A Self-Training Framework for Automatic Identification of Exploratory Dialogue

ZHONGYU WEI,[1] YULAN HE,[2] SIMON BUCKINGHAM SHUM,[3]
REBECCA FERGUSON,[3] WEI GAO,[4] AND KAM-FAI WONG[5]

[1] *The Chinese University of Hong Kong, Hong Kong*
[2] *Aston University, UK*
[3] *The Open University, UK*
[4] *Qatar Foundation, Qatar*
[5] *Key Laboratory of High Confidence Software Technologies, China*

ABSTRACT

The dramatic increase in online learning materials over the last decade has made it difficult for individuals to locate information they need. Until now, researchers in the field of Learning Analytics have had to rely on the use of manual approaches to identify exploratory dialogue. This type of dialogue is desirable in online learning environments, since training learners to use it has been shown to improve learning outcomes. In this paper, we frame the problem of exploratory dialogue detection as a binary classification task, classifying a given contribution to an online dialogue as exploratory or non-exploratory. We propose a self-training framework to identify exploratory dialogue. This framework combines cue-phrase matching and K-nearest neighbour (KNN) based instance selection, employing both discourse and topical features for classification. To do this, we first built a corpus from transcripts of synchronous online chat recorded at The Open University annual Learning and Technology Conference in June 2010. Experimental results from this corpus show that our proposed framework outperforms several competitive baselines.

KEYWORDS: *Exploratory dialogue identification, self-training, K-nearest neighbour, classification.*

1 INTRODUCTION

Exploratory dialogue is a form of discourse associated with deep learning and learners engaging with each other's ideas constructively. It is desirable because prompting learners to employ this type of dialogue has been shown to improve learning outcomes. [1] defined exploratory dialogue as follows:"*Exploratory dialogue* represents a joint, coordinated from of co-reasoning in language, with speakers sharing knowledge, challenging ideas, evaluating evidence and considering options in a reasoned and equitable way."

Since exploratory dialogue has been shown to be a productive type of dialogue in which knowledge is made publicly accountable and reasoning is visible, the study of exploratory dialogue identification has attracted increasing attention from learning analytics researchers. Mercer et al. [2] originally conducted research on dialogue collected in face-to-face settings and identified exploratory dialogue as a type of learner talk including elements such as evaluation, challenge, reasoning and extension. Ferguson and Buckingham Shum [3] analysed transcripts from online conferences to identify exploratory dialogue. They found that markers of exploratory dialogue can be used to distinguish meaningfully between discussions and to support evaluation of them. They manually identified 94 words and phrases that signaled the presence of elements of exploratory dialogue. Examples of cue phrases for exploratory dialogue include "*but if, my view, I think, good example, good point, that is why, next step*".

Table 1 shows an excerpt from an online discussion about distance learning. Apart from those contributed by "user3", all postings are classified as *Exploratory*. Words highlighted in italics are discourse cues indicating exploratory dialogue.

Table 1. Examples of exploratory and non-exploratory dialogue.

User Id	Postings	Label
user1	*I also think* opening up the course production and design process is the way to go, but it will be a big culture change!	Exploratory
user2	*I agree with* user1 - but there are so many drivers, not least money.	Exploratory
user3	Audio back to normal speed for me now.	Non-Exploratory
user4	*I think* the key is teachers recognising that their skills lie in Learning Design, in all its variations.	Exploratory

The obvious drawback of such a cue-phrase based approach is that it is not possible to enumerate all the possible key phrases signaling the presence of exploratory dialogue. Indeed, our preliminary experiments on the online conference dataset show that the cue-phrase based approach gives high precision but low recall. In this paper, instead of using cue-phrase based methods, we investigate machine learning approaches to the automatic identification of exploratory dialogue. The three main challenges we face are:

– Firstly, the annotated dataset is limited. Although there are abundant online discussions on a wide range of topics, there are almost no annotated corpora specifically designed for detection of exploratory dialogue. This lack of annotated data corpora makes it impractical to use supervised learning methods.

– Secondly, exploratory dialogue is a form of discourse indicating that learning is likely to be taking place and that learners are going beyond a simple accumulation of ideas. Discourse features are therefore important indicators signaling the existence of exploratory dialogue. The high precision results we obtained from our collected online conference corpus using the cue-phrase based method also reveal the significance of discourse features. Discourse-based classification is intrinsically different from traditional text classification problems which are typically topic driven.

– Thirdly, although the content of online learning discussions may cover a range of topics, knowing the discussion topics in a particular dialogue segment could help with the detection of exploratory dialogue. For example, in the case of two postings extracted from an online discussion forum on the topic of "cloud computing" as shown below, both contain cue phrases indicating the presence of exploratory dialogue (these cue phrases are highlighted in italics). However, only the first posting is a positive example of exploratory dialogue. The second posting deals with an off-topic issue. This implies that both discourse and topical features should be considered when identifying exploratory dialogue.
Posting 1: *I disagree*. Freemind is superb to use for cloud computing.
Posting 2: I would like to join you for dinner, *but if* my wife comes home earlier, I will not make it.

In this paper, we treat exploratory dialogue detection as a binary classification problem that is concerned with labeling a given posting as ex-

ploratory or non-exploratory. To address the three challenges outlined above, we propose a SElf-training from Labeled Features (SELF) framework to carry out automatic detection of exploratory dialogue from online content. Our proposed SELF framework makes use of a small set of annotated data and a large amount of un-annotated data. In addition, it employs both cue-phrase matching and KNN-based instance selection to incorporate discourse and topical features into classification model training. The SELF framework makes use of self-learned features instead of pseudo-labeled instances to train classifiers by constraining the models' predictions on unlabeled instances. It avoids the incestuous bias problem of traditional self-training approaches that use pseudo-labeled instances in the training loop. This problem arises when instances are consistently mislabeled, which makes the model worse instead of better in the next iteration.

2   RELATED WORK

Exploratory Dialogue Detection: Research into exploratory dialogue originates in the field of educational research, where this type of dialogue has been studied for more than a decade. In face-to-face settings, Mercer and his colleagues [4, 1] distinguished three social modes of thinking used by groups of learners: disputational, cumulative and exploratory. They proposed that exploratory dialogue is the type considered most educationally desirable [5].

Ferguson et al. [3] explored methods of detecting exploratory dialogue within online synchronous text chat. They manually identified a list of cue phrases indicative of the presence of exploratory dialogue. Despite the identification of these phrases, this manual approach cannot easily be generalised to other online texts.

Apart from detecting exploratory dialogue within online and offline discussions, there has also been research [6, 7] into different approaches to the detection of exploratory sections of texts. In particular, this research has focused on science papers and feedback reports. This context is different to that of chat because such documents are usually grammatically correct, carefully punctuated and formally structured.

Dialogue Act Detection: Since exploratory dialogue detection can be carried out using discourse cues, it is closely related to dialogue act classification, which aims to analyze the intentions of the speaker, for example instruction or explanation. Samuel et al. [8] identified a number of cue phrases automatically and showed these can be powerful indicators

of the associated dialogue acts. Webb et al. [9, 10] explored the use of cue phrases to carry out direct classification of dialogue act.

Using manually annotated datasets such as *Verbmobil* [11], many supervised machine learning approaches have been applied to dialogue act recognition, including Hidden Markov Models [12], the language model [13], Bayesian networks [14], Decision Trees [15] using features including n-grams, syntactic tags (such as dependency parse chunks or part of speech tags), and pragmatic information.

Self Training from Labeled Features: Traditional self training approaches employ *self-labeled instances* in the training loop. Although the current model might be improved by adding self-labeled examples with the highest confidence values generated at each iteration, this is not the case because instances might be mislabeled, making the model worse in the next iteration. In order to address this problem, research has been conducted to explore *labeled features* in model learning without labeled instances. Druck et al. [16] proposed training discriminative probabilistic models with labeled features and unlabeled instances using generalized expectation (GE) criteria. He and Zhou [17] also made use of the GE criteria for self training. They derived labeled features from a generic sentiment lexicon for sentiment classification.

To summarise, exploratory dialogue can be detected using either a set of pre-defined cue phrases signaling the existence of exploratory dialogue or supervised classifiers trained on an annotated corpus. Manually defining cue phrases is both time consuming and labour intensive. On the other hand, annotated corpora are difficult to obtain for practical applications. We therefore propose a feature-based self-learning framework which combines the advantages of cue-phrase based and supervised learning approaches. Further-more, integrating a KNN-based instance selection method into the framework offers an opportunity to reduce the mislabeled instances introduced through self-training.

3 SELF-TRAINING FROM LABELED FEATURES (SELF) FRAMEWORK

We propose a SElf-training from Labeled Features (SELF) framework for exploratory dialogue detection. This framework is shown in Figure 1. We first train an initial maximum entropy (MaxEnt) classifier based on generalized expectation (GE) criteria [16], using self-learned features extracted from a small set of annotated dataset. The trained classifier is then applied to a large amount of un-annotated data. We employ a cue-phrase matching method together with the classifier in order to select positive

examples (exploratory dialogue) and improve the labelling accuracy. In order to take into account topical features, a KNN-based instance selection method is used to select pseudo-labeled instances. These are added to the original annotated training set to derive self-learned features. In the next training loop, the classifier is re-trained using the self-learned features based on GE. Training iterations terminate after five iterations or when the number of label changes in the un-annotated dataset is less than 0.5% of the size of the un-annotated dataset (50 in our study).

3.1 *Classifier Training using Generalized Expectation Criteria*

For exploratory dialogue classification, we define a label set with L labels denoted by $\mathcal{L} = \{\text{exploratory}, \text{non-exploratory}\}$. In addition, we have a corpus with a collection of M postings denoted by $\mathcal{C} = \{\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_M\}$ where the bold-font variables denote the vectors. Each posting in the corpus is a vector of M_d features denoted by $\mathbf{d} = \{f_1, f_2, ..., f_{M_d}\}$.

In case of a classifier parameterized by θ, the label l of a dialogue post $\mathbf{d}$ is found by maximizing Equation 1.

$$\tilde{l} = \arg \max_l P(l|\mathbf{d}; \theta) \tag{1}$$

Assuming we have some labeled features with probability distribution on label set $\mathcal{L}$, we can construct a set of real-valued features of the observation to express some characteristic of the empirical distribution of the training data that should also hold for the model distribution.

$$F_{jk}(\mathbf{d}, l) = \sum_{i=1}^{M} \delta(l_d = j)\delta(k \in \mathbf{d}_i), \tag{2}$$

where $\delta(x)$ is an indicator function that takes a value of 1 if x is true, 0 otherwise. Equation 2 calculates how often feature k and dialogue label j co-occur in an instance.

We define the expectation of the feature as shown in Equation 3.

$$E_\theta[\mathbf{F}(\mathbf{d}, l)] = E_{\tilde{P}(\mathbf{d})}[E_{P(l|\mathbf{d};\theta)}[\mathbf{F}(\mathbf{d}, l)]], \tag{3}$$

where $\tilde{P}(\mathbf{d})$ is the empirical distribution of $\mathbf{d}$ in the dialogue corpus $\mathcal{C}$, $P(l|\mathbf{d}; \theta)$ is a conditional model distribution parameterized at θ, and $E_\theta[\mathbf{F}(\mathbf{d}, l)]$ is a matrix of size $L \times K$, where K is the total number of features used in model learning. The jk_{th} entry denotes the expected number of instances that contain feature k and have label j.
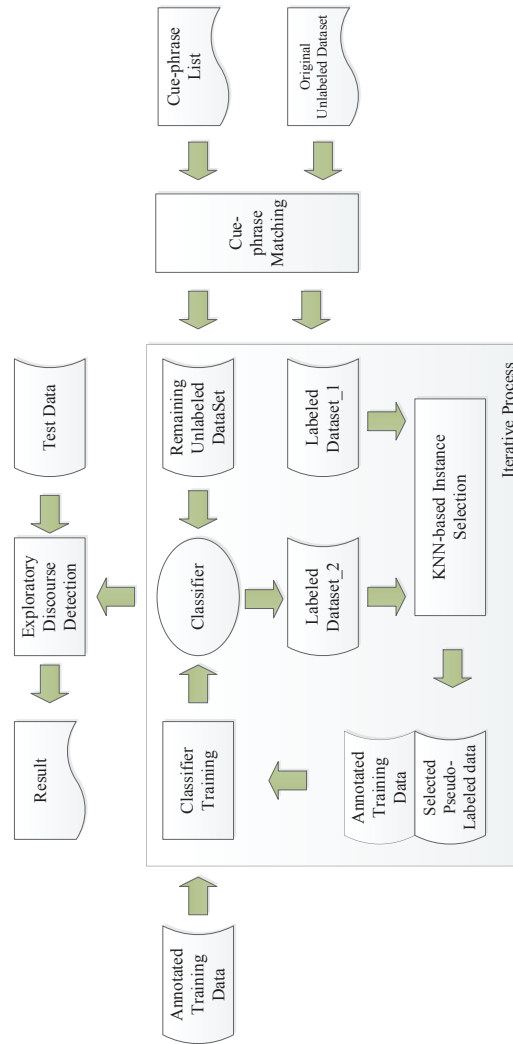
Fig. 1. A self-training framework for exploratory dialogue detection.

A criterion can be defined that minimises the KL divergence of the expected label distribution and a target expectation $\bar{\mathbf{F}}$, which is essentially an instance of generalized expectation criteria that penalizes the

divergence of a specific model expectation from a target value [16].

$$G(E_\theta | \mathbf{F}(\mathbf{d}, l)) = KL(\bar{\mathbf{F}} || E_\theta [\mathbf{F}(\mathbf{d}, l)]) \qquad (4)$$

We can use the target expectation $\bar{\mathbf{F}}$ to encode human or task prior knowledge. For example, the feature "but-if" (bi-gram feature of combining two words "but if")typically signifies an exploratory dialogue. We thus expect this feature to appear in an exploratory dialogue posting more often than in a posting that does not contain exploratory dialogue.

In our experiments, we built a MaxEnt classifier based on GE. In order to do so, we first had to select the indicative features for each class, decide on their respective class labels, and suggest the target or reference feature-class distribution for each of them.

Given a small set of annotated training data, information gain can be used to select representative features. Features with probability higher than threshold ρ are selected. The expected feature-class distribution for a given feature f is defined as a vector $\mathbf{F}(d)$ where

$$F(f, j) = \tilde{P}(j | f; \theta) \qquad (5)$$

That is, $F(f, j)$ element is the probability of a label $l = j$ being assigned given that feature f is present in a dialogue post. Such probabilities can be estimated directly from data.

3.2   *Incorporating Cue Phrases for Un-annotated Data Labelling*

In our preliminary experiments, the cue-phrase matching method based on the 94 cue phrases identified in [3] has been found to give a high precision over 95% when detecting exploratory dialogue. This suggests that discourse features based on cue phrases could potentially improve the accuracy of exploratory dialogue detection. In our proposed SELF framework, cue phrases can be utilised in two ways. One approach is to combine them with the features extracted from a small set of annotated data in order to train MaxEnt using GE. Another approach is to use them to select positive examples (exploratory dialogue) from un-annotated data, which can subsequently be combined with a small set of annotated data to train classifiers.

Our preliminary experimental results found that features selected from our small set of annotated data are typically in the range of thousands. Hence, merely combining 94 cue phrases with the selected features does not bring any obvious improvement in exploratory dialogue detection performance. Therefore, in this paper, we use cue phrases to identify exploratory dialogue within the un-annotated data and then add them to the originally labelled data set for subsequent classifier training.

3.3 *KNN-Based Instance Selection*

Within a self-training framework, pseudo-labeled instance selection is a crucial step, because adding consistently mislabeled instances to the training set can degrade the model in subsequent iterations. A straightforward way of selecting pseudo-labeled instances is only to select instances with confidence values generated by the current classifier that are above a certain threshold. Nevertheless, as mentioned in Section 1, we argue that topical features are also crucial to exploratory dialogue detection. Therefore, we propose a KNN-based instance selection method to utilise local topical features in order to reduce the number of mislabeled instances.

Once a classifier is trained, it is applied to the un-annotated data with a total of N postings $\mathcal{C}^U = \{\mathbf{d}_1^u, \mathbf{d}_2^u, ..., \mathbf{d}_N^u\}$, and it generates a corresponding label for each posting $\mathcal{L}^U = \{l_1^u, l_2^u, ..., l_N^u\}$ together with a confidence value $\mathcal{Z}^U = \{z_1^u, z_2^u, ..., z_N^u\}$ indicating how confident the classifier is when assigning the corresponding label.

We first select k nearest neighbors for each posting $\mathbf{d}_i^u \in \mathcal{C}^U$ based on the cosine similarity measurement as defined by Equation 6.

$$Sim(\mathbf{d}_i^u, \mathbf{d}_j^u) = \frac{\mathbf{d}_i^u \times \mathbf{d}_j^u}{||\mathbf{d}_i^u|| \times ||\mathbf{d}_j^u||} \tag{6}$$

This essentially selects postings that are topically similar to $\mathbf{d}_i^u$. We then decide whether the instance $\mathbf{d}_i^u$ should be selected for subsequent classifier training by considering the pseudo-labels of its k nearest neighbors. A support value s_i is calculated for instance selection.

$$s_i = \frac{\sum\limits_{j=1}^{k} \delta(l_i^u = l_j^u) z_j^u}{k} \tag{7}$$

where $\delta(x)$ is an indicator function which takes a value of 1 if x is true, 0 otherwise.

A pseudo-labeled instance $\mathbf{d}_i^u$ is selected only if its corresponding support value s_i higher than a threshold η. In our experiment, we empirically set η to 0.4 and k to 3.

4 EXPERIMENTS

4.1 *The Open University Conference 2010 Dataset*

The dataset for evaluating our proposed exploratory dialogue detection method was constructed from the Annual Learning and Technology Con-

ference: Learning in an Open World[6], run by the UK Open University (OU) in June 2010. Statistics relating to the OU Conference 2010 dataset (OUC2010) are provided in Table 2. The two-day conference was made up of four sessions - a morning session and an wvening session on each day. During the conference, 164 participants generated 2,636 postings within the synchronous text chat forum. These consisted of 6,689 distinct word tokens. These postings are typically short with a mean average of 10.14 word tokens in each one.

In addition to OUC2010, we constructed an additional un-annotated dataset from three open online courses, including 49 sessions containing 10,568 dialogue postings in total. Statistics relating to the un-annotated dataset are provided in the *Un-annotated* category of Table 2. We will make both the OUC2010 and un-annotated corpora available for public access.

Table 2. Statistics of the original OUC2010 and the un-annotated datasets.

	SessionID	Participant#	Posting#	Token#	Vocabulary#	Ave. Length
Annotated	OU_22AM	76	667	7204	2506	10.80
	OU_22PM	61	860	9073	3074	10.55
	OU_23AM	54	541	5517	2037	10.19
	OU_23PM	54	568	4937	1932	8.69
	total	164	2636	26731	6798	10.14
Un-annotated		1152	10568	97699	17268	9.244

We hired three graduate students with expertise in educational technology to annotate a subset of OUC2010. The task was to classify whether a dialogue posting was exploratory or not. The dialogue postings were presented in chronological order so that annotators could make decisions based on contextual information (i.e., postings before and after the current posting).

The Kappa coefficient [18] for inter-annotator agreement was 0.5977 for the binary classification of exploratory / non-exploratory. Statistics relating to the annotated OUC2010 dataset are presented in Table 3.

4.2 *Experimental Setup*

As shown in Table 2, the average length of each posting was relatively short. We therefore did not carry out stopwords removal or stemming.

---

[6] http://cloudworks.ac.uk/cloudscape/view/2012/

Table 3. Statistics of annotated OUC2010 dataset.

SessionID	Agreed Posting#	Exploratory#	Non-Exploratory#
OU_22AM	529	380	149
OU_22PM	661	508	153
OU_23AM	456	310	146
OU_23PM	441	219	222
total	2087	1417	670

Our preliminary experiments showed that combining unigrams with bi-grams and trigrams gave better performance than using any one or two of these three features. Therefore, in the experiments reported here, we use the combination of unigrams, bigrams and trigrams as features for classifier training and testing.

We compare our proposed framework with the following approaches in order to explore the effectiveness of the framework:

- **Cue phrase labelling (CP)**. Detect exploratory dialogue using cue phrases only.
- **MaxEnt**. Train a supervised MaxEnt classifier using annotated data.
- **GE**. Train a MaxEnt model using labeled features based on Generalized Expectation (GE) criteria. We select labeled features if their association probabilities with any one of the classes exceed 0.65.
- **Self-learned features (SF)**. The feature based self-learning framework without cue phrase matching and KNN instance selection. Documents labeled by the initial classifier are taken as labeled instances. Features are selected based on the information gain (IG) of the feature with the class label and the target expectation of each feature is re-estimated from the pseudo-labeled examples. A second classifier is then trained using these self-learned features using GE.
- **Self-learned features + KNN (SF+KNN)**. At each training iteration, the KNN-based instance selection method is used to select the pseudo-labeled instances for the derivation of self-labeled features.
- **Self-learned features + Cue-phrase + KNN (SF+CP+KNN)**. Our proposed method integrating both cue-phrase matching method and KNN based instance selection method within the self-training framework.

In each run of experiment, one session of the annotated OUC2010 was selected as the test set, and all or part of the remainder was used as the training set. The un-annotated dataset was used for self-training. For performance evaluation, all possible training and testing combinations

were tested and the results were averaged over all such runs. In each of the re-training iterations, pseudo-labeled instances were selected with the same ratio of exploratory to non-exploratory as in the initial training set. We evaluated our method using metrics including accuracy, precision, recall and F-measure.

4.3 *Results*

OVERALL PERFORMANCE Table 4 shows the exploratory dialogue classification results on the OUC2010 dataset using the methods described above. We used half a session from one of the four annotated sessions for training. The total amount of training postings ranged from 220 to 330. *CP* gives the highest precision of over 95%. However, it also generates the lowest recall value, only 42%. This indicates that the manually defined cue phrases are indeed accurate indicators of exploratory dialogue. However, they missed over half the positive exploratory dialogue.
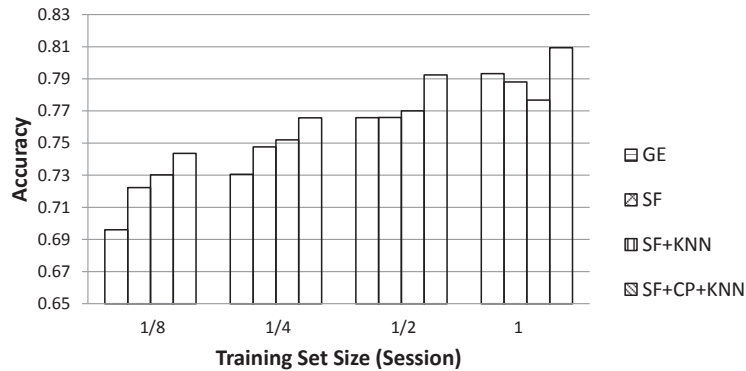
Training from labeled features only (*GE*) performs worse than the supervised classifier *MaxEnt*. The original self-learned features method, *SF*, presents a similar performance when compared to *GE*. *SF+KNN*, incorporating the KNN-based pseudo-labelled instances selection method, outperforms *SF*, showing the effectiveness of adding instances based on the labels of their k-nearest neighbours. Our proposed method, which is *SF+CP+KNN*, incorporating both cue phrase matching and KNN based instance selection, outperforms all the other baselines according to accuracy and F1 value, generating 3.4% and 4% improvement to accuracy compared to the *GE* method. Although the improvement seems modest compared to supervised learning methods such as MaxEnt, our significance test shows that the improvement is statistically significant. In addition, while supervised learning methods require annotated data for training, our proposed SELF framework only requires a small set of labelled features. This is important for exploratory dialogue detection because annotated data are scarce.

VARYING TRAINING SET SIZE To explore the influence of the amount of training data on accuracy and to investigate the effectiveness of two components within SELF, we varied the size of the annotated training set from 1/8 session to 1 session and compared the performance of different approaches. As shown in Figure 2, as the size of the training set increases, the performance of all approaches grow improves. *SF+CP+KNN* outperforms all the other methods with regard to accuracy across different sizes

Table 4. Exploratory dialogue classification results.

Approach	Accuracy	Precision	Recall	F1
CP	0.5389	**0.9523**	0.4241	0.5865
MaxEnt	0.7886	0.8262	0.8609	0.8301
GE	0.7658	0.7753	**0.8717**	0.8017
SF	0.7659	0.7572	0.8710	0.8062
SF+KNN	0.7701	0.7865	0.8539	0.8148
SF+CP+KNN	**0.7924**	0.8083	0.8688	**0.8331**

of training set. As the size of the training set increases, the accuracy of *GE* rises quickly exceeding both *SF* and *SF+KNN* when the size of the annotated data reaches 1 session. This shows that when annotated data are abundant, the effect of self-labeled feature learning and KNN-based instance selection diminishes. Nevertheless, incorporating both cue-phrase matching and KNN-based instance selection *SF+CP+KNN*, our proposed method performs significantly better than all other methods tested.

Fig. 2. Accuracy vs. training set size.

VARYING k IN KNN-BASED INSTANCE SELECTION To explore the impact of k in KNN based instance selection on the performance of our proposed SELF framework, we varied k, the number of neighbours, in *SF+CP+KNN*. Here, we only used half a session of the annotated dataset

for training. As shown in Table 5, the best performance is achieved when k is set to 3.

Table 5. Performance of proposed framework on different k

k	Accuracy	Precision	Recall	F1
1	0.7868	0.8007	0.8666	0.8282
3	**0.7924**	**0.8083**	**0.8688**	**0.8331**
5	0.7881	0.8005	0.8685	0.8292
7	0.7586	0.7505	0.8640	0.8001

5   CONCLUSIONS

In this paper, we have proposed a self-training framework for the detection of exploratory dialogue within online dialogue. Cue phrases have been employed to utilise discourse features for classification and a KNN-based instance selection method has been proposed to make use of topical features in order to reduce the erroneously-labeled instances introduced by self training. We have built the first annotated corpus for the detection of exploratory dialogue, OUC2010, from the OU Online Conference. Experimental results on OUC2010 show that our approach outperforms competitive baselines.

To the best of our knowledge, our study is the pioneer work on the automatic detection of exploratory dialogue. There are elements of this work that we would like to explore further. In the current paper, we have only focused on the use of n-grams. It would be possible to explore other features, such as the position of dialogue postings within one session. For example, dialogue exchanges at the beginning of sessions are likely to be non-exploratory because people tend to introduce themselves and greet each other when they first arrive. Moreover, if we know that one posting is exploratory, for example, if someone challenges a previous statement, then the next posting is also likely to be exploratory. Hence, contextual information such as previous and subsequent postings could be taken into account when classifying a posting. Another interesting direction will be to explore automatic ways of expanding the cue phrase list and combining it with machine learning methods for exploratory dialogue detection.

REFERENCES

1. Mercer, N., Littleton, K.: Dialogue and the development of children's thinking: A sociocultural approach. Taylor & Francis (2007)
2. Mercer, N.: Sociocultural discourse analysis. Journal of Applied Linguistics **1**(2) (2004) 137–168
3. Ferguson, R., Buckingham Shum, S.: Learning analytics to identify exploratory dialogue within synchronous text chat. In: Proceedings of the 1st International Conference on Learning Analytics and Knowledge. (2011) 99–103
4. Mercer, N.: Developing dialogues. Learning for life in the 21st century (2002) 141–153
5. Mercer, N., Wegerif, R.: Is "exploratory talk" productive talk? In Littleton, K., Light, P., eds.: Learning with computers. Routledge (1998) 79–101
6. Whitelock, D., Watt, S.: Open mentor: Supporting tutors with their feedback to students. In: 11th CAA International Computer Assisted Assessment Conference. (2007)
7. Sándor, Á., Vorndran, A.: Detecting key sentences for automatic assistance in peer reviewing research articles in educational sciences. In: Workshop on Text and Citation Analysis for Scholarly Digital Libraries. (2009) 36–44
8. Samuel, K., Carberry, S., Vijay-Shanker, K.: Automatically selecting useful phrases for dialogue act tagging. Arxiv preprint cs/9906016 (1999)
9. Webb, N., Hepple, M., Wilks, Y.: Dialogue act classification based on intra-utterance features. In: Proceedings of the AAAI Workshop on Spoken Language Understanding. (2005)
10. Webb, N., Liu, T.: Investigating the portability of corpus-derived cue phrases for dialogue act classification. In: 22nd International Conference on Computational Linguistics (COLING). (2008) 977–984
11. Jekat, S., Klein, A., Maier, E., Maleck, I., Mast, M., Quantz, J.J.: Dialogue acts in verbmobil. Technical report, DFKI GmbH (1995)
12. Levin, L., Langley, C., Lavie, A., Gates, D., Wallace, D., Peterson, K.: Domain specific speech acts for spoken language translation. In: Proceedings of 4th SIGdial Workshop on Discourse and Dialogue (SIGDIAL). (2003)
13. Reithinger, N., Klesen, M.: Dialogue act classification using language models. In: 5th European Conference on Speech Communication and Technology. (1997)
14. Ji, G., Bilmes, J.: Dialog act tagging using graphical models. In: ICASSP. (2005) 33–36
15. Verbree, D., Rienks, R., Heylen, D.: Dialogue-act tagging using smart feature selection; results on multiple corpora. In: IEEE Spoken Language Technology Workshop. (2006) 70–73
16. Druck, G., Mann, G., McCallum, A.: Learning from labeled features using generalized expectation criteria. In: SIGIR. (2008) 595–602
17. He, Y., Zhou, D.: Self-training from labeled features for sentiment analysis. Information Processing & Management **47**(4) (2011) 606–616

18. Carletta, J.: Assessing agreement on classification tasks: the kappa statistic. Computational linguistics **22**(2) (1996) 249–254

ZHONGYU WEI
THE CHINESE UNIVERSITY OF HONG KONG,
SHATIN, N.T., HONG KONG
E-MAIL: <YUTOUWEI@GMAIL.COM>

YULAN HE
SCHOOL OF ENGINEERING & APPLIED SCIENCE,
ASTON UNIVERSITY,
BIRMINGHAM, UK
E-MAIL: <Y.HE@CANTAB.NET>

SIMON BUCKINGHAM SHUM
KNOWLEDGE MEDIA INSTITUTE /
INSTITUTE OF EDUCATIONAL TECHNOLOGY,
THE OPEN UNIVERSITY,
MILTON KEYNES, UK
E-MAIL: <S.BUCKINGHAM.SHUM@OPEN.AC.UK>

REBECCA FERGUSON
KNOWLEDGE MEDIA INSTITUTE /
INSTITUTE OF EDUCATIONAL TECHNOLOGY,
THE OPEN UNIVERSITY,
MILTON KEYNES, UK
E-MAIL: <REBECCA.FERGUSON@OPEN.AC.UK>

WEI GAO
QATAR COMPUTING RESEARCH INSTITUTE,
QATAR FOUNDATION,
DOHA, QATAR
E-MAIL: <WGAO@QF.ORG.QA>

KAM-FAI WONG
KEY LABORATORY OF HIGH CONFIDENCE
SOFTWARE TECHNOLOGIES,
MINISTRY OF EDUCATION, CHINA
E-MAIL: <KFWONG@SE.CUHK.EDU.HK>