using an Automatic Speech Recognition (ASR) system [7].

The rest of the paper is organized as follows. We present previous works about dialog state tracking and topic modeling in Section 2. A description of our proposed approach is given in Section 3. The experimental protocol is briefly described in Section 4 while Section 5 presents and discusses the obtained results on the main task of the DSTC5 challenge. Finally, a conclusion including perspectives are proposed in Section 6.

2. RELATED WORKS

Dialogues, automatically translated in this particular DSTC5 task, may contain many errors due to language specificities. Directly using the word content may then lead to poor performance level. An elegant way to tackle these potential errors is to map dialogues in a thematic space in order to abstract the document content. Several approaches proposed to consider the document as a mixture of latent topics. These methods, such as Latent Semantic Analysis (LSA) [8, 9], Probabilistic LSA (PLSA) [10] or Latent Dirichlet Allocation (LDA) [4], build a higher-level representation of the document in a topic space. Documents (here dialogues' transcription) are then considered as a *bag-of-words* [11] where the word order is not taken into account.

The most known and used topic modelling approach is the Latent Dirichlet Allocation (LDA) [4] model. The LDA approach represents dialogues (or documents) as a mixture of latent topics. These methods demonstrated their performance on various tasks, such as sentence [12] or keyword [13] extraction. In opposition to a multinomial mixture model, LDA considers that a theme is associated to each occurrence of a word composing the document, rather than associate a topic with the complete document. Thereby, a document can change of topics from a word to another. However, the word occurrences are connected by a latent variable which controls the global respect of the distribution of the topics in the document. These latent topics are characterized by a distribution of word probabilities which are associated with them. PLSA and LDA models have been shown to generally outperform LSA on IR tasks [14]. Moreover, LDA provides a direct estimate of the relevance of a topic knowing a word set.

The generative process corresponds to the hierarchical Bayesian model shown, using plate notation, in Figure 1

(a). Several techniques, such as Variational Methods [4], Expectation-propagation [15] or Gibbs Sampling [16], have been proposed to estimate the parameters describing a LDA hidden space. Gibbs Sampling is a special case of Markov-chain Monte Carlo (MCMC) [17] and gives a simple algorithm for approximate inference in high-dimensional models such as LDA [18]. This overcomes the difficulty to directly and exactly estimate parameters that maximize the likelihood of the whole data collection defined as:

$$P(W|\overrightarrow{\alpha}, \overrightarrow{\beta}) = \prod_{w \in W} P(\overrightarrow{w}|\overrightarrow{\alpha}, \overrightarrow{\beta}) \tag{1}$$

for the whole data collection W knowing the Dirichlet parameters $\overrightarrow{\alpha}$ and $\overrightarrow{\beta}$.
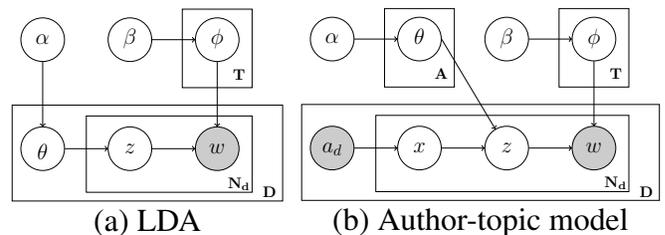

Fig. 1. Generative models in plate notation for Latent Dirichlet Allocation (LDA) (a) and Author-Topic (AT) (b) models.

The Gibbs Sampling, to estimate LDA, was firstly reported in [16]. A more comprehensive description of this method can be found in [18]. One can refer to these papers for a better understanding of this sampling technique. This method is used both to estimate the LDA parameters and to infer an unseen dialogue with a hidden space of T topics.

In the LDA technique, the topic z is drawn from a multinomial over θ which is drawn from a Dirichlet($\overrightarrow{\alpha}$). Thus, a set of p topic spaces are learned using LDA by varying the number of topics T to obtain p topic spaces of size T.

Gibbs Sampling allows one to both estimate the LDA parameters, in order to represent a new dialogue d with the rth topic space of size T, and obtain a feature vector $V_d^{z^r}$ of the topic representation of d. The jth feature:

$$V_d^{z_j^r} = \theta_{j,d}^r , \tag{2}$$

where $\theta_{j,d}^r = P(z_j^r|d)$ is the probability of topic z_j^r ($1 \leq j \leq T$) generated by the unseen dialogue d in the rth topic space of size T as described in Figure 2.

Example of a dialogue (identification of the major theme of a dialogue [7]) mapped into a topic space from LDA model is presented in Figure 2.

In the context of dialogue state tracking, different approaches have been proposed, such as Markovian discriminative modeling [19], sequence modelling with Conditional Random Fields (CRF) method [20], rule-based approaches [21]... More recently, deep neural network methods
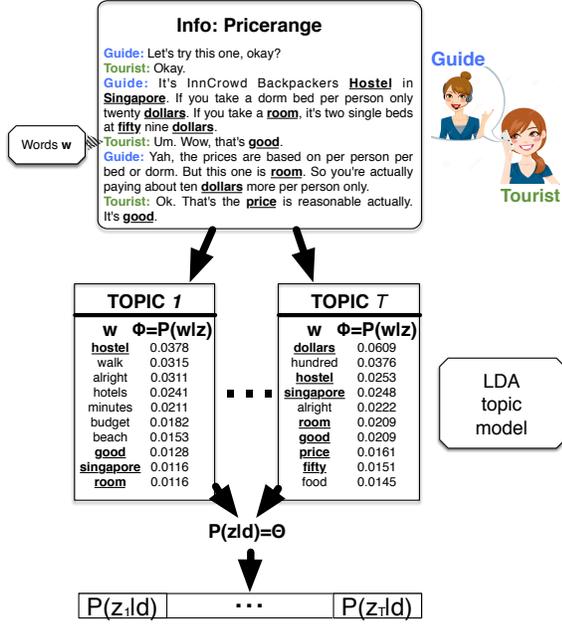
545

Figure 2 (left)

Info: Pricerange

Guide: Let's try this one, okay?
Tourist: Okay.
Guide: It's InnCrowd Backpackers **Hostel** in **Singapore**. If you take a dorm bed per person only twenty **dollars**. If you take a **room**, it's two single beds at **fifty** nine **dollars**.
Tourist: Um. Wow, that's **good**.
Guide: Yah, the prices are based on per person per bed or dorm. But this one is **room**. So you're actually paying about ten **dollars** more per person only.
Tourist: Ok. That's the **price** is reasonable actually. It's **good**.

Words **w**

Guide

Tourist

TOPIC 1		TOPIC T	
w	Φ=P(w|z)	w	Φ=P(w|z)
hostel	0.0378	dollars	0.0609
walk	0.0315	hundred	0.0376
alright	0.0311	hostel	0.0253
hotels	0.0241	singapore	0.0248
minutes	0.0211	alright	0.0222
budget	0.0182	room	0.0209
beach	0.0153	good	0.0209
good	0.0128	price	0.0161
singapore	0.0116	fifty	0.0151
room	0.0116	food	0.0145

LDA topic model

$P(z|d)=\Theta$

$P(z_1|d)$	$\cdots$	$P(z_T|d)$

Fig. 2. Example of a dialogue d from the DSTC5 corpus [3] mapped into a topic space of size n [7]. The the slot-value pair (or dialogue state) labeled to the dialogue is **INFO: Pricerange**

.

have gained a lot of attention [22, 23, 24]. From our knowledge, this is the first attempt of using a topic-based approach that includes relations between words and dialogue states (slot-value pairs) themselves using the Author-Topic (AT) model for tracking dialog states.

3. PROPOSED APPROACH

3.1. Author-Topic (AT) model

While many topic modelling approaches have been proposed, as reported in Section 2, these models do not code statistical relations between words contained into the document (here the dialogue transcription), and the label (here the dialogue state) that could be associated to it.

To go beyond this limit, the Author-Topic (AT) model has been proposed [5]. The AT model links both authors (here, the dialogue state) and documents content (words). The AT model, represented into its plate notation in Figure 1 (b), uses a topic-based representation to model both the document content (words distribution) and the authors (authors distribution). For each word w contained into a document d, an author a is uniformly chosen at random. Then, a topic z is chosen from a distribution over topics specific to that author, and the word is generated from the chosen topic.

In our considered DSTC5 dialogue state tracking task, a

document d is a sub-dialogue session on touristic information for Singapore collected from Skype calls between tour guides and tourists. Each sub-dialogue part has been manually annotated with a dialogue state (slot-value pairs), considered here as an author. Thus, each sub-dialogue d is composed with a set of words w and a dialogue state a. In this model, x indicates the author (*i.e.* the dialogue state) responsible for a given word, chosen from a_d. Each author is associated with a distribution over topics (θ), chosen from a symmetric Dirichlet prior ($\overrightarrow{\alpha}$), and a weighted mixture to select a topic z. A word is then generated according to the distribution ϕ corresponding to the topic z. This distribution ϕ is drawn from a Dirichlet ($\overrightarrow{\beta}$).

Figure 3 (right)

Author **a**

Words **w**

Info: Pricerange

Guide: Let's try this one, okay?
Tourist: Okay.
Guide: It's InnCrowd Backpackers **Hostel** in **Singapore**. If you take a dorm bed per person only twenty **dollars**. If you take a **room**, it's two single beds at **fifty** nine **dollars**.
Tourist: Um. Wow, that's **good**.
Guide: Yah, the prices are based on per person per bed or dorm. But this one is **room**. So you're actually paying about ten **dollars** more per person only.
Tourist: Ok. That's the **price** is reasonable actually. It's **good**.

Guide

Tourist

TOPIC 1		TOPIC T	
w	Φ=P(w|z)	w	Φ=P(w|z)
hostel	0.0378	dollars	0.0609
walk	0.0315	hundred	0.0376
alright	0.0311	hostel	0.0253
hotels	0.0241	singapore	0.0248
minutes	0.0211	alright	0.0222
budget	0.0182	room	0.0209
beach	0.0153	good	0.0209
good	0.0128	price	0.0161
singapore	0.0116	fifty	0.0151
room	0.0116	food	0.0145
a	Θ=P(a|z)	a	Θ=P(a|z)
Info: Exhibit	0.0399	Info: Facility	0.3613
Info: Delivery	0.0285	Info: Distance	0.0402
Info: Pricerange	0.0119	Place: Gallery Hotel	0.0258
Info: Room Type	0.0119	Type Place: Budget Hotel	0.0254
Neigh.: North	0.0049	Activity: Doctor Fish	0.0094
Type Place: Fast Food	0.0014	Info: Package	0.0088
Type Place: Mid Range Hotel	0.0012	Info: Preference	0.0057
…	…	…	…

Author-Topic model

$P(a|d)=\Sigma\Sigma\Theta\Phi$

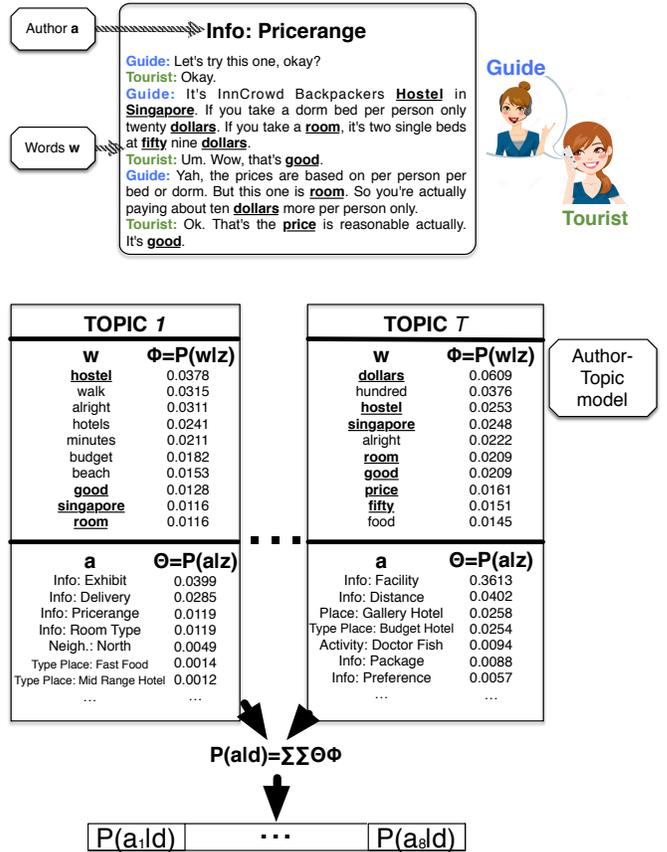| $P(a_1|d)$	$\cdots$	$P(a_8|d)$

Fig. 3. Example of a dialogue d from the DSTC5 corpus [3] mapped into Author-Topic (AT) model of size n [7]. The slot-value pair (or dialogue state) labeled to the dialogue is **INFO: Pricerange**. A slot-value pair is considered here as an author a.

The parameters ϕ and θ are estimated from a straightforward algorithm based on Gibbs Sampling such as LDA hyper-parameters estimation method. One can find more about Gibbs Sampling and author topic model in [5].

Figure 3 shows an example of the mapping process of a

made-up unseen dialogue d to identify the major theme of a conversation [7], into an author topic space of size T. Each sub-dialogue d is composed with a set of words w and a label (or dialogue state) a considered as the author in the AT model. Thus, this model allows one to code statistical dependencies between dialogue content (words w) and label (dialogue state a) through the distribution of the latent topics into the dialogue.

Gibbs Sampling allows us to estimate the AT model parameters, in order to represent an unseen dialogue d with the rth author topic space of size T, and to obtain a feature vector $V_d^{a_k^r} = P(a_k|d)$ of the topic representation an unseen dialogue d with the rth author topic space Δ_r^n of size T. The kth ($1 \leq k \leq A$) feature is:

$$V_d^{a_k^r} = \sum_{i=1}^{N_d} \sum_{j=1}^{T} \theta_{j,a_k}^r \phi_{j,i}^r \qquad (3)$$

where A is the number of authors (or dialogue states); $\theta_{j,a_k}^r = P(a_k|z_j^r)$ is the probability of author a_k to be generated by the topic z_j^r ($1 \leq j \leq T$) in the rth topic space of size T. $\phi_{j,i}^r = P(w_i|z_j^r)$ is the probability of the word w_i (N_d is the vocabulary size of d) to be generated by the topic z_j^r.

This original AT model has already successfully been applied on a similar task, focusing on identifying the major theme of a dialogue conversation [7].

3.2. SVM classification

A classification approach based on Support Vector Machines (SVM) is performed to find out the dialog state (slot-value pair) of a given sub-dialogue segment. As this classification task requires a multi-class classifier, the SVM *one-against-one* method is chosen with a linear kernel. This method gives a better accuracy than the *one-against-rest* [25]. In this multi-class problem, A denotes the number of slot-value pairs (dialog states) and $t_i, i = 1, \ldots, A$ denotes the A slot-value pairs. A binary classifier is used with a linear kernel for every pair of distinct slot-value pair. As a result, binary classifiers $A(A-1)/2$ are constructed all together. The binary classifier $C_{i,j}$ is trained from example data where t_i is a positive class and t_j a negative one ($i \neq j$).

For a vector representation of an unseen sub-dialogue segment d ($V_d^{a_k^r}$ for the AT model representation), if $C_{i,j}$ means that d is in the slot-value pair t_i, then the vote for the class t_i is added by one. Otherwise, the vote for the slot-value pair t_j is increased by one. After the vote of all classifiers, the sub-dialogue segment d is assigned to the slot-value pair having the highest number of votes.

4. EXPERIMENTAL PROTOCOL OF THE DSTC5 MAIN TASK

4.1. General overview

The goal of the main task of the fifth Dialog State Tracking Challenge (DSTC5) [3] is to track dialog states for sub-dialog segments in conversations between a guide and a tourist. Participants will then use the TourSG corpus provided by the DSTC5 organizers. The main task of the challenge is the one considered here by our LIA team. For each turn (or segment) in a given sub-dialog, the proposed solution must provide an hypothesis of slot-value pairs depending on the history of the dialogue prior to the turn. Then, for each segment of the considered sub-dialogue, an hypothesis can only be made by considering its "history" (and not the "future"). For example, if a sub-dialogue contains 10 segments, and a decision has to be made for the fifth, only the five first segments can be used to make a decision. Table 1 presents an example of a slot-value pair (or dialogue state) manually annotated for a sub-dialogue. These slot-value pairs are the information that should automatically be found in each segment of a considered sub-dialogue. In this example, the sub-dialogue is used either (when training) to train the AT model, or (when testing), to map the sub-dialogue segment into the AT model. The slot-value pair (or dialogue state), here **INFO: Priceramge**, is considered as the "author" in the proposed AT model.

Table 1. Example of a translated sub-dialogue example manually annotated with the slot-value pair (or dialogue state) **INFO: Pricerange** for the topic **Accomodation** [3].

Speaker	Transcription
Guide	Let's try this one, okay?
Tourist	Okay.
Guide	It's InnCrowd Backpackers Hostel in Singapore. If you take a dorm bed per person only twenty dollars. If you take a room, it's two single beds at fifty nine dollars.
Tourist	Um. Wow, that's good.
Guide	Yah, the prices are based on per person per bed or dorm. But this one is room. So it should be fifty nine for the two room. So you're actually paying about ten dollars more per person only.
Tourist	Ok. That's the price is reasonable actually. It's good.

The DSTC5 is split into two time periods. The first one, considered as the development phase, allows the participants to build and fine-tune their methods and systems using training and development data coming with manual annotations provided by the organizers. In the second one, considered as the test phase, participants are invited to run their systems on provided test data in a short time period (6 days). Data are split as follows:

- **Train data**: manual transcriptions and annotations at

both utterance and sub-dialog levels of 35 English dialogs. The automatic translations (5-best translations) of these dialogs into Chinese language are also provided.

- **Development data**: manual transcriptions and annotations at both utterance and sub-dialog levels of 2 Chinese dialogs. The automatic translations of these dialogs (5-best translations) into English language are also provided.

- **Test data**: manual transcriptions will be provided for 10 Chinese dialogs, as well as their automatic translations into English language (5-best translations), for evaluating the trackers.

Table 2 details the number of dialogues, sub-dialogues and segments of the DSTC5 for each dataset.

Table 2. Number of dialogues, sub-dialogues and segments of the DSTC5 for each dataset.

Dataset	# Dialogues	# Sub-dialogues	# Segments
Train	35	4,296	25,338
Dev	2	253	2,189
Test	10	1,387	12,290

The DSTC5 also offers four additional tasks (spoken language understanding, speech act prediction, spoken language generation, end-to-end system), which were not considered in this work. More information about these tasks could also be found in [3].

4.2. Evaluation

As described in [3], the performance will be evaluated by comparing, for every utterance, reference (*i.e.* manual) annotations and the ones given by the proposed automatic systems. As already explained in Section 4.1, for each segment of the considered sub-dialogue, an hypothesis can only be made by considering its "history" (and not the "future"). For example, if a sub-dialogue contains 10 segments, and a decision has to be made for the fifth, only the five first segments can be used to make a decision.

In the evaluation program, two "schedules" are proposed:

- **Schedule 1**: all turns are included.

- **Schedule 2**: only the turns at the end of segments are included.

As explained in the DSTC5 handbook, "*If some information is correctly predicted or recognized at an earlier turn in a given segment and well kept until the end of the segment, it will have higher accumulated scores than the other cases where the same information is filled at a later turn under schedule 1. On the other hand, the results under schedule*

2 indicate the correctness of the outputs after providing all the turns of the target segment."

Two types of evaluation metrics are finally considered:

- **Accuracy**: Proportion of sub-dialogue segments correctly labeled in comparison to the reference.

- **Precision/Recall/F-measure**:

  - **Precision**: Proportion of slot-value pairs hypothesis correctly labeled relative to the total number of slot-value pairs hypothesis.

  - **Recall**: Proportion of slot-value pairs hypothesis correctly labeled relative to the total number of slot-value pairs references.

  - **F-measure**: Harmonic mean of precision and recall.

5. RESULTS AND DISCUSSION

We propose, in this article, to track dialog states in sub-dialog segments using a topic-based representation obtained with an original Author-Topic (AT) model (see Section 3). An AT model of 20 topic spaces is elaborated using all the training data provided for the DSTC5 (see Section 4.1) and is learned with an homemade implementation. During the test phase, each turn including its history is mapped into the trained AT model to obtain a vector that will then be used to automatically identify a dialog state (slot-value pair) hypothesis using a SVM classification method, as explained in 3. While 5-best translations are provided, only the top-1 hypothesis has been used in the AT model.

In order to evaluate and compare the performance of the proposed method, the main task's organizers provided a simple baseline tracker. In this simple approach, the slot values are determined using string matching between the entries in the ontology and the word translations appearing at the beginning of a given segment to the current turn. Since two languages are studied in the DSTC5 (English and Chinese), the following methods are proposed [3]:

- **Method 1**: The automatic translation from Chinese to English is matched to the English words in the original ontology.

- **Method 2**: The Chinese utterances are matched to the automatically translated words in the ontology from English to Chinese.

For the two baseline methods, even if 5-best translations are given, only the top-1 hypothesis is used. Before analyzing the results, we can firstly note the major difference between our AT model approach and the baseline methods: our AT model should consider semantic aspect of the dialogues (higher-level representation of the word content), which is

totally ignored by the baseline methods that only use string matching.

Table 3 presents the results, both in terms of Accuracy and Precision/Recall/F-measure scores, obtained by the LIA team on the DTSC5 test set main task (dialogue state tracking) using the proposed Author-Topic model (LIA-AT), as well as the baseline results given by the main task's organizers. As a reminder, Schedule 1 takes into consideration all turns for evaluation, while Schedule 2 takes into account only the turns at the end of sub-dialogues.

Table 3. Results obtained by our team (LIA) on the DTSC5 test set main task (dialogue state tracking) using the proposed Author-Topic model (LIA-AT). Fo sake of comparison, baseline results (Baseline1 and Baseline2) submitted by the task's organizers are reported.

	Schedule 1			
	Acc.	**Prec.**	**Recall**	**F1**
Baseline1	**0.0250**	0.1148	0.1102	0.1124
Baseline2	0.0161	0.1743	**0.1279**	0.1475
LIA-AT	0.0192	**0.3130**	0.1048	**0.1570**
	Schedule 2			
	Acc.	**Prec.**	**Recall**	**F1**
Baseline1	**0.0321**	0.1425	0.1500	0.1462
Baseline2	0.0222	0.1979	**0.1774**	**0.1871**
LIA-AT	0.0214	**0.3021**	0.1046	0.1554

Globally, we can firstly see that our proposed AT model approach obtains close results considering the accuracy metric for both schedules, and a better F-measure score when all the segments are taken into account for the evaluation (schedule 1), which is not the case for the second schedule (only the last label of a sub-dialogue is taken into account), where a better F-measure score is only observed in comparison to the Baseline1. We can make two conclusions from these first observations: 1) the AT model seems to be interesting when small dialogue history is available (schedule 1), but do not perform globally better results (schedule 2); 2), the automatic translation quality should definitely have an impact on performance, as seen on the baseline methods, better results being observed when using automatic translations of the ontology at the F-measure level (Baseline2), while better results are observed on the accuracy metric when automatic translation is directly performed on the dialogue (Baseline1).

The second main observation given by Table 3 is relative to the precision of the proposed approach. We can note that a much better precision is observed no matter the schedule or the baseline method considered. We have two main assumptions about this fact. Firstly, we think that this topic-based approach may better deal with translation errors, which was our initial assumption about using such a method. And secondly, this may be due to the fact that we did not take into account

the global topic category of the sub-dialogue: indeed, in the ontology, topic is crucial since it determines a possible list of slot-value pairs. What we did was to take all the possible slot-value pairs and build our model, ignoring the topics category. At the end, we did not take any decision about the slot-value pair of a segment when the category did not match.

From all these observations, we think that there is much way of improvements for our AT model approach. The first one would be to consider the topic category when building the AT model: we observed a high precision score, which means that our method is robust, and then could easily increase the recall by including the topic category given for each sub-dialogue. Another interesting way to explore would be to go deeper into the translations hypothesis provided, since, for now, only the 1-best automatic translation of dialogues content from Chinese to English has been explored. And finally, a full evaluation of all the AT model parameters, such as the number of topics, has not yet been completed: only an AT model of 20 topics has been evaluated.

Note that all detailed results, especially the ones obtained by the other DSTC5 participants, can be found in [3].

6. CONCLUSION

Automatic translation systems, especially in spontaneous speech conditions such as human/human dialogues, make translation error. In the context of a dialog state tracking task, these errors inevitably have an impact on the global tracking performance. In this paper, an elegant way to deal with translation errors by mapping a sub-dialogue segment and its history into an Author-Topic (AT) space is proposed. In the past, the AT model demonstrated its interest on a conversation theme identification task. This high-level representation allows to take into consideration the semantic information contained in the dialogue while dealing with translation errors. Performance obtained on the fifth Dialogue State Tracking Challenge (DSTC5) showed that this AT model representation is promising, with a better F-measure score obtained when all segments are considered in the evaluation.

Much way of improvements seems obvious in regards to the unexploited possibilities offered in this dialog state tracking task and by the AT model representation. In the future, an interesting work is at the translation level, by using the 5-best translations (and not only the top-1) as well as the ontology translation that may have an impact on the performance. Fine-tuning all the AT-model parameters would also be interesting. Another perspective would be to compare this representation with other ones using a thematic representation, such as labeled LDA [26] or supervised LDA [27], since this last one is close to the AT representation. A last perspective is to add a new latent variable into the Author-Topic model, to allow this model to infer effectively an unseen sub-dialogue.

7. REFERENCES

[1] Mohamed Morchid, Richard Dufour, Pierre-Michel Bousquet, Mohamed Bouallegue, Georges Linarès, and Renato De Mori, "Improving dialogue classification using a topic space representation and a gaussian classifier based on the decision rule," in *ICASSP*, 2014.

[2] Konrad Scheffler and Steve Young, "Automatic learning of dialogue strategy using dialogue simulation and reinforcement learning," in *Proceedings of the second international conference on Human Language Technology Research*. Morgan Kaufmann Publishers Inc., 2002, pp. 12–19.

[3] Seokhwan Kim, Luis Fernando D'Haro, Rafael E. Banchs, Jason Williams, Matthew Henderson, and Koichiro Yoshino, "The Fifth Dialog State Tracking Challenge," in *Proceedings of the 2016 IEEE Workshop on Spoken Language Technology (SLT)*, 2016.

[4] David M. Blei, Andrew Y. Ng, and Michael I. Jordan, "Latent dirichlet allocation," *The Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[5] M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth, "The author-topic model for authors and documents," in *Proceedings of the 20th conference on Uncertainty in artificial intelligence*. AUAI Press, 2004, pp. 487–494.

[6] Vladimir Vapnik, "Pattern recognition using generalized portrait method," *Automation and Remote Control*, vol. 24, pp. 774–780, 1963.

[7] Mohamed Morchid, Richard Dufour, Mohamed Bouallegue, and Georges Linares, "Author-topic based representation of call-center conversations," in *IEEE Spoken Language Technology Workshop (SLT)*, 2014, pp. 218–223.

[8] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *Journal of the American society for information science*, vol. 41, no. 6, pp. 391–407, 1990.

[9] Jerome R. Bellegarda, "A latent semantic analysis framework for large-span language modeling," in *Fifth European Conference on Speech Communication and Technology*, 1997.

[10] Thomas Hofmann, "Probabilistic latent semantic analysis," in *Proc. of Uncertainty in Artificial Intelligence, UAI ' 99*. Citeseer, 1999, p. 21.

[11] Gerard Salton, "Automatic text processing: the transformation," *Analysis and Retrieval of Information by Computer*, 1989.

[12] Jerome R. Bellegarda, "Exploiting latent semantic information in statistical language modeling," *Proceedings of the IEEE*, vol. 88, no. 8, pp. 1279–1296, 2000.

[13] Yoshimi Suzuki, Fumiyo Fukumoto, and Yoshihiro Sekiguchi, "Keyword extraction using term-domain interdependence for dictation of radio news," in *17th international conference on Computational linguistics*. ACL, 1998, vol. 2, pp. 1272–1276.

[14] Thomas Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Machine Learning*, vol. 42, no. 1, pp. 177–196, 2001.

[15] Thomas Minka and John Lafferty, "Expectation-propagation for the generative aspect model," in *Conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 2002, pp. 352–359.

[16] Thomas L. Griffiths and Mark Steyvers, "Finding scientific topics," *Proceedings of the National academy of Sciences of the United States of America*, vol. 101, no. Suppl 1, pp. 5228–5235, 2004.

[17] Stuart Geman and Donald Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, , no. 6, pp. 721–741, 1984.

[18] Gregor Heinrich, "Parameter estimation for text analysis," *Web: http://www. arbylon. net/publications/text-est. pdf*, 2005.

[19] Hang Ren, Weiqun Xu, and Yonghong Yan, "Markovian discriminative modeling for dialog state tracking," in *15th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 2014, p. 327.

[20] Hang Ren, Weiqun Xu, Yan Zhang, and Yonghong Yan, "Dialog state tracking using conditional random fields," in *Proceedings of SIGDIAL*, 2013.

[21] Kai Sun, Lu Chen, Su Zhu, and Kai Yu, "A generalized rule based tracker for dialogue state tracking," in *IEEE Spoken Language Technology Workshop (SLT)*, 2014, pp. 330–335.

[22] Matthew Henderson, Blaise Thomson, and Steve Young, "Deep neural network approach for the dialog state tracking challenge," in *Proceedings of the SIGDIAL 2013 Conference*, 2013, pp. 467–471.

[23] Matthew Henderson, Blaise Thomson, and Steve Young, "Robust dialog state tracking using delexicalised recurrent neural networks and unsupervised adaptation," in *Spoken Language Technology Workshop (SLT), 2014 IEEE*. IEEE, 2014, pp. 360–365.

[24] Nikola Mrkvsić, Diarmuid O Séaghdha, Blaise Thomson, Milica Gavsić, Pei-Hao Su, David Vandyke, Tsung-Hsien Wen, and Steve Young, "Multi-domain dialog state tracking using recurrent neural networks," *arXiv preprint arXiv:1506.07190*, 2015.

[25] Guo-Xun Yuan, Chia-Hua Ho, and Chih-Jen Lin, "Recent advances of large-scale linear classification," vol. 100, no. 9, pp. 2584–2603, 2012.

[26] Daniel Ramage, David Hall, Ramesh Nallapati, and Christopher D. Manning, "Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora," in *Conference on Empirical Methods in Natural Language Processing*, 2009, pp. 248–256.

[27] Jon D. Mcauliffe and David M. Blei, "Supervised topic models," in *Advances in neural information processing systems*, 2008, pp. 121–128.