ECNU at 2016 eHealth Task 1: Handover Information Extraction

Yang Song[1,2], Yun He[1,2], Hongyu Liu[1,2], Yueyao Wang[1,2], Qinmin Hu[1,2], Liang He[1,2], and Guihua Luo[3,4]

[1] Shanghai Key Laboratory of Multidimensional Information Processing
[2] Department of Computer Science & Technology, East China Normal University, Shanghai, 200241, China
{ysong,yhe,liuhy,yywang}@ica.stc.sh.cn, {qmhu,lhe}@cs.ecnu.edu.cn
[3] Science and Technology Commission of Shanghai Municipality
[4] Shanghai Digital Trade Co. Ltd, Shanghai, 200241,China
ghluo@metinform.cn

Abstract. The CLEF eHealth 2016 Task 1 is set to automatically assign pre-defined medical tag to each word in the patient case records. The difficulty of the task is that many classes have little training data. This paper presents our work on the 2016 CLEF eHealth Task 1. In particular, we propose an optimized Conditional Random Field algorithm to better fulfill the task. We also utilize the information extracted through association rules and MetaMap to boost the performance of our results. The evaluation results show our runs outperform the four official baselines in this difficulty task.

Keywords: Conditional Random Field, Association Rule, MetaMap, Information Extraction

1 Introduction

CLEF eHealth 2016 Task 1 addresses clinical information extraction, related to Australian nursing shift change[1][2]. This extends the 2015 task 1a of converting verbal nursing handover to written free-text records. In 2016, our participants are challenged to maximize the correctness in structuring these written free-text records by pre-filling a handover form by automatically identifying relevant text-snippets for each slot of the form.

The data set utilized in this task is called NICTA Synthetic Nursing Handover Data [3][4]. It has been developed for clinical speech recognition and information extraction related to nursing shift-change handover at NICTA from 2012. This data set contains 200 synthetic patient cases which can be used for training and validation. The patient cases record the patient's profile and health information.

In this task, the organizers provide us with 36 tags which are related to the categories of Patient Introduction, My Shift, Appointments, Medication and

Future Case in the handover form. We are asked to assign one of the 36 tags to each word in the patient case records. The 36 tags are summarized in table 1.

Table 1. Summarization of the Tags

Patient Introduction	Medication
PatientIntroduction_AdmissionReason/Diagnosis	Medication_Dosage
PatientIntroduction_Ageinyears	Medication_Medicine
PatientIntroduction_Allergy	Medication_Status
PatientIntroduction_CarePlan	
PatientIntroduction_ChronicCondition	
PatientIntroduction_CurrentBed	
PatientIntroduction_CurrentRoom	
PatientIntroduction_Disease/ProblemHistory	
PatientIntroduction_Gender	
PatientIntroduction_GivenNames/Initials	
PatientIntroduction_Lastname	
PatientIntroduction_UnderDr_GivenNames/Initials	
PatientIntroduction_UnderDr_Lastname	
Appointment/procedure	My Shift
Appointment/Procedure_City	MyShift_RiskManagement
Appointment/Procedure_ClinicianGivenNames/Initials	MyShift_Contraption
Appointment/Procedure_ClinicianLastname	MyShift_Input/Diet
Appointment/Procedure_Day	MyShift_OtherObservation
Appointment/Procedure_Description	MyShift_Status
Appointment/Procedure_Status	MyShift_Wounds/Skin
Appointment/Procedure_Time	MyShift_Output/Diuresis /BowelMovement
Appointment/Procedure_Ward	MyShift_Activities OfDailyLiving
Future Care	NA
Future_Alert/Warning/AbnormalResult	NA
Future_Discharge/TransferPlan	
Future_Goal/TaskToBeCompleted/ExpectedOutcome	

Our system architecture is proposed in Figure 1. We mainly utilize the Conditional Random Field (CRF) model to achieve the results [6][7]. To better satisfy the assignment, we design strategy to automatically select features to train different tags. Furthermore, we use association rules [8][9] to extract information of patient name, age, gender, room number, bed number and doctor name to enhance our outputs. At last, we apply MetaMap to recognize the Unified Medical Language System (UMLS) concepts in the patient case records. The MetaMap tags are utilized as the supplement to decide which word should have the tag of Medication_Medicine.

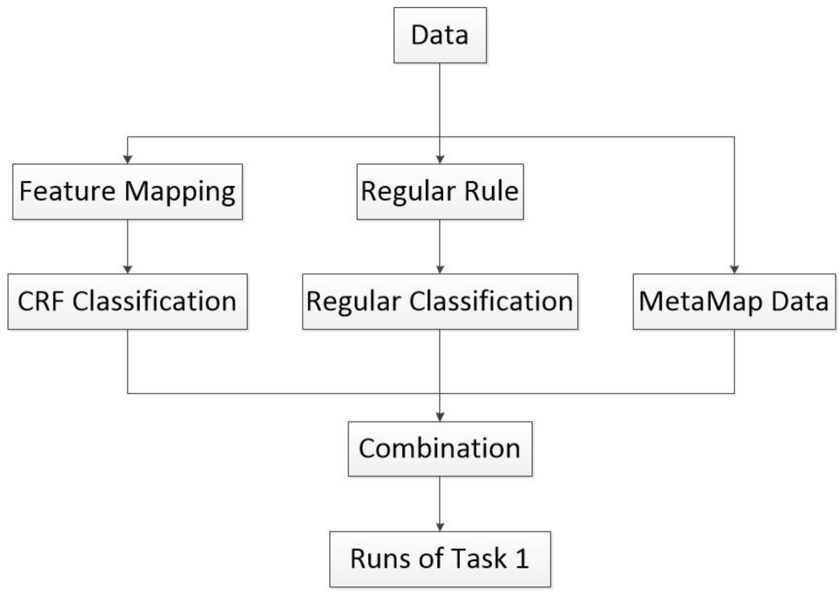

Fig. 1. System Architecture

2 Methodology

2.1 CRF Model Training

The organizers provide us with 17 features for each word in the health care records, which are listed in table 2 [5]. All of these features are relevant to the 36 health care information tags. We intent to utilize CRF model to implement the labeling task. However, different feature set have different precision and recall rate. We are motivated to solve this problem by select the proper feature set for each tag to train the CRF model.

Suppose $A = \{f_i\}, i = 1, 2, ..., 17$ is the feature set which contains the whole 17 features, and B is the feature set which contains features should be removed from A. For a tag, such as PatientIntroduction_CarePlan, we utilize the following method to discover its suitable feature set for the CRF model training. Denote C as the suitable feature set for the PatientIntroduction_CarePlan tag.

- Utilizing feature set A to train CRF model in data set 1, and using data set 2 to test the model. denote the precision of PatientIntroduction_CarePlan as p.
- $\forall f_i \in A, i = 1, 2, ..., 17$, remove f_i from A. Train the CRF model by utilizing the data set 1 based on the feature set A without f_i. Test the learned CRF model on data set 2. Denote the precision on the PatientIntroduction_CarePlan tag as p_{f_i}.
- Calculate the mean value of $p_{f_i}, i = 1, 2, ..., 17$ and denote it as p_m. If $p_{f_i} > p$ and $p_{f_i} - p_m > 10\%$, then put the feature f_i into set B.

Table 2. Experimented Syntactic Features

ID	Name	Definition	Example	Software
1	Word	Word itself	"Patients" or "had"	None
2	Lemma	Lemma of the word	"patients" or "have"	CoreNLP
3	NER (named entity recognition)	NER tag of the word for named entities(ie, person, location, organization, other proper name) and numerical entities (ie. date, time, money, number)	"number" for "5"	CoreNLP
4	POS (part of speech)	POS tag of the word	"IN"(ie, perposition) for "in", "NN"(ir, c-ommon noun as opp-osed to Proper Name, "PN") for "bed", "CN"(ie, cardinal number) for "5"	CoreNLP
5	Parse tree	Parse tree of the sentence from the root to the current word	"ROOT-NP-NN" (ie, root-noun phrase-common noun). For "5" in "In bed 5 we have..."	CoreNLP
6	Basic dependents	Basic dependents of the word	"Cardinal number 5" that refers to the bed ID for "bed" in "In the bed 5 we have..."	CoreNLP
7	Basic governors	Basic governors of the word	Preposition "in" and subject "we" for "have" in "In bed 5 we have..."	CoreNLP
8	Phrase	Phrase that contains this word	"In bed 5" for "bed" in "In bed 5 we have"	MetaMap
9	Top 5 candidates	Top 5 candidates retrie-ved from UMLS	"BP" may refer to, for example, "Bachelor of Pharmacy", "bed-pan", "before present", "birth-place", or "blood pressure"	MetaMap
10	Top mapping	Top UMLS mapping for the concept that is the best match with a given text snippet	"pneumonia" is a type "respiratory tract infection"	MetaMap

ID	Name	Definition	Example	Software
11	Medication score	1 if the word is a full term in ATCL (Anatomical Therapeutic Chemical List); else 0.5 if it can be found in ATCL; 0 otherwise	1 for “acetylsalicylic acid”	NICTA
12	Location	Location of the word on a tenpoint scale from the beginning of the document to its end	“1” for the first word and “10” for the last word	NICTA
13	Normalized term frequency	Number of times a given term occurs in a document divided by the maximum of this term frequency over all terms in the document		NICTA
14	Top 5 candidates	As 9 using SNOMED-CT-AU (Systematized Nomenclature of Medicine-Clinical Terms-Australian Release)		Ontoserver
15	Top mapping	As 1using SNOMED-CT-AU		Ontoserver
16	Top 5 candidates	As 9 using AMT[5]		Ontoserver
17	Tom mapping	As 10 using AMT		Ontoserver

– The suitable set $C = A - B$. This means that the features in the set B will decrease the precision and ought to be removed from the features collection.

Thus, we obtain a suitable feature set for each tag. The suitable feature sets are noted as $C_i, i = 1, 2, ..., 36$, corresponding to the 36 medical tags. Next step, we utilize $C_i, i = 1, 2, ..., 36$, to train CRF model respectively, and we obtain 36 results by using those learned CRF model to perform labelling task in a patient case record. Then, we combine these results through the method of voting. Note the 36 medical tags as $t_i, i = 1, 2, ..., 36$. For a word w in the patient case record, suppose the tag t_i appears n_i times. We select the tag which have the highest voting count as the final tag of the word w. Suppose t_A is the label of w obtained by the CRF model which is trained by utilizing the feature set A. If two or more tags which share the same highest appear time we select t_A as the final tag of w.

2.2 Utilizing Association Rules to Extract Information

We analyze the test and training set and discover that information about patient name, room, bed, gender, age and doctor name have the same expressing format. Thus, we utilize association rules to extract relevant information and assign them with the relevant tags automatically. The association rules are listed in table 3

2.3 MetaMap Application

We use MetaMap[6] to obtain the Unified Medical Language System (UMLS) concepts in the patient case records. Meanwhile, we identify the words with MetaMap tag “phsu” and assign them with the tag “Medication_Medicine”.

2.4 Combination

We combine our results achieved from the CRF model, association rules and MetaMap in order to get a better performance. Runs obtained by CRF models are utilized as our basic runs. We use the results achieved by using rules and MetaMap to modify the tags in CRF runs.

Suppose S_C, S_R, S_M are the results achieved by CRF, association rules and MetaMap respectively. For a word in a patient case record, if its tag in CRF run is different from its tag in rule and MetaMap runs, we utilize the tag in association rules and MetaMap runs to replace the tag in CRF run. Note that, there is no conflict between the tags resulted from the association rules and the MetaMap, since the two methods extract information for different tags.

[6] https://mmtx.nlm.nih.gov/

Table 3. Association Rules

Tag	Rule
PatientIntroduction_CurrentBed	"bed" + No.
PatientIntroduction_CurrentRoom	"room" + NO.
PatientIntroduction_Gender	word in {male,him,his,he,gentleman, gentlemen, man,men,boy,boys,himself } or {female,her,she, hers,lady,ladies, woman,women,girl,girls,herself }
PatientIntroduction_GivenNames/Initials	without the word "under" + words capitalized the first letter
PatientIntroduction_Lastname	without the word "under" + words capitalized the first letter
PatientIntroduction_UnderDr _GivenNames/Initials	"under" + ("Dr") + name
PatientIntroduction_UnderDr_Lastname	"under" + ("Dr") + name
PatientIntroduction_Ageinyears	NO. + "years old" / "yrs old" / "yr old"

3 Experiments and Evaluation

We utilize the features in the file "CRF_Matrix_noLabel.data" provided by organizers to train our CRF model. We implement the training and tagging process of CRF by a open-source toolkit named "CRF++-0.58". Data set 1 and data set 2 are all used for training a CRF model to tag each word in the data set 3. Specifically, we submit six runs based on two methods, where the description for each method is as follows.

– Method A: We use rules mainly based on regular expressions to extract information of bed number, room number age and doctor's name. Then, we use all features provided to run CRF model and obtain label for each word. Finally, we combine the result achieved by CRF model with result obtained by the association rules.
– Method B: We use the method, which is detailed in section 2.1, to select the best suitable feature set for different label among the features provided by the organizers. Then, we train the CRF model based on those feature sets. At last, we use the method of voting to determine the label for each word. We also use association rules to extract information of bed number, room number, age and doctor's name. The final submission is the combination of results obtained by the association rules and the voting methods.

The primary evaluation measure of this year is the macro-precision(MaAPrec), macro- recall(MaARec), macro-F1(MaAF1), micro-precision(MiAPrec), micro-recall(MiARec), micro-F1(MiAF1), precision of NA tag(NA-Prec), recall of NA tag(NA-Recall), F1 value of NA tag(NA-F1). Data set 1 is used for training. Data set 2 is used for validation and data set 3 is designed for test. Evaluation of our methods over the data set 1, 2 and 3 is summarized in Table 3, where A and B imply our method A and method B respectively.

4 Conclusions and Future Work

In 2016 CLEF eHealth task 1, we propose an optimized CRF model and utilize the association rules and MetaMap tag to achieve the better performance for the handover form automatically filling. All of our submissions outperform the four baseline methods. In the future, we will continue on the research of handover form automatically filling methods.

5 Acknowledgement

This research is funded by National Key Technology Support Program (No. 2015BAH12F01-04) and Science and Technology Commission of Shanghai Municipality (No.14DZ1101700). We also thank anonymous reviewers for their review comments on this paper.

Table 4. Evaluation of our submissions

Classifier	Set	MaAPrec	MaARec	MaAF1
NA	Training	0	0	0
NA	Validation	0	0	0
NA	Test	0	0	0
Majority	Training	0.002	0.029	0.003
Majority	Validation	0.001	0.029	0.003
Majority	Test	0	0.029	0.001
Random	Training	0.017	0.027	0.017
Random	Validation	0.018	0.025	0.018
Random	Test	0.018	0.028	0.019
NICTA	Training	1	0.976	0.98
NICTA	Validation	0.485	0.297	0.324
NICTA	Test	0.435	0.233	0.246
A	Training	0.995	0.992	0.994
A	Validation	0.467	0.329	0.345
A	Test	0.493	0.406	0.374
B	Training	0.454	0.328	0.344
B	Validation	0.483	0.313	0.331
B	Test	0.428	0.292	0.297

Classifier	Set	MiAPrec	MiARec	MiAF1
NA	Training	0	0	0
NA	Validation	0	0	0
NA	Test	0	0	0
Majority	Training	0.058	0.105	0.075
Majority	Validation	0.05	0.085	0.063
Majority	Test	0.016	0.027	0.02
Random	Training	0.017	0.03	0.022
Random	Validation	0.018	0.03	0.022
Random	Test	0.018	0.03	0.022
NICTA	Training	1	0.914	0.955
NICTA	Validation	0.649	0.398	0.493
NICTA	Test	0.433	0.368	0.398
A	Training	0.995	0.991	0.993
A	Validation	0.655	0.478	0.553
A	Test	0.51	0.522	0.516
B	Training	0.461	0.528	0.492
B	Validation	0.603	0.454	0.518
B	Test	0.581	0.459	0.513

Classifier	Set	NA-Prec	NA-Recall	NA-F1
NA	Training	0.444	1	0.615
NA	Validation	0.409	1	0.58
NA	Test	0.407	1	0.579
Majority	Training	0	0	0
Majority	Validation	0	0	0
Majority	Test	0	0	0
Random	Training	0.49	0.032	0.06
Random	Validation	0.437	0.031	0.057
Random	Test	0.405	0.03	0.055
NICTA	Training	0.903	1	0.949
NICTA	Validation	0.597	0.931	0.727
NICTA	Test	0.682	0.831	0.749
A	Training	0.993	0.998	0.995
A	Validation	0.667	0.927	0.775
A	Test	0.816	0.788	0.802
B	Training	0.864	0.706	0.777
B	Validation	0.677	0.92	0.78
B	Test	0.675	0.881	0.764

References

1. Kelly, L., Goeuriot, L., Suominen, H., Nvol, A., Palotti, J., Zuccon, G.: Overview of the CLEF eHealth Evaluation Lab 2016. CLEF 2016 - 7th Conference and Labs of the Evaluation Forum, Lecture Notes in Computer Science (LNCS), Springer, September, 2016.
2. Suominen, H., Zhou, L.Y., Goeuriot, L., Kelly, L.: Task 1 of the CLEF eHealth evaluation lab 2016: Handover information extraction. CLEF 2016 Evaluation Labs and Workshop: Online Working Notes, CEUR-WS, September, 2016.
3. http://www.nicta.com.au/nicta-synthetic-nursing-handover-open-software/
4. http://www.nicta.com.au/open-data/
5. Suominen, H., Zhou, L.Y., Hanlen, L., Ferraro, G.: Benchmarking Clinical Speech Recognition and Information Extraction: New Data, Methods, and Evaluations." JMIR Medical Informatics 3.2 (2015).
6. Lafferty, J., Andrew M., and Fernando C.P.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence data. (2001).
7. Sha, F., Pereira, F.: Shallow Parsing with Conditional Random Fields. In: Proceedings of the 2003 Conference of the North American Chapter of Association for Computational Linguistics on Human Language Technology, Vol.1 (2003).
8. Bing, L., Wynne H., Yiming, M.: Integrating classification and association rule mining. Proceedings of the fourth international conference on knowledge discovery and data mining. (1998).
9. Vaidya, J., Chris C.: Privacy preserving association rule mining in vertically partitioned data. Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, (2002).