NTCIR-12: Temporal intent disambiguation subtask: Naive Bayesian Classifier to predict temporal classes

Gandhimathi Moharasan[1] and Ho Tu Bao[1,2]

[1] Japan Advanced Institute of Science and Technology, Japan.

{s1460012, bao}@jaist.ac.jp

[2] John von Neumann Institute, VNU-HCM, Ho Chi Minh City, Vietnam.

bao.ho@jvn.edu.vn

ABSTRACT

The Holab team from Japan Advanced Institute of Science and Technology (JAIST) participated in NTCIR-12 Temporal Intent Disambiguation (TID) subtask. Objective of this task is to predict temporal classes of a query which is extended from NTCIR-11 Temporal Intent Query Classification (TIQC) subtask [6] , [7]. We exploited most famous and well-defined Naive-Bayes classifier to accomplish our objective. In TID subtask, firstly, we generated different level of features from given query, later we used classifier to calculate the distribution and classify the temporal classes. In this report, we discussed about varies features, that have been used to estimate the probability distribution of four temporal intent classes (atemporal, past, recent, or future) under the temporal intent disambiguation subtask. Also we discussed about experimental results and comparative analysis with other systems submitted by different participants.

Team

Holab

Temporal Intent Disambiguation

Temporal information processing denotes relating to time. Time plays a significant role in processing and retrieving relevant document according to user requirement or necessity (e.g. Stock market rates, Exchange rates). Set of retrieved relevant document, subject to change in terms of time and temporal intent. Hence, time period in Temporal Intent Classification task has been classified into four classes such as Atemporal, Present, Past and Recency. Objective of Temporal Intent Disambiguation (TID) subtask focused to calculate the probability distribution among four temporal classes (atemporal, future, past and recency) of a given query.

Keywords

Temporal query processing, Query classification, Temporal intent disambiguation (TID), Temporal Information Access

1. INTRODUCTION

The Holab team from JAIST involved in the temporal intent classification subtask of the NTCIR-12 Temporalia data processing [7]. In recent years, temporal information retrieval and searching relevant documents has been increased due to vast usage of Internet and world wide web (WWW)[8] [5]. Despite there are ways to check information, People likes to query and read information in Internet due to development of web search engines and fast and detailed update. For example news update minutes by minutes has been done very quickly in the current Internet world. Especially TID subtask focused on time-related information processing. Even though most of people likes to search about recent information, some users search about the past and future information (e.g Recent earthquakes in Japan, weather forecast tokyo tomorrow). Hence detecting temporal aspects are inevitable and unavoidable.

Temporal Information Access task from NTCIR-12 provided a chance to work on Temporal query processing. It has two subtasks: Temporal Intent Disambiguation subtask which is extended from NTCIR-11: Temporal Qurey Intent Classification (TQIC) pilot task [7] and Temporally Diversified Retrieval (TDR) Subtask, which is focused on retrieving set of relevant documents according to temporal class of given query. The other temporal data processing shared tasks organized by various groups such as I2b2 Temporal relations [12], [14] focused on determining temporal relationships among temporal events and expressions from clinical text, TREC Temporal Summarization Task [1] focused on tracking and updating events over time, especially track the event attributes (significance) and update from reliable sources.

The remainder of this paper is structured as follows, in section 2 we briefly summarized the related research works on temporal query processing and intent disambiguation, in section 3 we discussed on feature generation and proposed method to estimate the probability distribution and query classification, in section 4 we discussed obtained results,in section 5 we briefly provided discussion and finally in section 6, we provided conclusion and some idea to improve classification accuracy in future works.

2. RELATED WORKS

As we mentioned earlier, TID subtask of NTCIR-12 is extension of NTCIR-11 TQIC subtask. There are some research works established on NTCIR-11 TQIC and temporal information retrieval(TIR) subtask [3] [9] [13].

Abhishek.*et al.,* [11] established three systems Naive bayes classifier with various features, Support Vector Machine (SVM) and Decision Trees. Hou.*et al.,* [4] developed rule-based method and machine-learning method with multi-classifier voting.

Filannino.*et al.,* [2] developed different machine learning models (SVM with linear, polynomial and RBF kernel, Naive Bayes, C4.5 decision tree and Random Forests) using various features such as POS-tags (Penn Tree bank) with Wordnet related attributes.

3. TEMPORAL INTENT DISAMBIGUATION

Objective of TID subtask is to classify user query into four classes: *past, recency, future and atemporal* [7]. Our system is established the well-defined Naive Bayes classifier to classify the temporal classes.

3.1 Feature generation

Before applying Naive Bayes classifier, we followed some preprocessing steps to prepare training file and generates the extensive features. Query processing time and query string probability of temporal classes were extracted from given raw data. After that query string is parsed through Stanford NLP[1] [10] to tag POS features. Stanford taggar uses the Penn treebank tagset.Later we extracted tense of the query string as a feature from POS tags. To pull out "Tense of query", we extracted all words that have a POS tag that starts with a "V" (Especially verbs: VB-base form, VBD-past tense, VBG-present participle, VBN- Past participle, VBP, and VBZ). Also number of words and query processing time from the raw string considered as a feature. We believe that the extensive consideration of features helps to improve the probability distribution and accuracy of temporal classes classification.

To apply machine learning method, we created a training file with Query string, Query processing time, Tense of the query, Number of words in the query and finally probability distribution of temporal classes from the raw data. We trained the model to learn the probability distribution with prepared training file. In the same way we prepared the testing file to estimate the probability distribution of temporal classes for formal run queries.

[1]http://nlp.stanford.edu/software/tagger.shtml

3.2 Method: Naive Bayesian Classifier

We used Naive bayes classifier from Weka toolkit[2] to train the classifier. With the consideration of supervised learning classification problem, we are focusing to estimate the probability distribution for the unknown query string $f : X \rightarrow Y$ or equally P(Y|X). In other words, X = $\{ X_1, X_2, X_3...X_n \}$ denotes the features of given query string (Query string, Query processing time, Tense of the query, Number of words in the query) and Y (Atemporal, Present, Past, Recency) denotes the label of temporal class. One way is to develop the classifier model P(Y|X) is to exploit the training data (Query string) to estimate the P(X|Y) and P(Y). The below steps we followed to train the naive bayes classifier and estimate probability distribution for unknown queries.

1. Step 1: Preprocessing the training and testing data file such as extracting query string, query issue time and probability distribution of temporal classes from raw text
2. Step 2: Generate features: Various feature generation such as tense of the query, number of words in the query, base form of the verb and etc.,
3. Step 3: Train Naive Bayes classifier using weka with prepared training input file
4. Step 4: Estimate the probability distribution for formal run data using developed model
5. Step 5: Evaluate the established naive bayes classifier model

4. RESULTS AND ANALYSIS

In this section we discuss about the results achieved through developed Naive Bayes classifier. After established model, we estimated the probability distribution for the formal run (Testing file) data. Table 1: Denotes the accuracy of classification with averaged per-class absolute loss and cosine similarity. Figure 1: Denotes the probability distribution among classes for each query which is contributed to determine the temporal classes. Whereas Figure 2: Shows the temporal classes distribution for TID formal run evaluation results. From Figure 2, we can see that classification of Past, Recent and Atemporal classes is high compared to Future class. It is very difficult to classify past and recent classes, as it has the thin line of time differences. However the accuracy of classification is still questionable as from Table 1: We can see that the our system accuracy is low compared to other systems. From the evaluation we can see that the cosine similarity and averaged per-class absolute loss are vary inversely one increases as the other decreases. We evaluated the classification accuracy of the developed model with confusion matrix. Table 2: Shows the confusion matrix for TID formal run and true positive value is highlighted.

From Table 2 : We are able to see that the major sources of classification mistakes from frequency:

[2]http://www.cs.waikato.ac.nz/ml/weka/

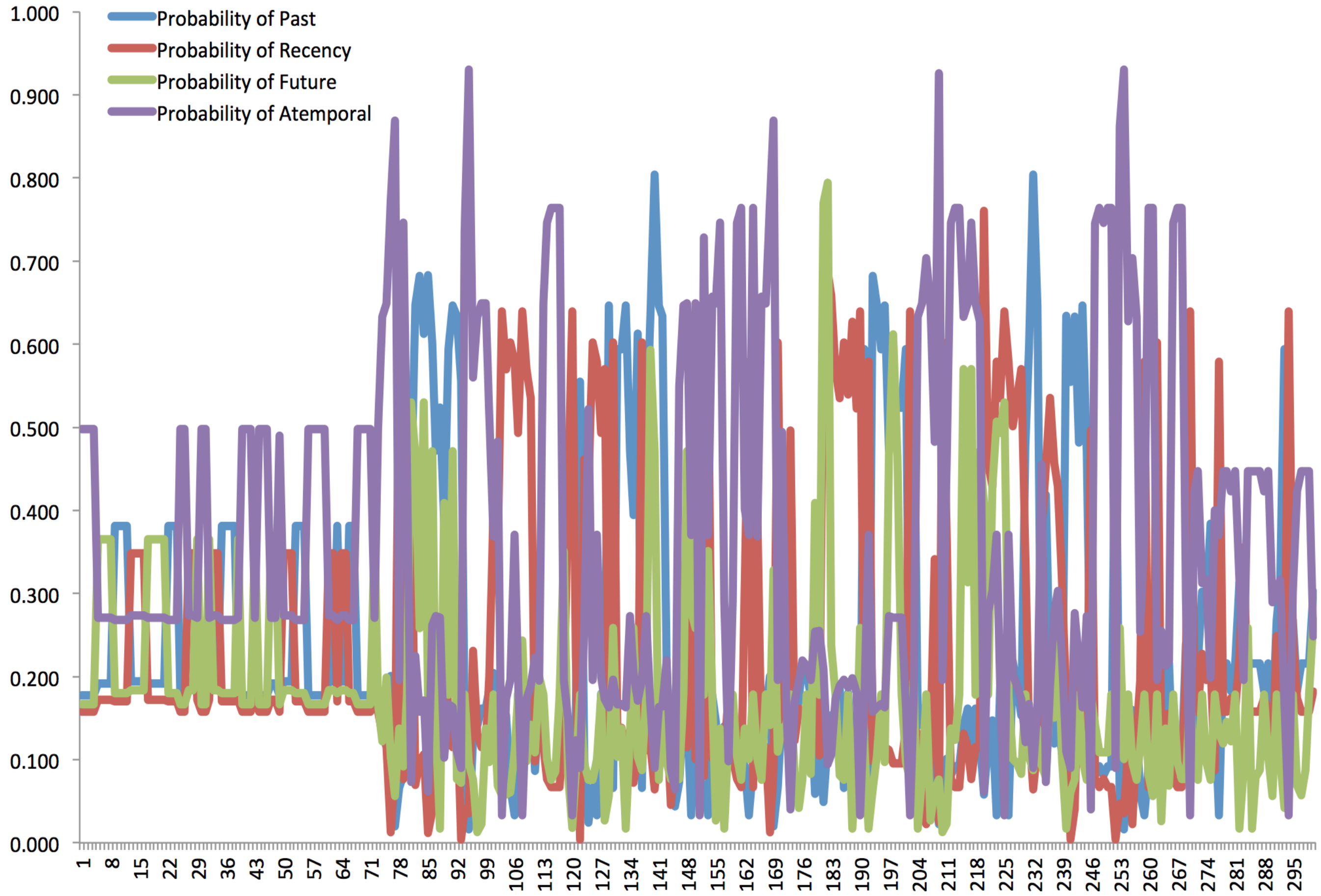

Figure 1: Probability distribution among four temporal c lasses for TID formal run

Run-ID	AvgAbsLoss	AvgCosin
Holab-TID-E-1	0.27789	0.58609

Table 1: Experimental results for TID formal run

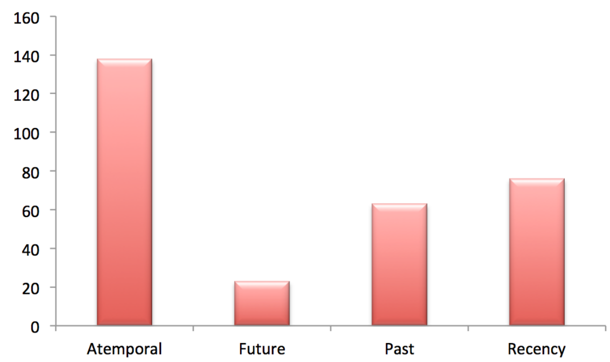

Figure 2: Temporal classes distribution for TID formal run

Past as Recency and Atemporal:

48 Past instances have been misclassified to 18 as Recency and 30 as Atemporal classes respectively. Some of the examples for Past as Atemporal misclassified queries are : *"cabernet sauvignon 2003"* and *"What Is Kony 2012 Scam"*.Some of past events have been classified as Recency. For example: *"release date xbox one"* and *"when was television invented"* which have all been submitted with only considering the tense. Hence the queries have been classified the temporal orientation as Atemporal and Recency.

Recency as Past and Atemporal:

55 Recency instances have been misclassified to 13 as Past and 42 as Atemporal classses respectively. Some of the examples for Recency as Atemporal misclassified queries are

Actual	Predicted			
	Past	Recency	Present	Atemporal
Past	**19**	18	8	30
Recency	13	**18**	2	42
Present	16	24	**8**	27
Atemporal	15	16	5	**39**

Table 2: Confusion matrix for TID formal run

: *"NFL Schedules"* and *"24 Hour Fitness Class Schedule"*. Some of Recency events classified to Past class, simply *"price hike in bangladesh"*, *"adidas Originals sheep shoes"* , *"NCAA Baseball Regionals Scores"* and *"new movie chinese"* which all have been submitted with the query processing time 1st May 2013. We can see that the first two queries are clearly denoting the recency class and our system is clearly classified to Recency class as well. Search results are strictly relating to recent time period, but the other two queries in the examples are misclassified due to time period confusion. Hence our model misclassified to Past class. New movie chinese may denote the upcoming movie or recently released one.

Atemporal as Past and Recency:

31 Atemporal instances have been classified to 15 Past and 16 Recency classes. But only very few queries has been classified to Present class due to thin time period difference between Recency, Present and Future class. Atemporal queries are not considering temporal aspects and orientation such as *"Energy Managment"* , *"samsung logo"* , *"causes of global warming"* and *"noaa weather service"*. First two queries are misclassified to Recency class and the last two queries are misclassified to Past class.

5. DISCUSSION

We developed a system(Naive Bayes classifier) with various features successfully classifiers to classify temporal intent of a given query. We established the system for supervised learning scenarios with the usage of training file. In future we are planning to make our system flexible, such like our system should able to extend with semi-supervised and unsupervised scenarios. Also as we mentioned earlier the misclassification between Past, Recency and Atemporal classes are high. To overcome this problem, we will focus to develop the hand-coded rules with naive bayes classifier to determine the time differences in future. In case of feature generation, both temporal aspects and orientation are the key factor to consider for TID task. Hence we will focus to find a new way to detect the temporal aspect feature for a given query in future.

6. CONCLUSIONS

In this paper, we discussed about the proposed method to classify the temporal intent disambiguation (TID), one of the subtasks from NTCIR-12. Nature of query string in NTCIR-12 is short compared to NTCIR-11. Even though the NTCIR-11 data helps to train the classifier, detecting temporal aspects and orientation from query string is a non-trivial task . Despite of this nature, we managed to achieve better performance than some other systems. However the error analysis provides inside that the mis-classification of temporal classes occurred between the Atemporal and Future. Also the results shows that extensive analysis of reasons should figured out, which is causing the mis-classifications. As far as from our analysis, the results can be improved using hand-coded rules especially for time period with detecting accurate temporal aspects and orientation in future.

7. REFERENCES

[1] J. A. Aslam, M. Ekstrand-Abueg, V. Pavlu, F. Diaz, and T. Sakai. Trec 2013 temporal summarization. In *TREC*, 2013.

[2] M. Filannino and G. Nenadic. Using machine learning to predict temporal orientation of search engines' queries in the temporalia challenge. In *NTCIR*, 2014.

[3] M. Hasanuzzaman, G. Dias, and S. Ferrari. Hultech at the ntcir-11 temporalia task: Ensemble learning for temporal query intent classification. In *The 11th NTCIR Conference on Evaluation of Information Access Technologies*, pages p–478, 2014.

[4] Y. Hou, C. Tan, J. Xu, Y. Pan, Q. Chen, and X. Wang. Hitsz-icrc at ntcir-11 temporalia task. In *NTCIR*, 2014.

[5] H. Joho, A. Jatowt, and R. Blanco. Ntcir temporalia: a test collection for temporal information access research. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion*, pages 845–850. International World Wide Web Conferences Steering Committee, 2014.

[6] H. Joho, A. Jatowt, R. Blanco, Y. Haitao, and S. Yamamoto. Overview of ntcir-12 temporal information access (temporalia) task. In *In Proceedings of the 12th NTCIR Conference on Evaluation of Information Access technologies*. Citeseer, 2016.

[7] H. Joho, A. Jatowt, R. Blanco, H. Naka, and S. Yamamoto. Overview of ntcir-11 temporal information access (temporalia) task. In *NTCIR*. Citeseer, 2014.

[8] H. Joho, A. Jatowt, and B. Roi. A survey of temporal web search experience. In *Proceedings of the 22nd*

international conference on World Wide Web companion, pages 1101–1108. International World Wide Web Conferences Steering Committee, 2013.

[9] R. R. Larson. A logistic regression approach for ntcir-11 temporalia. In *NTCIR*, 2014.

[10] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky. The stanford corenlp natural language processing toolkit. In *ACL (System Demonstrations)*, pages 55–60, 2014.

[11] A. Shah, D. Shah, and P. Majumder. Andd7@ ntcir-11 temporal information access task. In *NTCIR*, 2014.

[12] U. O. Sun Weiyi, Rumshisky Anna. Evaluating temporal relations in clinical text: 2012 i2b2 challenge,. *Journal of the American Medical Informatics Association*, pages amiajnl–2013, 2013.

[13] H. Yu, X. Kang, and F. Ren. Tuta1 at the ntcir-11 temporalia task. 2014.

[14] J. C.-Y. W. J.-M. C. R. T.-H. T. W.-L. H. Yung-Chun Chang, Hong-Jie Dai. Tempting system: A hybrid method of rule and machine learning for temporal relation extraction in patient discharge summaries. *Journal of Biomedical Informatics*, 46:s54–s62, 2013.