TEES 2.1: Automated Annotation Scheme Learning in the BioNLP 2013 Shared Task

Jari Björne and **Tapio Salakoski**
Department of Information Technology, University of Turku
Turku Centre for Computer Science (TUCS)
Joukahaisenkatu 3-5, 20520 Turku, Finland
firstname.lastname@utu.fi

Abstract

We participate in the BioNLP 2013 Shared Task with Turku Event Extraction System (TEES) version 2.1. TEES is a support vector machine (SVM) based text mining system for the extraction of events and relations from natural language texts. In version 2.1 we introduce an automated annotation scheme learning system, which derives task-specific event rules and constraints from the training data, and uses these to automatically adapt the system for new corpora with no additional programming required. TEES 2.1 is shown to have good generalizability and good performance across the BioNLP 2013 task corpora, achieving first place in four out of eight tasks.

1 Introduction

Biomedical event extraction concerns the detection of statements of biological relations from scientific texts. Events are a formalism for accurately annotating the content of any natural language sentence. They are characterized by typed, directed arguments, annotated trigger words and the ability to nest other events as arguments, leading to flexible, complex structures. Compared to the more straightforward approach of binary relation extraction, the aim of event extraction is to utilize the added complexity to more accurately depict the content of natural language statements and to produce more detailed text mining results.

The BioNLP Shared Task is the primary forum for international evaluation of different event extraction technologies. Organized for the first time in 2009, it has since been held in 2011 and now in 2013 (Kim et al., 2009; Kim et al., 2011). Starting from the single GENIA corpus on NF-kB, it has since been extended to varied domain tasks, such as epigenetics and bacteria-host interactions. The theme of the 2013 task is "knowledge base construction", defining several domain tasks relevant for different aspects of this overall goal.

The Turku Event Extraction System (TEES)[1] is a generalized biomedical text mining tool, developed at University of Turku and characterized by the use of a unified graph representation and a stepwise machine learning approach based on support vector machines (SVM). TEES has participated in all BioNLP Shared Tasks, achieving first place in 2009, first place in four out of eight tasks in 2011 and now in 2013 again first place in four out of eight tasks (Björne et al., 2011; Björne et al., 2012). It has been available as an open source project since 2009, and has also been used by other research groups (Jamieson et al., 2012; Neves et al., 2013).

The BioNLP Shared Tasks have recorded the progress of various event extraction approaches. Where TEES 1.0 achieved an F-score of 51.95% in 2009, in 2011 the best performing system by team FAUST on the extended, but similar GENIA task achieved an F-score of 56.0% (Riedel et al., 2011). Interesting approaches have been demonstrated also in the interim of the Shared Tasks, for example with the EventMine system of Miwa et al. (2010) achieving an F-score of 56.00% on the 2009 GENIA corpus, and with the extremely computationally efficient system of Bui et al. (2012) based on automatically learning extraction rules from event templates. The GENIA task of 2013 has been considerably extended and the scope of the corpus is different, so a direct comparison with the earlier GENIA tasks is not possible.

In the BioNLP 2013 Shared Task the goal of the TEES project is to continue the generalization of event extraction techniques introduced in 2011 by fully automating task-specific adaptation via auto-

[1] http://jbjorne.github.com/TEES/

Proceedings of the BioNLP Shared Task 2013 Workshop, pages 16–25,
Sofia, Bulgaria, August 9 2013. ©2013 Association for Computational Linguistics

mated learning of event annotation rules. As an open source project TEES should also be easily applicable by any team interested in this task, so TEES 2.1 analyses were provided for all interested participants during the system development phase of the competition.

2 Methods

2.1 Turku Event Extraction System 2.1

TEES is a machine-learning based tool for extracting text-bound graphs from natural language articles. It represents both binary relations and events with a unified graph format where named entities and triggers are nodes and relations and event arguments are edges. This representation is commonly stored in the "interaction XML" format, an extensible XML representation applicable to various corpora (Björne et al., 2012; Pyysalo et al., 2008; Segura-Bedmar et al., 2013).

TEES approaches event extraction as a classification task, breaking the complex graph generation task into smaller steps that can be performed with multiclass classification. The SVM^{multiclass} support vector machine[2] (Tsochantaridis et al., 2005) with a linear kernel is used as the classifier in all machine learning steps.

To start with the BioNLP Shared Task, TEES conversion tools are used to convert the shared task format (txt/a1/a2) corpora into the interaction XML format. Equivalence annotations are resolved into independent events in the process.

Figure 1 shows an overview of the TEES event extraction process. In real-world applications, external programs are used to split sentences, detect protein/gene named entities and parse text, but in the BioNLP Shared Tasks these analyses are provided by the organizers. As in previous Shared Tasks, we used the tokenisations and the McCCJ parses converted into the collapsed CC-processed Stanford dependency scheme (Stenetorp et al., 2013; McClosky, 2010).

With the preprocessing done, TEES uses three primary processing steps to detect events. First, event trigger words are detected by classifying each non-named entity token into one of the trigger classes or as a negative. Then, for each (optionally directed) pair of named entity and trigger nodes a relation/argument edge candidate

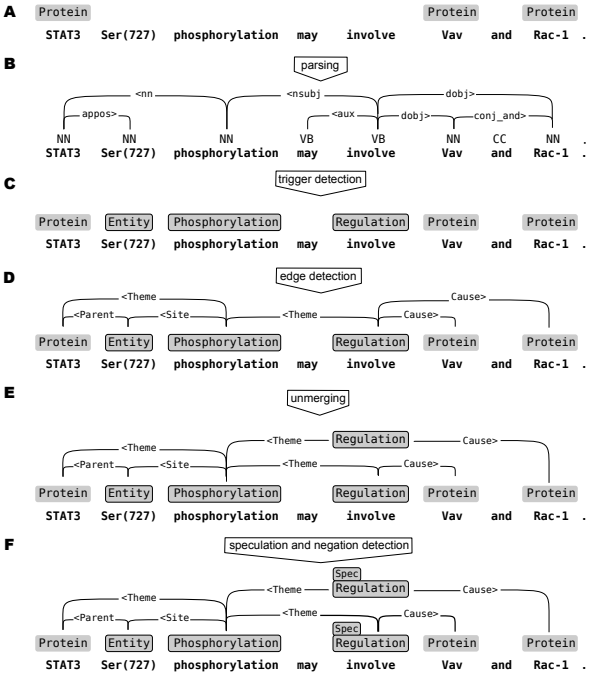

Figure 1: TEES event extraction process. Preprocessing steps A–C are achieved in the shared task with data provided by organizers. Event extraction steps D–F are all performed as consecutive, independent SVM classification steps. (Adapted from Björne et. al (2012).)

is generated and classified into one of the relation/argument classes or as a negative. Finally, for each event trigger node, for each valid set of outgoing argument edges an *unmerging* example is generated and classified as a true event or not, separating overlapping events into structurally valid ones. For tasks where events can have modifiers, a final modifier detection step can be performed. To better fit the trigger detection step into the overall task, a recall adjustment parameter is experimentally determined to increase the amount of triggers generated before edges are detected. The feature representations and basic approach of the system are largely unchanged from the 2011 entry, and for a more detailed overview we refer to Björne et. al (2012).

The main change in TEES 2.1, described in this paper, is the automated annotation scheme learning system, which enables the optimal use of the system on any interaction XML format corpus. This preprocessing step results in an annotation scheme definition which is used throughout the machine learning steps and the impact of which is described in detail in the following sections.

[2] http://svmlight.joachims.org/svm_multiclass.html

2.2 Automated Annotation Scheme Learning

In previous versions of TEES, task specific rules needed to be defined in code. The most important of these were the event annotation schemes of each task, which define the type and number of arguments that are valid for each event type. This limited straightforward application of TEES only to corpora that were part of the shared tasks. In TEES 2.1, the event scheme rules and constraints are learned automatically. All event types and argument combinations seen in the known training data are considered valid for the current task. The result of this analysis for the GE (GENIA) task is shown in Table 1.

The automatically generated annotation scheme analysis lists all entities, events, relations and modifiers detected in the corpus. Entities are simply a type of node and relations can be directed or undirected but are always defined as a single edge connecting two nodes. Events consist of a trigger node whose type is equal to the type of the event itself and a set of arguments, for which are defined also valid argument counts.

The interaction XML graph format represents both event arguments and binary relations as edge elements. To distinguish these annotations, a prerequisite for automated detection of valid event structures, elements that are part of events are labeled as such in the TEES 2.1 interaction XML graph. Those node and argument types that are not annotated also for the test set become the prediction targets, and the rest of the annotation can be used as known data to help in predicting them.

The annotation scheme analysis is stored in the TEES model file/directory, is available at runtime via a class interface and is used in the machine learning steps to enforce task-specific constraints. The availability of the learned annotation scheme impacts mostly the edge and unmerging detectors.

2.3 TEES 2.1 Edge Detection

The primary task specific specialization required in TEES 2.0 was the set of rules defining valid node combinations for edges. TEES detects edges (relations or arguments) by defining one edge candidate for each directed (or undirected) pair of nodes. While the system could be used without task-specific specialization to generate edge candidates for all pairs, due to the potentially large number of nodes in event-containing sentences this approach led to an inflated amount of negatives and reduced SVM performance. In the BioNLP Shared Task, e.g. the common *Protein* entities can only ever have incoming edges, so even such a simple limitation could considerably reduce the amount of edge candidates, but these task-specific rules had to be written into the Python-code. With the automatically learned annotation scheme, the edge detector checks for each node pair whether it constitutes a valid edge candidate as learned from the training data, automating and generalizing this task-specific optimization.

2.4 TEES 2.1 Unmerging

The TEES module most affected by the learned annotation scheme is the unmerging detector, which takes the merged event graph (where overlapping events share the same trigger node) and attempts to define which node/argument combinations constitute valid events (See Figure 1 E). One example is generated for each potential event, and nodes and edges are duplicated as needed for those classified as positives. In TEES 2.0, only the GE (GENIA), EPI (Epigenetics and Post-translational Modifications) and ID (Infectious Diseases) tasks from 2009 and 2011 were supported, with valid argument combinations defined in the code. In TEES 2.1 invalid argument combinations, as determined by the learned annotation scheme, are automatically removed before classification. Even if an event is structurally valid, it may of course not be a correct event, but reducing the number of negatives by removing invalid ones is an important optimization step also in the case of unmerging classification.

2.5 Unified site-argument representation

Representing the BioNLP Shared Task *site*-arguments in the interaction XML format has been problematic. The sites are arguments of arguments, linking a separate site-entity to a primary argument. In the graph format all arguments are edges, and while technically all edges could be defined as having a central node to which site-arguments could connect, this would result in a multi-step edge detection system, where site-argument edges could only be predicted after primary argument edges are predicted. To avoid this situation, in TEES 2.0 site arguments were defined as edges connecting the site entity either to the protein node (See Figure 2 A) or to the trigger node (See Figure 2 B). The second case was the most straightforward, and we assume closest to the

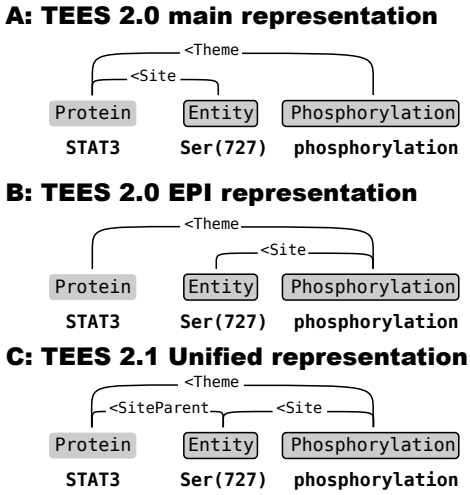

Figure 2: A unified representation (C) is introduced for site-arguments, replacing the different TEES 2.0 representations and enabling site-arguments to be processed as any other event arguments.

syntactic structure, as demonstrated by the good performance on the 2011 EPI task (Björne et al., 2012). However, in tasks where events can have multiple primary arguments, the approach shown in Fig. 2 B becomes problematic, as a primary/site argument pair cannot be determined unambiguously. In the approach shown in Fig. 2 A, the connection between the event and the site argument is indirect, meaning that the TEES 2.1 automated annotation scheme learning system cannot determine valid site argument constraints for events.

In TEES 2.1 this problem is solved with a unified approach where regardless of task, the site arguments are comparable to primary argument edges in all aspects, enabling consistent event analysis and simplifying site argument processing (See Figure 2 C). Additional *SiteParent* edges are defined to connect the entity and the protein it belongs to. In ambiguous cases, these are used to connect the right site to the right primary argument when converting to the final Shared Task format.

2.6 Validating final predictions

The current implementation of the automated annotation scheme learning system in TEES 2.1 has a shortcoming occasionally resulting in invalid event structures being produced. Consider an event with multiple optional arguments, such as *Cell_differentiation* from the CG task with 0–1 *AtLoc* arguments and 0–1 *Theme* arguments. While it can be possible that such an event can exist without any arguments at all, it is often the case that at least one of the optional arguments must be present. This is not detected by the current system, and would require the addition of learning rules for such groups of mandatory arguments.

The result of this and other small limitations in conforming to task rules is the occasional invalid predicted event. The Shared Task test set evaluation servers will not accept any invalid events, so these errors had to be resolved in some way. As this problem was detected at a late stage in the shared task, there was no more time to fix the underlying causes. However, these errors could not either be fixed by looking at the test set and correcting the events preventing the acceptance of the submission, as that would result in *de facto* manual annotation of the test set and an information leak. Therefore, we never looked at the document triggering the error, and used the following, consistent approach to resolve the invalid events. If the server would both report an invalid argument and a missing argument for the same event, the invalid argument was first replaced with the missing one. This was only the case with the GRN task. If the server would only report an invalid argument, we first removed the argument, and if this did not resolve the conflict, we removed the entire event. Following this, all events recursively pointing to removed invalid events were also removed. This approach could be implemented with a system processing the validation tools' output, but the better approach which we aim to pursue is to fix the limitations of the automated annotation scheme learning system, thus producing a tool usable on any corpora. In practice only a few invalid events were produced for each task where they occurred, so the impact on performance is likely to be negligible.

2.7 Public dataset

TEES 2.0, published in summer 2012 was a potentially useful tool for the BioNLP 2013 Shared Task, but at the same time required specific code extensions to be adapted for the task, leading to a situation where the program was available, but was not likely to be of practical value with new corpora. To resolve this problem the automated annotation scheme learning system was developed, taking the generalization approaches developed for the 2011 task and making them automatically applicable for new corpora. As using TEES can still

be difficult for people not familiar with the system, and as re-training the program is quite time consuming, we also published our event predictions for the 2013 task during the system development period, for other teams to make use of. Development set analyses were made available on February 26th, and test set analyses during the test period on April 13th. With only a few downloads, the data did not enjoy wide popularity, and due to the complexity of the tasks utilizing the data in other systems could very well have been too time consuming. TEES was also used to produce public analyses for the DDIExtraction 2013 Shared Task, where the data was used more, maybe due to easier integration into a binary relation extraction task (Segura-Bedmar et al., 2013; Björne et al., 2013).

3 Tasks and Results

TEES 2.1 could be applied as is to almost all the 2013 tasks with no task specific development required. Only subtask 1 of the Bacteria Biotopes task, concerning the assignment of ontology concepts, falls outside the scope of the current system. TEES 2.1 was the system to participate in most tasks, with good general performance, demonstrating the utility of abstracting away task-specific details. Official results for each task are shown in Table 2 and system performance relative to other entries in Figure 3.

Task	#	R	P	F	SER
GE	2/10	46.17	56.32	50.74	
CG	1/6	48.76	64.17	55.41	
PC	2/2	47.15	55.78	51.10	
GRO	1/1	15.22	36.58	21.50	
GRN	3/5	33	78	46	0.86
BBT1	0/4				
BBT2	1/4	28	82	42	
BBT3	1/2	12	18	14	

Table 2: Official test set results for the BioNLP 2013 tasks. Performance is shown in (R)ecall, (P)recision and (F)-score, and also SER for the GRN task. BB task 1 falls outside the scope of TEES 2.1. Rank is indicated by #.

3.1 GENIA (GE)

The GENIA task is the central task of the BioNLP Shared Task series, having been organized in all three Shared Tasks. It has also enjoyed the largest number of contributions and as such could be viewed as the primary task for testing different event extraction approaches. In 2013 the GENIA task annotation has been considerably extended and the coreference annotation that in 2011 formed its own supporting task is integrated in the main GENIA corpus (Kim et al., 2013a).

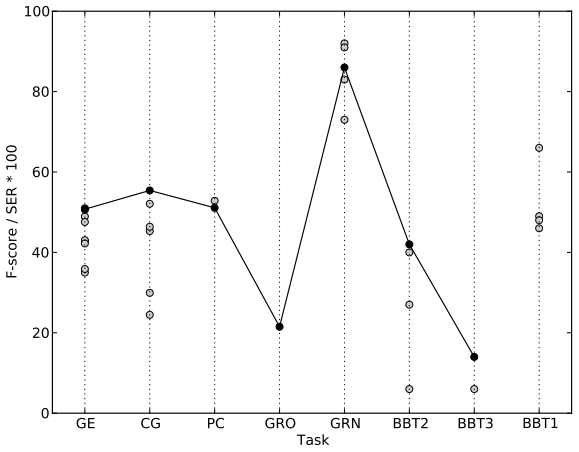

Figure 3: Performance of the systems participating in the BioNLP'13 Shared Task. Our results are marked with black dots. Please note that the performance metric for tasks GRN and BBT1 is SER*100, where a smaller score is better.

The GENIA task is a good example for demonstrating the usefulness of automatically learning the event annotation scheme. The task uses 11 different event types, pairwise binary coreference relations and modality annotation for both speculation and negation. Previous versions of TEES would have encoded all of this information in the program, but with TEES 2.1 the annotation rules are detected automatically and stored in a separate datafile external to the program. Table 1 shows the automatically learned event scheme. It should however be noted that while the learned scheme accurately describes the known annotation, it may not exactly correspond to the corpus annotation rules. For example, the *Binding* event, when learned from the data, can have one or two *Theme* arguments, when in the official rules it simply has one or more *Theme* arguments.

In some GENIA Coreference relations (45 out of 338 in train and devel data) at least one of the endpoints is an event trigger. While such relations could indeed be linked to event trigger nodes, TEES makes no distinction between triggers and events and would link them to the event annotation when converting back to the Shared Task format,

so we chose to skip them.

TEES 2.1 achieved a performance of 50.74%, placing second in the GENIA task. The first place was reached by team EVEX (Hakala et al., 2013), with a system that utilizes the publicly available TEES 2.1 program. This result further highlights the value of open sourcing scientific code and underlines the importance of incorporating existing solutions into future systems.

3.2 Cancer Genetics (CG)

The CG task is a domain-specific event extraction task targeting the recovery of information related to cancer (Pyysalo et al., 2013; Pyysalo et al., 2012). It is characterized by a large number of entity and event types. Despite a heterogeneous annotation scheme, TEES 2.1 achieved a performance of 55.41% F-score, placing first in this task. On some event categories TEES achieved a performance notably higher than usual for it in event extraction tasks, such as the 77.20% F-score for the Anatomy-group events. The impact of more common, and as such more easily detected classes on the micro-averaged F-score is certainly important, but it is interesting to speculate that maybe the very detailed annotation scheme led to a more focused and thus more consistent annotation, making machine learning easier on this task.

3.3 Pathway Curation (PC)

The PC task aims to produce events suitable for pathway curation (Ohta et al., 2013). Its extraction targets are based on existing pathway models and ontologies such as the Systems Biology Ontology (SBO). The dataset has only a few entity types, but similar to the CG task, a large number of event types. With 51.10% F-score TEES 2.1 placed second, behind team NaCTeM by 1.74 percentage points (Miwa and Ananiadou, 2013). On the CG task team NaCTeM placed second, 3.32 percentage points lower than TEES 2.1. Even with the only two participants in the PC task having very close performance, compared to the results of the same teams on the CG task, we speculate the PC and CG tasks are of similar complexity.

3.4 Gene Regulation Ontology (GRO)

The GRO task concerns the automatic annotation of documents with Gene Regulation Ontology (GRO) concepts (Kim et al., 2013b). The annotation is very detailed, with 145 entity and 81 event types. This results in a large number of small classes which are independent in SVM classification and thus hard to learn. TEES did not detect most of the small classes, and generally, the larger the class, the higher the performance. It is possible that classification performance might be improved by merging some of the smaller classes and disambiguating the predictions with a rule-based step, similar to the TEES approach in the EPI 2011 task.

Overall performance was at 21.50% F-score but as TEES 2.1 was the only system in this task, not many conclusions can be drawn from it. However, the system was also exactly the same as applied in the other tasks. With decent performance on some of the larger classes, we speculate that with a larger training corpus, and with a system adapted for the GRO task, performance comparable to the GE, CG and PC tasks could be reached.

3.5 Gene Regulation Network (GRN)

GRN is a task where event extraction is utilized as an optional, intermediate step in the construction of a large regulation network (Bossy et al., 2013a). The annotation consists of 11 entity types, 12 binary relation types and a single *Action* event type. The predicted events can be automatically converted to the regulation network, or the network can be produced by other means. In either case, the final evaluation is performed on the network, using the Slot Error Rate (SER) metric (Makhoul et al., 1999), where lower is better and a value of less than one is expected for decent predictions.

TEES 2.1 produced the event format submission, and with conversion to the regulation network achieved an SER of 0.86, placing in the middle of the five teams, all of which had an SER of less than one. A downloadable evaluator program was provided early enough in the development period to be integrated in TEES 2.1, allowing direct optimization against the official task metrics. As SER was a metric not used before with TEES, the relaxed F-score was instead chosen as the optimization target, with the assumption that it would provide a predictable result also on the hidden test set. In training it was also observed that the parameters for the optimal relaxed F-score also produced the optimal SER result.

3.6 Bacteria Biotopes (BB)

Along with the GENIA task, the BB task is the only task to continue from earlier BioNLP Shared Tasks. The BB task concerns the detection of statements about bacteria habitats and relevant en-

vironmental properties and is divided into three subtasks (Bossy et al., 2013b).

In task 1 the goal is to detect boundaries of bacteria habitat entities and for each entity, assign one or more terms from 1700 concepts in the OntoBiotope ontology. While the TEES entity detector could be used to detect the entities, assigning the types falls outside the scope of the system, and is not directly approachable as the sort of classification task used in TEES. Therefore, BB task 1 was the only task for which TEES 2.1 was not applied.

BB tasks 2 and 3 are a direct continuation of the 2011 BB task, with the goal being extraction of relations between bacteria entities and habitat and geographical places entities. Only three entity and two relation types are used in the annotation. In task 2 all entities are provided and only relations are detected, in task 3 also the entities must be predicted. The BB task was the only 2013 task in which we used (limited) task specific resources, as TEES 2.0 resources developed for the 2011 BB task were directly applicable to the 2013 tasks. A dictionary of bacteria name tokens, derived from the List of Prokaryotic names with Standing in Nomenclature[3] (Euzéby, 1997) was used to improve entity detection performance. Unlike the 2011 task, WordNet features were not used.

TEES 2.1 achieved F-scores of 42% and 14% for tasks 2 and 3 respectively, reaching first place in both tasks. The low overall performance is however indicative of the complexity of these tasks.

4 Conclusions

We applied TEES version 2.1 to the BioNLP 2013 Shared Task. An automated annotation scheme learning system was built to speed up development and enable application of the system to novel event corpora. The system could be used as is in almost all BioNLP 2013 tasks, achieving good overall performance, including several first places.

The GRO task highlighted the limitations of a purely classification based approach in situations with very many small classes, in a sense the same issue as with the ontology concept application in BB task 1. Despite these minor limitations, the basic stepwise SVM based approach of TEES continues to demonstrate good generalization ability and high performance.

We made our system public during the task development phase and provided precalculated analyses to all participants. While we consider it unfortunate that these analyses did not enjoy greater popularity, we are also looking forward to the varied approaches and methods developed by the participating teams. However, the encouraging results of the GENIA task, not to mention earlier positive reports on system combination (Kano et al., 2011; Riedel et al., 2011) indicate that there is untapped potential in merging together the strong points of various systems.

TEES 2.1 had very good performance on many tasks, but it must be considered that as an established system it was already capable of doing much of the basic processing that many other teams had to develop for their approaches. In particular, previous BioNLP Shared Tasks have shown that the TEES internal micro-averaged edge-detection F-score provides a very good approximation of the official metrics of most tasks. It is unfortunate that official evaluator programs were only available in some tasks, and often only at the end of the development period, potentially leading to a situation where different teams were optimizing for different goals. In our opinion it is of paramount importance that in shared tasks not only the official evaluation metric is known well ahead of time, but a downloadable evaluator program is provided, as the complexity of the tasks means that independent implementations of the evaluation metric are error prone and an unnecessary burden on the participating teams.

As with previous versions of TEES, the 2.1 version is publicly available both as a downloadable program and as a full, open source code repository. We intend to continue developing TEES, and will hopefully in the near future improve the automated annotation learning system to overcome its current limitations. We find the results of the BioNLP 2013 Shared Task encouraging, but as with previous iterations, note that there is still a long way to go for truly reliable text mining. We think more novel approaches, better machine learning systems and careful utilization of the research so far will likely lead the field of biomedical event extraction forward.

Acknowledgments

We thank CSC — IT Center for Science Ltd, Espoo, Finland for providing computational resources.

[3] http://www.bacterio.cict.fr/

Type	Name	Arguments
ENTITY	Anaphora	
ENTITY	Entity	
ENTITY	Protein	
EVENT	Binding	Site[0,1](Entity) / Theme[1,2](Protein)
EVENT	Gene_expression	Theme[1,1](Protein)
EVENT	Localization	Theme[1,1](Protein) / ToLoc[0,1](Entity)
EVENT	Negative_regulation	Cause[0,1](Acetylation, Binding, Gene_expression, Negative_regulation, Phosphorylation, Positive_regulation, Protein, Protein_catabolism, Regulation, Ubiquitination) / Site[0,1](Entity) / Theme[1,1](Binding, Gene_expression, Localization, Negative_regulation, Phosphorylation, Positive_regulation, Protein, Protein_catabolism, Regulation, Transcription, Ubiquitination)
EVENT	Phosphorylation	Cause[0,1](Protein) / Site[0,1](Entity) / Theme[1,1](Protein)
EVENT	Positive_regulation	Cause[0,1](Acetylation, Binding, Gene_expression, Negative_regulation, Phosphorylation, Positive_regulation, Protein, Protein_catabolism, Regulation, Ubiquitination) / Site[0,1](Entity) / Theme[1,1](Binding, Deacetylation, Gene_expression, Localization, Negative_regulation, Phosphorylation, Positive_regulation, Protein, Protein_catabolism, Protein_modification, Regulation, Transcription, Ubiquitination)
EVENT	Protein_catabolism	Theme[1,1](Protein)
EVENT	Protein_modification	Theme[1,1](Protein)
EVENT	Regulation	Cause[0,1](Binding, Gene_expression, Localization, Negative_regulation, Phosphorylation, Positive_regulation, Protein, Protein_modification, Regulation) / Site[0,1](Entity) / Theme[1,1](Binding, Gene_expression, Localization, Negative_regulation, Phosphorylation, Positive_regulation, Protein, Protein_catabolism, Protein_modification, Regulation, Transcription)
EVENT	Transcription	Theme[1,1](Protein)
EVENT	Ubiquitination	Cause[0,1](Protein) / Theme[1,1](Protein)
RELATION	Coreference, directed	Subject(Anaphora) / Object(Anaphora, Entity, Protein)
RELATION	SiteParent, directed	Arg1(Entity) / Arg2(Protein)
MODIFIER	negation	Binding, Gene_expression, Localization, Negative_regulation, Phosphorylation, Positive_regulation, Protein_catabolism, Regulation, Transcription
MODIFIER	speculation	Binding, Gene_expression, Localization, Negative_regulation, Phosphorylation, Positive_regulation, Protein_catabolism, Regulation, Transcription, Ubiquitination
TARGET	ENTITY	Acetylation, Anaphora, Binding, Deacetylation, Entity, Gene_expression, Localization, Negative_regulation, Phosphorylation, Positive_regulation, Protein_catabolism, Protein_modification, Regulation, Transcription, Ubiquitination
TARGET	INTERACTION	Cause, Coreference, Site, SiteParent, Theme, ToLoc

Table 1: Automatically learned GENIA 2013 task event annotation scheme. The *entities* are the nodes of the graph. *Targets* define the types of nodes and edges to be automatically extracted. *Events* and *relations* are defined by their type and arguments. Relations are optionally directed, and always have two arguments, with specific valid target node types. Events can have multiple arguments, and in addition to valid target node types, the minimum and maximum amount of each argument per event are defined. *Modifiers* are binary attributes defined by their type and the types of nodes they can be defined for.

References

Jari Björne, Juho Heimonen, Filip Ginter, Antti Airola, Tapio Pahikkala, and Tapio Salakoski. 2011. Extracting Contextualized Complex Biological Events with Rich Graph-Based Feature Sets. *Computational Intelligence, Special issue on Extracting Biomolecular Events from Literature*. Accepted in 2009.

Jari Björne, Filip Ginter, and Tapio Salakoski. 2012. University of Turku in the BioNLP'11 Shared Task. *BMC Bioinformatics*, 13(Suppl 11):S4.

Jari Björne, Suwisa Kaewphan, and Tapio Salakoski. 2013. UTurku: Drug Named Entity Detection and Drug-drug Interaction Extraction Using SVM Classification and Domain Knowledge. In *Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013)*.

Robert Bossy, Philippe Bessiŕes, and Claire Nédellec. 2013a. BioNLP shared task 2013 - an overview of the genic regulation network task. In *Proceedings of BioNLP Shared Task 2013 Workshop*, Sofia, Bulgaria, August. Association for Computational Linguistics.

Robert Bossy, Wiktoria Golik, Zorana Ratkovic, Philippe Bessiŕes, and Claire Nédellec. 2013b. BioNLP shared task 2013 - an overview of the bacteria biotope task. In *Proceedings of BioNLP Shared Task 2013 Workshop*, Sofia, Bulgaria, August. Association for Computational Linguistics.

Quoc-Chinh Bui and Peter M.A. Sloot. 2012. A robust approach to extract biomedical events from literature. *Bioinformatics*, 28(20):2654–2661, October.

Jean Paul Marie Euzéby. 1997. List of Bacterial Names with Standing in Nomenclature: a Folder Available on the Internet. *Int J Syst Bacteriol*, 47(2):590–592.

Kai Hakala, Sofie Van Landeghem, Tapio Salakoski, Yves Van de Peer, and Filip Ginter. 2013. EVEX in ST'13: Application of a large-scale text mining resource to event extraction and network construction. In *Proceedings of BioNLP Shared Task 2013 Workshop*, Sofia, Bulgaria, August. Association for Computational Linguistics.

Daniel G. Jamieson, Martin Gerner, Farzaneh Sarafraz, Goran Nenadic, and David L. Robertson. 2012. Towards semi-automated curation: using text mining to recreate the hiv-1, human protein interaction database. *Database*, 2012.

Yoshinobu Kano, Jari Björne, Filip Ginter, Tapio Salakoski, Ekaterina Buyko, Udo Hahn, K Bretonnel Cohen, Karin Verspoor, Christophe Roeder, Lawrence Hunter, Halil Kilicoglu, Sabine Bergler, Sofie Van Landeghem, Thomas Van Parys, Yves Van de Peer, Makoto Miwa, Sophia Ananiadou, Mariana Neves, Alberto Pascual-Montano, Arzucan Ozgur, Dragomir Radev, Sebastian Riedel, Rune Saetre, Hong-Woo Chun, Jin-Dong Kim, Sampo Pyysalo, Tomoko Ohta, and Jun'ichi Tsujii. 2011. U-compare bio-event meta-service: compatible bionlp event extraction services. *BMC Bioinformatics*, 12(1):481.

Jin-Dong Kim, Tomoko Ohta, Sampo Pyysalo, Yoshinobu Kano, and Jun'ichi Tsujii. 2009. Overview of BioNLP'09 Shared Task on Event Extraction. In *Proceedings of the BioNLP 2009 Workshop Companion Volume for Shared Task*, pages 1–9, Boulder, Colorado. ACL.

Jin-Dong Kim, Sampo Pyysalo, Tomoko Ohta, Robert Bossy, and Jun'ichi Tsujii. 2011. Overview of BioNLP Shared Task 2011. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task*, Portland, Oregon, June. Association for Computational Linguistics.

Jin-Dong Kim, Yue Wang, and Yamamoto Yasunori. 2013a. The genia event extraction shared task, 2013 edition - overview. In *Proceedings of BioNLP Shared Task 2013 Workshop*, Sofia, Bulgaria, August. Association for Computational Linguistics.

Jung-Jae Kim, Xu Han, Vivian Lee, and Dietrich Rebholz-Schuhmann. 2013b. GRO task: Populating the gene regulation ontology with events and relations. In *Proceedings of BioNLP Shared Task 2013 Workshop*, Sofia, Bulgaria, August. Association for Computational Linguistics.

John Makhoul, Francis Kubala, Richard Schwartz, and Ralph Weischedel. 1999. Performance measures for information extraction. In *Proceedings of DARPA Broadcast News Workshop*, pages 249–252.

David McClosky. 2010. *Any domain parsing: automatic domain adaptation for natural language parsing*. Ph.D. thesis, Department of Computer Science, Brown University.

Makoto Miwa and Sophia Ananiadou. 2013. NaCTeM EventMine for BioNLP 2013 CG and PC tasks. In *Proceedings of BioNLP Shared Task 2013 Workshop*, Sofia, Bulgaria, August. Association for Computational Linguistics.

Makoto Miwa, Sampo Pyysalo, Tadayoshi Hara, and Jun'ichi Tsujii. 2010. A comparative study of syntactic parsers for event extraction. In *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing*, BioNLP '10, pages 37–45, Stroudsburg, PA, USA. Association for Computational Linguistics.

Mariana Neves, Alexander Damaschun, Nancy Mah, Fritz Lekschas, Stefanie Seltmann, Harald Stachelscheid, Jean-Fred Fontaine, Andreas Kurtz, and Ulf Leser. 2013. Preliminary evaluation of the cellfinder literature curation pipeline for gene expression in kidney cells and anatomical parts. *Database*, 2013.

Tomoko Ohta, Sampo Pyysalo, Rafal Rak, Andrew Rowley, Hong-Woo Chun, Sung-Jae Jung, Sung-Pil Choi, and Sophia Ananiadou. 2013. Overview of the pathway curation (PC) task of bioNLP shared task 2013. In *Proceedings of BioNLP Shared Task 2013 Workshop*, Sofia, Bulgaria, August. Association for Computational Linguistics.

Sampo Pyysalo, Antti Airola, Juho Heimonen, Jari Björne, Filip Ginter, and Tapio Salakoski. 2008. Comparative analysis of five protein-protein interaction corpora. *BMC Bioinformatics*, 9(Suppl 3):S6.

Sampo Pyysalo, Tomoko Ohta, Makoto Miwa, Han-Cheol Cho, Jun'ichi Tsujii, and Sophia Ananiadou. 2012. Event extraction across multiple levels of biological organization. *Bioinformatics*, 28(18):i575–i581.

Sampo Pyysalo, Tomoko Ohta, and Sophia Ananiadou. 2013. Overview of the cancer genetics (CG) task of bioNLP shared task 2013. In *Proceedings of BioNLP Shared Task 2013 Workshop*, Sofia, Bulgaria, August. Association for Computational Linguistics.

Sebastian Riedel, David McClosky, Mihai Surdeanu, Andrew McCallum, and Christopher D. Manning. 2011. Model combination for event extraction in bionlp 2011. In *Proceedings of the BioNLP Shared Task 2011 Workshop*, BioNLP Shared Task '11, pages 51–55, Stroudsburg, PA, USA. Association for Computational Linguistics.

Isabel Segura-Bedmar, Paloma Martínez, and Maria Herrero-Zazo. 2013. SemEval-2013 Task 9: Extraction of Drug-Drug Interactions from Biomedical Texts. In *Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013)*.

Pontus Stenetorp, Wiktoria Golik, Thierry Hamon, Donald C. Comeau, Rezarta Islamaj Dogan, Haibin Liu, and W. John Wilbur. 2013. BioNLP shared task 2013: Supporting resources. In *Proceedings of BioNLP Shared Task 2013 Workshop*, Sofia, Bulgaria, August. Association for Computational Linguistics.

Ioannis Tsochantaridis, Thorsten Joachims, Thomas Hofmann, and Yasemin Altun. 2005. Large margin methods for structured and interdependent output variables. *Journal of Machine Learning Research (JMLR)*, 6(Sep):1453–1484.