CMedTEX: A Rule-based Temporal Expression Extraction and Normalization System for Chinese Clinical Notes

Zengjian Liu, MS[1], Buzhou Tang, PhD[1*], Xiaolong Wang, PhD[1], Qingcai Chen, PhD[1], Haodi Li, MS[1], Junzhao Bu, MS[1], Jingzhi Jiang, MS[1], Qiwen Deng, MS[2], Suisong Zhu, MS[2]
[1]Key Laboratory of Network Oriented Intelligent Computation, Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, China; [2]The Sixth People's Hospital of Shenzhen, Shenzhen, China

Abstract

Time is an important aspect of information and is very useful for information utilization. The goal of this study was to analyze the challenges of temporal expression (TE) extraction and normalization in Chinese clinical notes by assessing the performance of a rule-based system developed by us on a manually annotated corpus (including 1,778 clinical notes of 281 hospitalized patients). In order to develop system conveniently, we divided TEs into three categories: direct, indirect and uncertain TEs, and designed different rules for each category of them. Evaluation on the independent test set shows that our system achieves an F-score of 93.40% on TE extraction, and an accuracy of 92.58% on TE normalization under "exact-match" criterion. Compared with HeidelTime for Chinese newswire text, our system is much better, indicating that it is necessary to develop a specific TE extraction and normalization system for Chinese clinical notes because of domain difference.

Introduction

Time is one of the most important aspects of information as information is usually event-centred and organized in order of time. To understand progress of events, people have to capture temporal information related to them, including temporal expressions (TEs) and temporal relations. Because of this, temporal information extraction as a fundamental task of natural language processing (NLP) has been always attracting a great deal of attention. A large number of systems using various methods, including rule-based and machine learning (ML)-based, have been developed to extract temporal information in text in different languages (e.g., English, Chinese, French, etc.) from different domains such as newswire and social media. In the clinical domain, there also have been a few temporal information extraction systems. However, most of them are designed for English clinical notes from electronic medical record (EMR) systems, and very limited studies have been carried out on Chinese clinical notes.

In recent years, with the rapid growth of EMRs in China, information extraction from Chinese clinical notes has attracted more and more attention of Chinese researchers. In this study, we developed a temporal expression (TE) extraction and normalization system for Chinese clinical notes, assessed its performance on a manually annotated corpus of Chinese clinical notes, and published our system online. To the best of knowledge, this is one of the earliest studies on TE extraction and normalization for Chinese clinical notes, our system is the first publicly available system, and we believe it will provide valuable insights into NLP research in Chinese clinical text.

Background

TE extraction and normalization are integral parts of information extraction and have been extensively studied in multilingual text in multiple domains. The development of related technology was mainly driven by several public challenges such as the TE recognition and normalization (TERN) challenge[1] in 2004, the automatic content extraction (ACE) challenges[2] in both 2005 and 2007, and a series of TempEval challenges: TempEval-1[3] in 2007, TempEval-2[4, 5] in 2010 and TempEval-3[6] in 2013. In the early challenges (e.g., the 2004 TERN challenge and the 2007 ACE challenge), only English text in newswire domain was considered. Subsequently text in other languages such as Chinese was also gradually included, and the text was not limited to newswire text. For example, extracting and normalizing TEs in Chinese text from broadcast news, newswire and weblogs was added as a new subtask of the 2005 ACE challenge. In these challenges, both rule-based and ML-based approaches have been investigated, and the rule-based ones almost always showed better performance than the ML-based ones. For example, in the 2010 TempEval-2 challenge, the best TE extraction and normalization system (i.e., HeidelTime[7-9]) was a rule-based system which significantly outperformed other state-of-the-art ML-based systems such as two conditional random

[*] Corresponding author; Email: tangbuzhou@gmail.com

fields (CRF)-based systems: TIPSem[10] and TRIOS[11]. In the rule-based system, not only word, but also syntactic and semantic information such as part-of-the-speech and semantic role of words were used for rule design[10, 12, 13].

In the clinical domain, with the development of clinical NLP, a number of attempts have been proposed for English clinical notes. The early representative clinical NLP systems such as ConText[14], CNTRO[15, 16], MedLEE[17], CTakes[18], etc. usually include an individual rule-based module for TE extraction and normalization. As there was no benchmark corpus available for TE extraction and normalization system evaluation, it was difficult to compare them. In 2012, the i2b2 (Center of Informatics for Integrating Biology and Bedside) organized a clinical NLP challenge[19] of temporal information extraction in clinical text, including a subtask of TE extraction and normalization. This challenge provided a benchmark corpus to evaluate different systems developed by researchers all over the world. In this challenge, both rule-based and ML-based methods were proposed for TE extraction and normalization, and finally the rule-based system developed by Sunghwan Sohn et al[20] achieved the best performance, slightly better than the CRF-based system developed by Yan Xu et al[21]. Subsequently, the SemEval challenge in 2015 included a clinical TempEval task[22, 23] including a subtask of TE extraction and normalization in clinical text. Most of systems were rule-based. The information used to design rules included word, pos, clinical entity, etc.

In recent years, with the rapid growth of EMRs in China, the need of information extraction from Chinese clinical notes has become more and more strident, including temporal information extraction. However, very limited studies have been proposed for TE extraction and normalization in Chinese clinical text. Zhou et al[24] proposed a simple rule-based method to extract and normalize TEs in Chinese clinical text and evaluated the method on a very small manually annotated dataset with only 1,207 TEs from 147 medical records. Despite the important contributions of previous studies on TE extraction and normalization in Chinese clinical text, none has systematically analyzed the difference of the TE extraction and normalization systems for clinical text in Chinese and English, and the difference of the TE extraction and normalization systems for Chinese clinical text in the newswire domain and clinical domain, which is very important for system porting from English clinical text or Chinese newswire text to Chinese clinical text.

In this paper, we analyzed the characteristics of Chinese clinical notes, developed a rule-based TE extraction and normalization system for Chinese clinical text (available at http://icrc.hitsz.edu.cn:8096/CMedTEX), and analyzed which cases were the new cases in Chinese clinical text compared to English clinical text and Chinese newswire text. To the best of our knowledge, this is one of the earliest comprehensive studies on TE extraction and normalization for Chinese clinical notes, and our system is the first publicly available TE extraction and normalization system for Chinese clinical notes, whose source code was also released for application at: http://icrc.hitsz.edu.cn/Article/show/147.html.

Methods

Dataset and annotation

A real world dataset of 1,778 de-identified clinical notes of 281 hospitalized patients were collected from the sixth people's hospital of Shenzhen, Shenzhen, China. Two master students with computer science and medical informatics background were recruited to annotate TEs in the clinical notes by following annotation guidelines designed for this study. When creating the annotation guidelines, we referred to the TE annotation guidelines of TimeML[25, 26] for English newswire text and the 2012 i2b2 NLP challenge[27] for English clinical text, and defined the same three attributes for each TE, i.e., type, value and modifier, where there were four types (DATE, TIME, DURATION and FREQUENCY) and seven modifiers (APPROX, START, END, MIDDLE, MORE, LESS and NA (default value)), and the value attribute is the standard temporal value defined in ISO8601. Some examples were shown in Figure 1.

To calculate the inter-rater annotation agreement of the two annotators, we asked them to annotate 300 out of 1,778 notes at the same time. When there was any disagreement between the two annotators, a third judge was brought in to select or confirm the different annotations. Following the 2012 NLP challenge[19, 27], the inter-rater annotation agreement of the two annotators was calculated under two criteria: "exact-match" and "partial-match", where "exact-match" and "partial-match" denote that two time expressions are correctly matched if and only if their text spans are exactly the same and overlap with each other respectively. Based on these 300 notes, inter-rater agreements of TE span, type and modifier using Kappa statistics[28] were 0.9035, 0.9681 and 0.9195, the inter-rater annotation agreements of type value and modifier using accuracy were 0.9756, 0.8773 and 0.9879 under the exact-match criterion (see Table 1), indicating that our annotation was reliable. The remaining 1,478 notes were annotated by one annotator only.

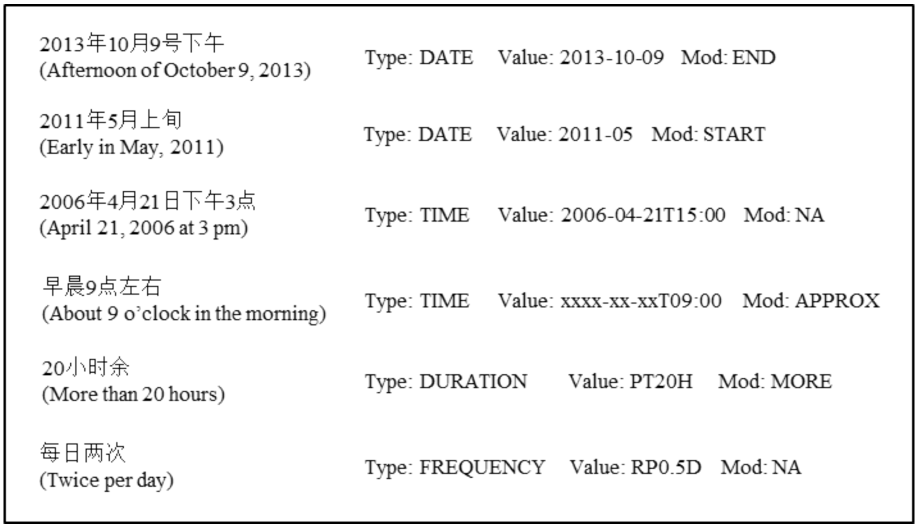

Figure 1. Examples of TEs with attributes. 'x' denotes the missing information needing to be determined during normalization.

Table 1. Inter-annotator agreement under both "exact-match" and "partial-match" criteria. "/" denotes there is no available value.

	Exact-match		**Partial-match**	
	Accuracy	**Kappa**	**Accuracy**	**Kappa**
Span	/	0.9035	/	0.9432
Type	0.9756	0.9681	0.9589	0.9438
Value	0.8773	/	0.8625	/
Modifier	0.9879	0.9195	0.9707	0.8309

After annotation, we obtained 46,508 TEs. The annotated dataset was randomly divided into training and test sets. The training set consisted of 1076 records with 28,442 TEs of 170 hospitalized patients, while the test set consisted of 702 records with 18,066 TEs of 111 hospitalized patients. The statistics of the dataset were shown in Table 2 in detail, where "#*" denotes the number of '*'. Among four types of TEs, dates accounted for the highest proportion of 36.56% (17,005/46,508), and times accounted for the lowest proportion of 15.84% (7,368/46,508).

Table 2. Statistics of the annotated dataset for TE extraction and normalization.

	#Patient	**#Record**	**#Temporal expression**				
			#Date	**#Time**	**#Duration**	**#Frequency**	**#All**
Training set	170	1,076	10,205	4,500	6,821	6,916	28,442
Test set	111	702	6,800	2,868	4,214	4,184	18,066
Total	281	1,778	17,005	7,368	11,035	11,100	46,508

TE extraction and normalization

Although a large number of both machine learning-based and rule-based methods have been proposed for TE extraction, experiments in TempEval-2 and 2012 i2b2 NLP challenges demonstrated that the rule-based methods obviously outperformed the machine learning-base methods, and the best systems of the two challenges were all rule-based[4, 19]. For TE normalization, almost all systems were rule-based. Therefore, in this study, we investigated the rule-based methods for TE extraction and normalization. In order to develop our system conveniently, we analyzed the characteristics of Chinese clinical notes and classified them into three categories: direct, indirect and uncertain, which denoted complete and exact TEs, incomplete or implicit but exact TEs, and inexact TEs. Figure 2 listed some examples. For each category of TEs, we designed specific rules, which were presented in the following paragraphs in detail.

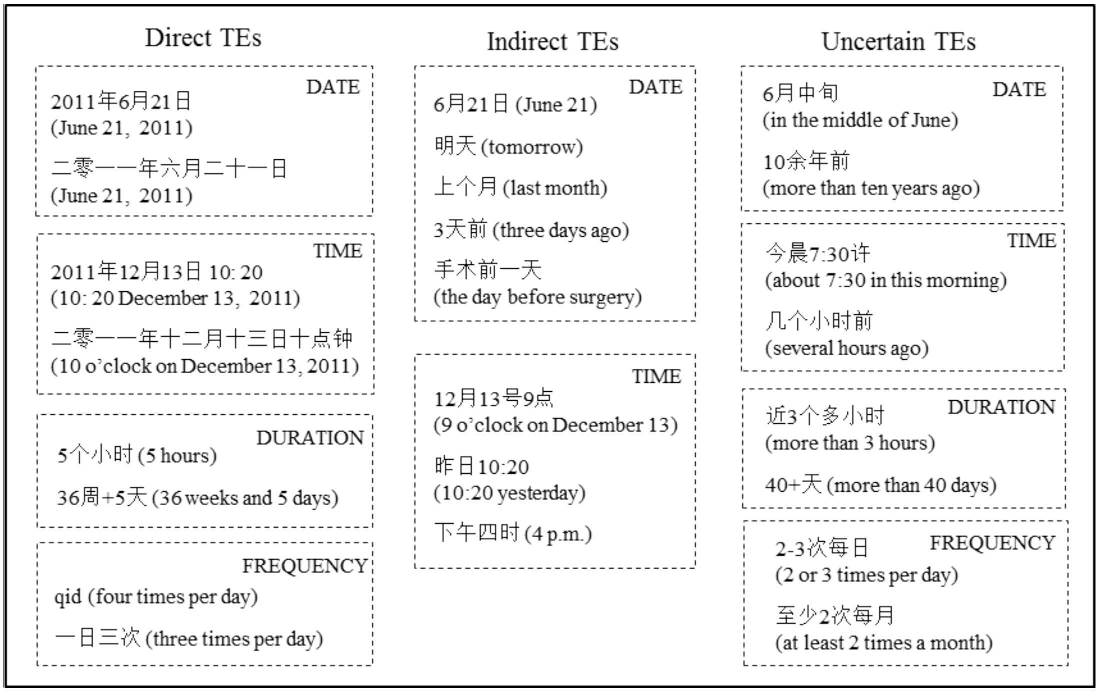

Figure 2. Examples of TEs in the three categories: direct, indirect and uncertain.

1. Direct TEs

It is easy to extract and normalize direct TEs in clinical text as they usually appear in common patterns explicitly and do not need further processing for normalization. For each type of direct TEs, we defined a large number of rules. Taking dates for example, they are calendar dates in three patterns as shown in Figure 3, where "YearDigit", "MonthNumber" and "DayNumber" are sets of regular expressions for year, month and day numbers respectively (including Chinese numbers such as "零" (zero), "一" (one), "二" (two), etc.), '年', '月' and "日/号" are the units of year, month and day, and "NormYear", "NormMonth", "NormDay" are mapping functions from "YearDigit", "MonthNumber" and "DayNumber" to their standard forms respectively.

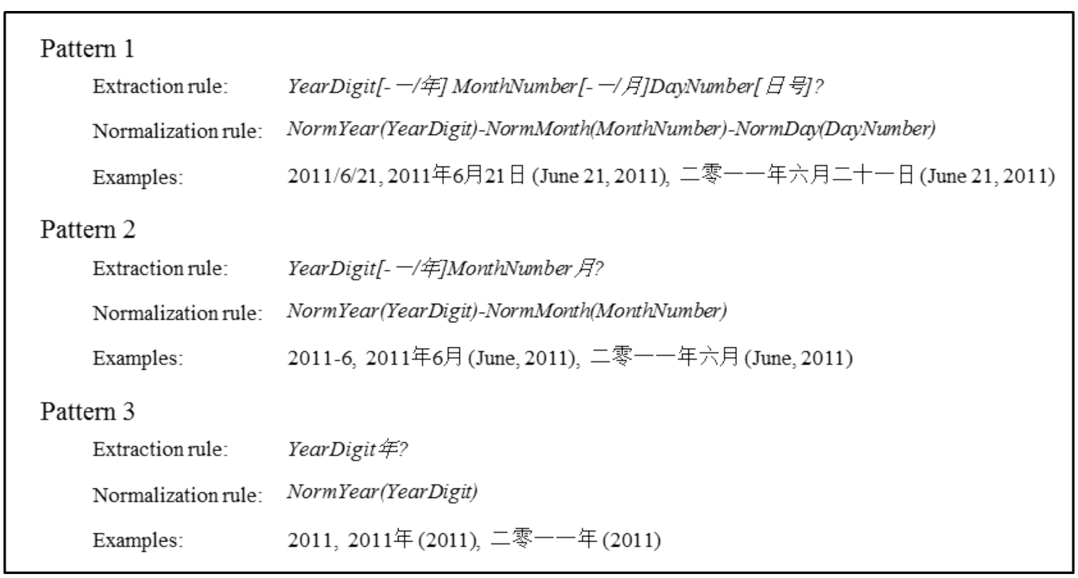

Figure 3. Extraction and normalization rules for direct date expressions.

2. Indirect TEs

Compared to direct TEs, indirect TEs are also exact but incomplete or implicit. There are two cases: 1) incomplete temporal expressions missing some parts such as year and month. For example, “6 月 21 日” (“June 21”) is a date expression missing year; 2) implicit temporal expressions representing relative time such as “3 天前” (“three days ago”) and “手术前一天” (“the day before surgery”). To exact indirect TEs, we modified rules for direct TEs. To normalize indirect TEs, we need further inference to complete the missing part of them or to calculate the absolute time of them. For example, we need to complete the year information for “6 月 21 日” (“June 21”), and to calculate the absolute time of “3 天前” (“three days ago”) and that of “手术前一天” (“the day before surgery”). The key of indirect TE normalization is to find out the baseline time of them which may be a calendar time or the occurring time of some event such as “手术” (“surgery”) in “手术前一天” (“the day before surgery”). In this study, we tried the following three strategies gradually to determine the baseline time of indirect TEs:

1) The baseline time of all indirect TEs in a note is the admission/discharge/create time (for admission notes, discharge summaries and progress notes respectively.) mentioned in the note, which is easy to determine as it usually appears at the beginning of a note.

2) The baseline time of an indirect TE is the nearest TE in front of it in the same sentence or a previous sentence.

3) The baseline time, which is the occurring time of some events (e.g., surgery, operation, transfer, etc.), is determined by the previous two strategies when they first appear.

Figure 4 gave some examples using different strategies for indirect TE normalization, where there was a fragment of an admission note with two direct TEs and five indirect TEs enclosed in square brackets. It is clear that the five indirect TEs were wrongly normalized except the last one when only using strategy 1 mentioned above. When strategy 2 was added, the first three indirect TEs were correctly normalized but the last two TEs not. Nevertheless, strategy 3 corrected the last two TEs normalized by strategy 2 since they were times relative to “手术” (“the operation”) and “入院” (“admission”) with baseline times: 2011-10-18 and 2012-02-03T15:00, respectively. It should be noted that the time of “手术” (“the operation”) is “2011-10-18” not “2011-10-19” since the nearest TE in front of “手术” (“the operation”) that first appears is “18 日” (“18th”) not “一天后” (“one day later”).

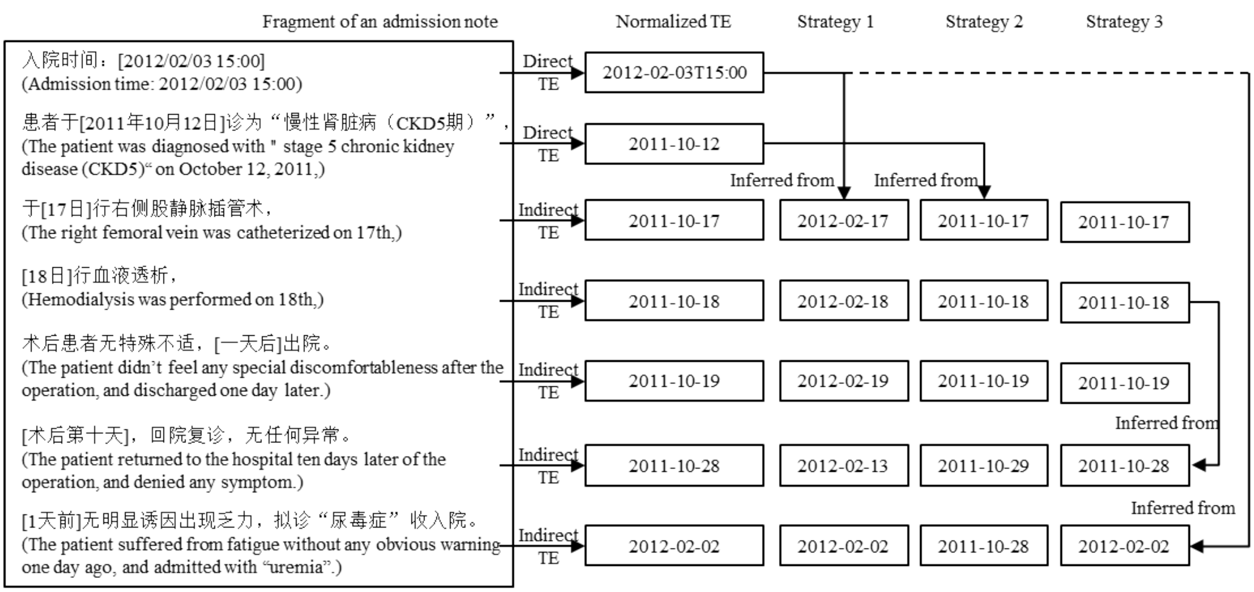

Figure 4. Examples of indirect TE normalization using different strategies.

3. Uncertain TEs

There are two types of uncertain TEs: direct/indirect TEs modified by fuzzy adjectives or adverbs, and fuzzy temporal phrases (e.g., more than, about, in the morning, etc.). For the first type of uncertain TEs, rules for extraction were also modified from direct/indirect TEs’ by adding fuzzy adjectives or adverbs; and the key of their normalization lies in a list of modifiers and in which way an uncertain TE and a modifier should be combined together. For the second type of uncertain TEs, we needed to collect a list of fuzzy temporal phrases with modifier attribute. Figure 5 listed some examples of uncertain TEs with modifier value.

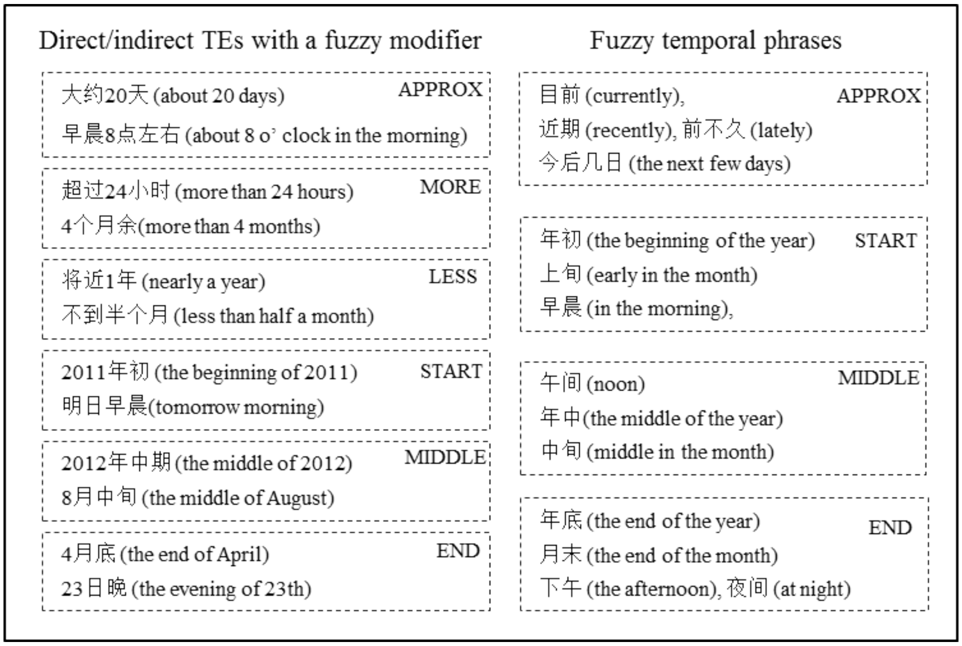

Figure 5. Examples of uncertain TEs with modifier value.

Experiments and evaluation

We developed our system on the training set, and then evaluated its performance on the independent test set. In order to investigate the effects of the three categories of TEs, we started with the baseline system only considering direct TEs, and then progressively added indirect TEs and uncertain TEs.

The micro-averaged precision, recall and F-measure under exact/partial-match criteria were used to evaluate system performance for TE extraction, and the micro-averaged accuracy for TE normalization. Under exact-match criterion, an entity is correctly predicted if and only if its text span is exactly the same as that of one entity in the gold standard, while, under partial-match criterion, an entity is correctly predicted if it overlaps with any entity in the gold standard. We developed an evaluation tool based on the official evaluation program provided by the 2012 i2b2 NLP organizers.

Results

Table 3. Performance of our system on the test set when different categories of TEs were considered (%). "Val" and "Mod" denote the value and modifier attributes.

Category	Exact-match				Partial-match			
	Span	Type	Val	Mod	Span	Type	Val	Mod
Direct	68.09(79.46/59.57)	59.31	57.88	59.50	81.18(95.11/70.81)	63.91	61.50	66.44
Direct+indirect	82.04(87.00/77.62)	77.28	74.04	77.12	88.66(94.32/83.64)	82.01	77.58	78.76
Direct+indirect+uncertain	**93.40**(90.83/96.11)	95.72	**92.58**	95.39	96.42(94.01/98.96)	97.65	93.53	97.42

Table 3 showed the performance of our system on the test set when different categories of TEs were considered. The number in columns 2 and 6 were F-measures followed by corresponding precision and recall in parentheses for TE extraction under exact-match and partial-match criteria, while the numbers in other columns were accuracies for identifying corresponding attributes such as numbers in columns 4 and 8 for TE normalization under two criteria. Under exact-match criterion, our system ("direct") achieved a precision of 79.46%, a recall of 59.57% and an F-measure of 68.09% on TE extraction, and an accuracy of 57.88% on TE normalization when it only considered direct TEs. When indirect TEs were added ("direct + indirect"), the precision, recall and F-measure of our system on TE extraction were improved to 87.00%, 77.62% and 82.04%, and the accuracy on TE normalization was improved to 74.04%. When further considering uncertain TEs ("direct + indirect + uncertain"), our system achieved highest precision of 90.83%, highest recall of 96.11% and highest F-measure of 93.40% on TE extraction and highest accuracy of 92.58% on TE normalization. These results demonstrated that both indirect and uncertain TEs played an important role in TE extraction and normalization.

The detailed results of the best system for each type of TE are shown in Table 4. On TE extraction, F-measures ranged from 88.45% to 97.92% under exact-match criterion. Among four types of TEs, our system performed best for times and worse for durations. On TE normalization, accuracies ranged from 87.69% to 96.48% under exact-match criterion, and our system performed best for times and worse for frequencies.

Table 4. Detailed results of the best system for each type of TE (%).

Type	Exact-match				Partial-match			
	Span	Type	Value	Modifier	Span	Type	Value	Modifier
Overall	93.40(90.83/96.11)	95.72	92.58	95.39	96.42(94.01/98.96)	97.65	93.53	97.42
Date	93.65(90.79/96.69)	96.69	91.99	95.68	94.65(91.80/97.68)	97.68	92.37	96.32
Time	**97.92**(97.98/97.87)	97.87	**96.48**	97.25	98.76(98.81/98.71)	98.71	97.32	97.45
Duration	88.45(82.86/94.85)	94.85	94.71	94.35	91.66(85.88/98.27)	98.27	96.23	97.15
Frequency	93.30(92.97/93.62)	93.62	87.69	93.28	97.00(96.89/97.11)	97.11	89.13	95.84

Discussion

In this paper, we comprehensively investigated the rule-based methods for TE extraction and normalization in Chinese clinical notes. We manually created an annotated dataset of 1,778 clinical notes in Chinese for system development and test. Experiments on the dataset showed that our system achieved the highest F-measure of 93.40% on TE extraction, and the highest accuracy of 92.58% on TE normalization under exact-match criterion. We published our system online (http://icrc.hitsz.edu.cn:8096/CMedTEX), which is the first public available TE extraction and normalization system for Chinese clinical notes and will be useful for future Chinese clinical NLP studies. The source code of our system was available for application at: http://icrc.hitsz.edu.cn/Article/show/147.html.

It is clear that our system performance is progressively improved when gradually adding specific rules for direct, indirect and uncertain TEs as shown in Table 3. The main reason lies in that when the rules for direct TEs are deployed to indirect and uncertain TEs, parts of indirect and uncertain TEs will be recognized as direct TEs, which will cause some errors. For example, when applying rules for direct TEs to "2011 年 6 月 21 日下午" ("afternoon of June 21, 2011"), an indirect TE, and "3 天以前" ("3 days ago"), an uncertain TE, we will only obtain two direct TEs: a date of "2011 年 6 月 21 日" ("June 21, 2011") and a duration of "3 天" ("3 days"), which are two errors. Therefore, we designed specific rules for indirect and uncertain TEs, respectively, which are derived from the rules for direct TEs, to correct these errors.

Compared to TE extraction and normalization systems in other domains (e.g., newswire) or other languages (e.g., English), our system is promising. For example, the best system[8] on the Chinese corpus of TempEval-2 in newswire domain (i.e., HeidelTime) was rule-based and achieved an F-measure of 89.30% on TE extraction and an accuracy of 87% on TE normalization under exact-match criterion. In the 2012 i2b2 challenge about temporal information extraction for English clinical text, the highest F-measure and accuracy on TE extraction and normalization were 91.44%[21] and 73.13%[20] respectively. Our system showed better performance than the best systems of those two challenges even though the comparison between them is not very fair as they were not evaluated under the same conditions.

In order to investigate the effect of domain difference on TE extraction and normalization systems, we compared HeidelTime, a state-of-the-art TE extraction and normalization system for Chinese newswire text, with our system on the test set. Under exact-match criterion, the F-measure and accuracy on TE extraction and normalization of HeidelTime were only 38.35% and 8.60%, much lower than those of our system, indicating that TEs in clinical text are much different from TEs in other domains (especially newswire domain). Therefore, it is necessary to design new rules and systems to extract and normalize TEs in clinical text. The reasons why HeidelTime performed so bad for Chinese clinical notes may lie in the following two aspects: 1) there are a large number of proper TEs in clinical text such as qid ("一天四次", four times per day), which cannot be recognized by HeidelTime; and 2) many indirect TEs in clinical text are relative to clinical events such as operation and transfer, which are not considered by HeidelTime.

Furthermore, we analyzed the errors caused by our system, and found that most of the errors were related to word ambiguities. For example, "9-10" can be either a date "9 月 10 号" ("September 10") or a duration "9 至 10" ("from 9 to 11"), "9.10" can be either a date "9 月 10 号" ("September 10") or a number, "10 个月" ("10 months") can be either a duration or a child's age, "夜间" ("at night") is usually a duration, but not a separate TE when it is a part of problem "夜间盗汗" ("night sweats"), and "一周" ("one week") is usually a duration, but not a TE when it appears in problem "脐带绕颈一周" ("a nuchal cord") with another meaning "one round" in Chinese. Among these errors, the first three are common problems in both English and Chinese clinical text and the last two are special problems in Chinese clinical text. To tackle these problems, we may need to take other information, such as syntax and clinical entities, into account. However, it is not easy to obtain these information as the studies on Chinese clinical NLP has just began.

In the future, we plan to focus on the following two aspects for further improvement: 1) consider more information, such as syntactic and semantic information, to fix current errors; and 2) attempt machine learning-based methods to enhance our system.

Conclusion

In this study, we developed a rule-based time expression extraction and normalization system for Chinese clinical notes. The time expressions were summarized into three different categories: direct, indirect and uncertain expressions, which facilitated the construction of rules more easily and clearly. Our system achieved a span F-measure of 93.40% and a value accuracy of 92.58% under "exact-match" criterion. The online system is publicly available and its source code is released for application.

Acknowledgements

This paper is supported in part by grants: National 863 Program of China (2015AA015405), NSFCs (National Natural Science Foundations of China) (61573118, 61402128, 61473101, 61472428, 61173075 and 61272383), Strategic Emerging Industry Development Special Funds of Shenzhen (JCYJ20140508161040764, JCYJ20140417172417105, JCYJ20140627163809422 and JSGG20151015161015297), Innovation Fund of Harbin Institute of Technology (HIT.NSRIF.2017052) and Program from the Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education (93K172016K12).

References

1. Negri M, Marseglia L. Recognition and normalization of time expressions: ITC-irst at TERN 2004. Rapport interne. 2004.
2. Doddington GR, Mitchell A, Przybocki MA, Ramshaw LA, Strassel S, Weischedel RM. The automatic content extraction (ACE) program-tasks, data, and evaluation. LREC; 2004.
3. Verhagen M, Gaizauskas R, Schilder F, Hepple M, Katz G, Pustejovsky J. Semeval-2007 task 15: tempeval temporal relation identification. Proceedings of the 4th International Workshop on Semantic Evaluations; 2007:75-80.
4. Verhagen M, Sauri R, Caselli T, Pustejovsky J. SemEval-2010 task 13: tempEval-2. Proceedings of the 5th international workshop on semantic evaluation; 2010:57-62.
5. Pustejovsky J, Verhagen M. SemEval-2010 task 13: evaluating events, time expressions, and temporal relations (TempEval-2). Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions; 2009:112-6.
6. UzZaman N, Llorens H, Allen J, Derczynski L, Verhagen M, Pustejovsky J. Tempeval-3: Evaluating events, time expressions, and temporal relations. arXiv preprint arXiv:12065333. 2012.
7. Strötgen J, Gertz M. A baseline temporal tagger for all languages. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing; 2015:541-7.
8. Li H, Strötgen J, Zell J, Gertz M. Chinese temporal tagging with HeidelTime. EACL 2014; 2014:133-7.
9. Strötgen J, Gertz M. HeidelTime: High quality rule-based extraction and normalization of temporal expressions. Proceedings of the 5th International Workshop on Semantic Evaluation; 2010:321-4.

10. Llorens H, Saquete E, Navarro B. Tipsem (english and spanish): evaluating crfs and semantic roles in tempeval-2. Proceedings of the 5th International Workshop on Semantic Evaluation; 2010:284-91.
11. UzZaman N, Allen JF. TRIPS and TRIOS system for TempEval-2: Extracting temporal information from text. Proceedings of the 5th International Workshop on Semantic Evaluation; 2010:276-83.
12. Llorens H, Saquete E, Navarro B, Li L, He Z. Data-driven approach based on semantic roles for recognizing temporal expressions and events in Chinese. Natural Language Processing and Information Systems; 2011:88-99.
13. Negri M, Saquete E, Martínez-Barco P, Munoz R. Evaluating knowledge-based approaches to the multilingual extension of a temporal expression normalizer. Proceedings of the Workshop on Annotating and Reasoning about Time and Events; 2006:30-7.
14. Chapman WW, Chu D, Dowling JN. ConText: An algorithm for identifying contextual features from clinical text. Proceedings of the Workshop on BioNLP 2007; 2007:81-8.
15. Tao C, Solbrig HR, Chute CG. CNTRO 2.0: a harmonized semantic web ontology for temporal relation inferencing in clinical narratives. AMIA Summits on Translational Science Proceedings; 2011:64-8.
16. Tao C, Wei W-Q, Solbrig HR, Savova G, Chute CG. CNTRO: a semantic web ontology for temporal relation inferencing in clinical narratives. AMIA Annual Symposium Proceedings; 2010:787-91.
17. Friedman C, Alderson PO, Austin J, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. Journal of the American Medical Informatics Association. 1994 ;1(2):161-74.
18. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. Journal of the American Medical Informatics Association. 2010 ;17(5):507-13.
19. Sun W, Rumshisky A, Uzuner O. Evaluating temporal relations in clinical text: 2012 i2b2 challenge. Journal of the American Medical Informatics Association. 2013 ;20(5):806-13.
20. Sohn S, Wagholikar KB, Li D, Jonnalagadda SR, Tao C, Elayavilli RK, et al. Comprehensive temporal information detection from clinical text: medical events, time, and TLINK identification. Journal of the American Medical Informatics Association. 2013 ;20(5):836-42.
21. Xu Y, Wang Y, Liu T, Tsujii J, Eric I, Chang C. An end-to-end system to identify temporal relation in discharge summaries: 2012 i2b2 challenge. Journal of the American Medical Informatics Association. 2013 ;20(5):849-58.
22. Velupillai S, Mowery DL, Abdelrahman S, Christensen L, Chapman WW. BluLab: Temporal information extraction for the 2015 clinical TempEval challenge. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015); 2015:815-9.
23. Bethard S, Derczynski L, Pustejovsky J, Verhagen M. Semeval-2015 task 6: Clinical tempeval. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015); 2015:806-14.
24. Zhou Xiaojia ZQ, Li Haomin, Lv Xudong. Automatic Annotation of Temporal Information in Chinese Medical Narrative Records. Chinese Journal of Biomedical Engineering. 2012 ;31(3):434-9.
25. Saurı R, Littman J, Knippen B, Gaizauskas R, Setzer A, Pustejovsky J. TimeML annotation guidelines, version 1.2. 2006.
26. Ferro L, Gerber L, Mani I, Sundheim B, Wilson G. TIDES 2005 standard for the annotation of temporal expressions. Technical report, MITRE; 2005.
27. Sun W, Rumshisky A, Uzuner O. Annotating temporal information in clinical narratives. Journal of biomedical informatics. 2013 ;46:S5-S12.
28. Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions. John Wiley & Sons; 2013.