Semantic Interpretation of Events in Live Soccer Commentaries

Anne-Lyse Minard[1,2], Manuela Speranza[1], Bernardo Magnini[1], Mohammed R. H. Qwaider[1]
[1] Fondazione Bruno Kessler, Trento, Italy
[2] Dept. of Information Engineering, University of Brescia, Italy
{minard,manspera,magnini,qwaider}@fbk.eu

Abstract

English. In the context of semantic interpretation of live soccer commentaries in Italian, we propose an annotation schema for relevant events and their argument structure, on whose basis we annotated a reference evaluation corpus. We investigated automatic event classification and used Active Learning to reduce the cost of acquiring domain-specific training data.

Italiano. *Nel contesto dell'interpretazione di commenti calcistici in diretta, proponiamo uno schema per l'annotazione di eventi (e relativa struttura argomentativa), sulla base del quale abbiamo creato un corpus di valutazione di riferimento. Ci siamo occupati della classificazione automatica di eventi utilizzando Active Learning per ridurre lo sforzo per l'acquisizione di dati annotati specifici del dominio.*

1 Introduction

This work focuses on understanding the content of live commentaries of sport games. This form of written reporting has become very popular in recent years, and almost every national Italian online newspaper has a section dedicated to live sport commentaries. Live commentaries have several interesting properties: (i) they are short descriptions of an event written by professionals while the event is happening; their form is much simpler than a full spoken running commentary; (ii) they have a clear and simple structure, typically based on the timing of the sport event; (iii) they are often associated with metadata (e.g. *La Roma passa in vantaggio* [Roma takes the lead] is associated with the metadata GOAL); (iv) finally, they describe visual scenes, which is relevant to automatic alignment of multimedia content (e.g. align a sequence of frames in a video with the corresponding commentary), a topic of emerging interest in Computational Linguistics (see, for instance, (Song et al., 2016)). Our work is part of a larger cross-disciplinary project, Understanding Multimedia Content, currently involving several research groups at FBK.

In this paper we first define an annotation framework for the semantic interpretation of online soccer commentaries in Italian (Section 3), which includes the detection and classification of relevant events, as well as the identification of their argument structure. Based on this annotation schema, which could also be used for the annotation of tweets or other short online comments, we manually annotated a collection of commentaries in Italian to be used as a gold standard (Section 4). As a first step towards a comprehensive system for automatic interpretation of soccer events we focused on event detection and classification (i.e. event extraction), and used Active Learning to build a training corpus (Section 5). We show that this procedure is very effective, allowing our system to reach an F1 of 77.25, with considerable savings of annotation time (Section 6).

2 Related Work

Most of the work on event detection and classification focuses either on the news (UzZaman et al., 2012) or medical domains (Sun et al., 2013). For Italian, two corpora annotated with events following the It-TimeML framework (Caselli et al., 2011a) are available: EVENTI (Caselli et al., 2014) and WItaC (Speranza and Minard, 2015).

Event detection and classification on news has been of interest for English, Italian and Spanish in the TempEval evaluation campaigns (Verhagen et al., 2010; UzZaman et al., 2012) and for Italian in the EVENTI task at Evalita 2014 (Caselli et al., 2014). As part of these evaluation campaigns, several event extraction systems, mainly supervised, have been implemented (Caselli et al., 2011b; Jung and Stent, 2013; Bethard, 2013; Mirza and Minard, 2014). The development of supervised systems requires a significant amount of training data, whose creation is very time consuming. The effort needed to annotate these data can be reduced by using Active Learning methods, i.e. methods where instances to be annotated are selected according to their predicted impact on the model learned for a specific task. Active Learning has been used in various linguistic annotation tasks, such as Named Entity Recognition (Shen et al., 2004) and Part-of-Speech tagging (Ringger et al., 2007).

The surging interest of the NLP community for event detection and classification in the sport domain, on the other hand, is shown by the hackathon recently organized on extraction of soccer events from Tweets in French, English and Arabic (`http://hackatal.github.io/2016/`).

Fort and Claveau (2012) present a corpus of match commentaries and transcripts of video commentaries of soccer games in French, which has been annotated with entities (e.g. players, referees), events (e.g. corner, penalty) and some relations (e.g. pass, replace player) and van Oorschot et al. (2012) propose a method to extract relevant events of games in Dutch using the quantity of tweets posted per minute.

Event extraction in the sport domain is even more important as far as analysis of video (Xu et al., 2008; Han et al., 2008) and audio (Cabasson and Divakaran, 2003) data is concerned.

In the domain of automatic alignment of multimedia content, the analysis of both texts, videos and audio is necessary, and the research focuses on the alignment of the events detected in the three media (Malmaud et al., 2015; Regneri et al., 2013).

3 Task Definition and Annotation Framework

In our annotation framework, semantic interpretation of soccer events consists of the following steps: (i) soccer event recognition and classification, (ii) recognition and classification of the entities involved in the soccer event, and (iii) identification of the argument relations between the soccer event and the participant entities.

3.1 Event Recognition and Classification

Soccer event annotation is inspired by the It-TimeML definition of event and follows its minimal chunk rule, according to which only the head of the event phrase is included in the annotated text span (Caselli et al., 2011a). The main difference with the It-TimeML framework is that we restrict it to verbal and nominal events and to a semantically defined set of relevant events.

In particular, we identified six semantic categories of events relevant to the soccer domain (and a number of sub-categories):

`Referee decision` includes events that are characterized as such due to a referee's intervention; examples of subcategories are `Yellow card` and `Offside`;

`Kick` includes events in which the ball is kicked by a player; examples of subcategories are `Penalty`, `Corner`, `Pass` (e.g. *apre* in (1)), `Shot on goal`, and `Free kick`;

`Interruption` includes events in which a player interrupts the action of the opposing team examples of subcategories are `Clearance` and `Intercept` (e.g. *anticipato* in (1));

`Possession` includes events where the ball, although moving, does not go from one player to another; as subcategories we find, for example, `Dribbling` and `Holding possession`;

`Goal` includes events where a team scores (we did not devise subcategories for `Goal`);

`No ball` includes (i) events where a player doesn't have the ball (e.g. *inserimento* in (1)), and (ii) events not involving the ball, such as pushing or knocking to the ground (no subcategorization).

(1) 71: Griezmann passa a Pogba che *apre* per Matuidi, *inserimento* in area del centrocampista del Psg, che viene *anticipato*. [Griezman for Pogba who in turn passes to Matuidi, the Psg midfield player makes a forward run for the ball but gets beaten to it]

3.2 Entity Recognition and Classification

In order to annotate entities relevant to the soccer domain, we identified four categories, i.e. `Player`, `Team`, `Referee`, and `Coach`. Entities include both named entities (e.g. *Griezmann* and *Psg* in (1)) and nominal entities (e.g. *centrocampista* [middle field player] in (1)) and textual span is identified according to the minimal chunk rule (as was done for events).

3.3 Argument Structure Identification

The annotation of the argument structure of an event is performed through the creation of links called `ARG_rel` between each event and its arguments (which can be either entities or events). Inspired by PropBank (Bonial et al., 2010), we also defined four numbered arguments to be assigned to each `ARG_rel` in the form of an attribute: `ARG_0` and `ARG_1` correspond to the required arguments of a predicate, e.g. agent and patient respectively, while `ARG_2` and `ARG_3` correspond to arguments that occur with high-frequency for a certain predicate.

In (1), for instance, we have an `ARG_rel` between *passa* and *Griezmann* (`ARG_0`) and an `ARG_rel` between *passa* and *Pogba* (`ARG_2`).

4 Reference Annotated Corpus for Event Interpretation

Based on the annotation schema described in Section 3, we manually annotated a corpus of nine soccer games (five games from the Euro 2016 competition and four games from *Campionato di Serie A* 2015-2016) collected from La Repubblica,[1] Tuttosport,[2] and Eurosport.[3] Annotation was performed using the CAT tool (Bartalesi Lenzi et al., 2012). The result is a reference corpus for the evaluation of semantic interpretation of soccer events consisting of around 13,500 tokens, for a total of 1,372 annotated events and 1,600 argument relations (see Table 1).

We computed the inter-annotator agreement (IAA) over 46 commentaries annotated by two annotators (two halves from two different games). In terms of Dice's coefficient (Dice, 1945) we obtained an IAA of 0.70 and 0.96 (micro average) for event and entity classification respectively, and 0.69 for relation recognition (between events and entities marked by both annotators).

5 Event Extraction

In order to extract and classify soccer events in online commentaries, we used a supervised machine learning approach. We had a system for event detection (trained on news articles annotated following It-TimeML) available, which did not perform well on the soccer domain (it obtained an F1 of 40.8 and recall of 50.1 on our reference corpus).

[1] `http://www.repubblica.it/`
[2] `http://www.tuttosport.com/`
[3] `http://it.eurosport.com/`

	ref. corpus	training corpus
Games	9	101
Commentaries	652	1,377
Commentaries/game	72	14
Tokens	13,567	31,955
Tokens/com.	20.8	23.2
Goal	66	168
Kick	666	1,425
Interruption	274	390
Possession	71	181
Referee decision	254	807
No Ball Event	41	181
Player	1317	-
Referee	21	-
Coach	10	-
Team	291	-
ARG_rel	1,600	-

Table 1: Dataset statistics.

As a consequence, a training corpus specifically developed for this task was needed.

We therefore exploited the TEXTPRO-AL Active Learning platform (Magnini et al., 2016) which selects the most informative samples from an unlabeled set. More precisely, TEXTPRO-AL selects commentaries containing events that the system was not able to recognize correctly, pre-annotates them and asks the annotator to check them.

As illustrated in Figure 1, an AL cycle consists of the following steps:[4]

1. Train a model using the annotated commentaries[5] (step 3);
2. Repeat the following cycle until the batch[6] is full:
   (a) Select, from an unlabeled database of commentaries (see Section 5.1), a commentary that matches the first event string in the error queue[7] (i.e. the event with the lowest confidence) (step 4);
   (b) Pre-annotate the example (step 5);
   (c) Correct the annotation (done manually by an annotator) (step 1);
   (d) Add the annotated example to the batch (step 2a);

[4] The AL cycle is repeated until a stopping criteria is verified; for instance, until the system reaches a pre-defined performance.

[5] At the beginning the training corpus is empty, so the first commentary is randomly selected and added to the batch.

[6] The batch size was set to 2 for the first 24 examples and then to 10. These values were chosen to enable frequent retraining of the model and an update of the confidence scores and system errors.

[7] The error queue (or system global memory) contains the history of the system errors corrected by the annotator.

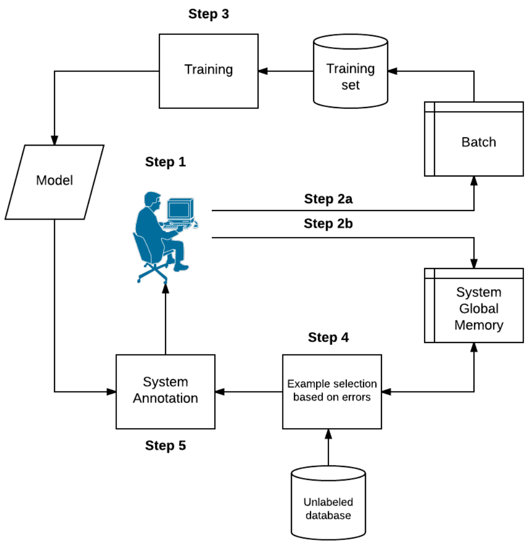

Figure 1: Active Learning schema adopted to build the training corpus.

(e) Save in the error queue the annotated events with their model confidence score (step 2b);

Our system is highly customizable: the event detection classification system can easily be substituted by a different system for different classification tasks, like NER and PoS tagging.

5.1 Unlabeled Database

The unlabeled database used in the AL procedure is composed of commentaries of 101 soccer games from DirettaGoal,[8] La Repubblica,[9] Tuttosport,[10] and Eurosport.[11] We extracted the online commentaries of all games of the Euro 2016 Cup and of the final 6 rounds of *Campionato di Serie A* 2015-2016. In total 6,573 commentaries were collected, with 155,005 tokens.

5.2 Error Selection

The error-based selection process exploits the idea that the corrections done by the annotator can be used to select new examples more efficiently. The system has a memory in which the events contained in the checked commentaries are stored, together with the system's confidence score and the indication of whether the system was right or wrong.

[8] http://www.direttagoal.it/
[9] http://www.repubblica.it/
[10] http://www.tuttosport.com/
[11] http://it.eurosport.com/

5.3 Event Detection and Classification

The system for event detection and classification is based on machine learning, using the SVM algorithm implemented in TinySVM and included in Yamcha (Kudo and Matsumoto, 2003). The task is treated as a multi-class classification task, where each token has to be classified in one of the 7 predefined classes.[12] The features used are those defined in the system of Mirza and Minard (2014), which took part in the EVENTI task at Evalita 2014 (Caselli et al., 2014), obtaining an F1 of 0.86 for the task of event detection and an F1 of 0.67 for event classification.

5.4 Annotation Editor

For the manual revision of linguistic annotations within the Active Learning method, we adapted an existing editor, MTEqual[13] (Girardi et al., 2014), originally developed for assessing the quality of machine translation.

6 Evaluation

The AL system described in the previous section was used by a non-expert annotator who annotated events in soccer commentaries for seven working days. This resulted in a training corpus of 1,377 commentaries, that is, around 200 commentaries per day (see Table 1).

The evaluation of our system was performed by comparing it to the reference annotated corpus described in Section 4. The learning curve in Figure 2 represents the results obtained by the system in terms of precision, recall and F1-measure as the training set was progressively extended. At the beginning the training set was empty, so the performance of the system was null. After the annotation of 200 commentaries, the system reached 53.27 F1, and after 800 commentaries it obtained 70.94 F1. At the end of our experiment, almost 1,400 commentaries had been annotated and the system's performance was 76.65 F1 (73.42 of recall and 80.16 of precision). The peak performance is 77.25 F1 and was reached with 1,347 commentaries (i.e. almost 32,000 tokens).

7 Conclusion and Future Work

We presented a new annotation framework for the interpretation of online soccer commentaries, as

[12] Referee decision, Kick, Interruption, Possession, Goal, No Ball Event and O for tokens that are not part of an event.
[13] https://github.com/hltfbk/MT-EQuAl

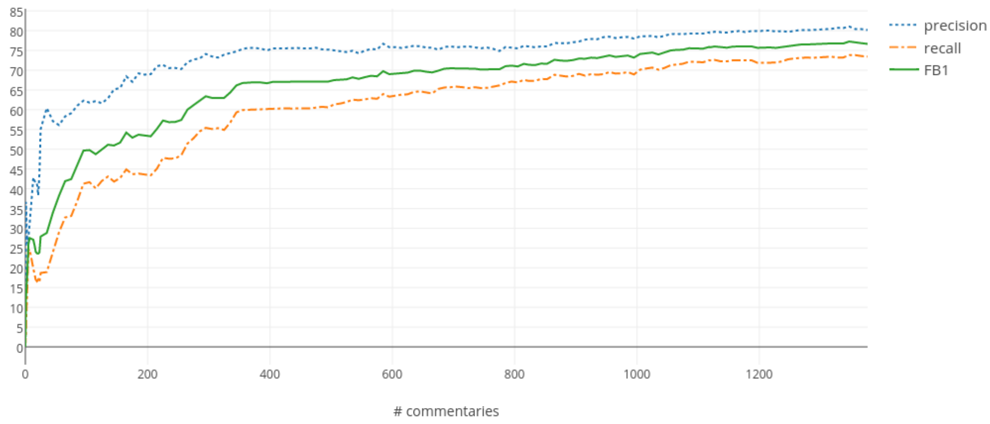

Figure 2: Event extraction performance as the training set was extended.

well as the reference annotated corpus we created.[14] We also described our system for event extraction from live soccer commentaries in Italian. It exploits the TEXTPRO-AL Active Learning platform, which allowed us to reach a significant F1 (77.25) in seven working days of a non-expert annotator. The annotation was performed for Italian but the method and the annotation schema we devised can be applied to other languages. The only language dependent component is the feature extractor used by the event detection module.

As for ongoing work, we are working at parameter optimization on the Active Learning framework (particularly, we are interested in the relations between the size of the unlabeled dataset, the frequency of the re-training, and the confidence score used by the selection procedure). We also plan to extend the current system by adding the detection of the argument structure of events.

Acknowledgments

We thank Valentino Frasnelli for his contribution, which consisted of manually annotating the data. This work has been partially funded by the Euclip project, in collaboration with Euregio.[15]

References

Valentina Bartalesi Lenzi, Giovanni Moretti, and Rachele Sprugnoli. 2012. CAT: the CELCT Annotation Tool. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12)*, pages 333–338, Istanbul, Turkey, May. European Language Resources Association (ELRA).

Steven Bethard. 2013. Cleartk-timeml: A minimalist approach to tempeval 2013. In *Proceedings of the Seventh International Workshop on Semantic Evaluation*, SemEval '13, Atlanta, Georgia, USA.

Claire Bonial, Olga Babko-Malaya, Jinho D. Choi, Jena Hwang, and Martha Palmer. 2010. Propbank annotation guidelines, version 3.0. Technical report, Center for Computational Language and Education Research, Institute of Cognitive Science, University of Colorado at Boulder. `http://clear.colorado.edu/compsem/documents/propbank_guidelines.pdf`.

Romain Cabasson and Ajay Divakaran. 2003. Automatic extraction of soccer video highlights using a combination of motion and audio features. In *Storage and Retrieval for Media Databases 2003, Santa Clara, CA, USA, January 22, 2003*, pages 272–276.

Tommaso Caselli, Valentina Bartalesi Lenzi, Rachele Sprugnoli, Emanuele Pianta, and Irina Prodanof. 2011a. Annotating Events, Temporal Expressions and Relations in Italian: the It-TimeML Experience for the Ita-TimeBank. In *Linguistic Annotation Workshop*, pages 143–151.

Tommaso Caselli, Hector Llorens, Borja Navarro-Colorado, and Estela Saquete. 2011b. Data-driven approach using semantics for recognizing and classifying timeml events in italian. In *Recent Advances in Natural Language Processing, RANLP 2011, 12-14 September, 2011, Hissar, Bulgaria*, pages 533–538.

Tommaso Caselli, Rachele Sprugnoli, Manuela Speranza, and Monica Monachini. 2014. EVENTI

[14] Currently the annotated data are not be distributed due to copyright issues.

[15] `http://www.euregio.it`

EValuation of Events and Temporal INformation at Evalita 2014. In *Proceedings of the Fourth International Workshop EVALITA 2014*.

Lee Raymond Dice. 1945. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, July.

Karën Fort and Vincent Claveau. 2012. Annotating football matches: Influence of the source medium on manual annotation. In *Proceedings of LREC 2012*, Istanbul, Turkey, may. European Language Resources Association (ELRA).

Christian Girardi, Luisa Bentivogli, Mohammad Amin Farajian, and Marcello Federico. 2014. Mt-equal: a toolkit for human assessment of machine translation output. In *COLING 2014, 25th International Conference on Computational Linguistics, Proceedings of the Conference System Demonstrations, August 23-29, 2014, Dublin, Ireland*, pages 120–123.

Yina Han, Guizhong Liu, and Gérard Chollet. 2008. Goal event detection in broadcast soccer videos by combining heuristic rules with unsupervised fuzzy c-means algorithm. In *Proceedings of ICARCV 2008, Hanoi, Vietnam, 17-20 December 2008, Proceedings*, pages 888–891.

Hyuckchul Jung and Amanda Stent. 2013. Att1: Temporal annotation using big windows and rich syntactic and semantic features. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of SemEval 2013*, pages 20–24, Atlanta, Georgia, USA, June. Association for Computational Linguistics.

Taku Kudo and Yuji Matsumoto. 2003. Fast Methods for Kernel-based Text Analysis. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, ACL '03, pages 24–31, Stroudsburg, PA, USA.

Bernardo Magnini, Anne-Lyse Minard, Mohammed R. H. Qwaider, and Manuela Speranza. 2016. TEXTPRO-AL: An Active Learning Platform for Flexible and Efficient Production of Training Data for NLP Tasks. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*.

Jonathan Malmaud, Jonathan Huang, Vivek Rathod, Nick Johnston, Andrew Rabinovich, and Kevin Murphy. 2015. What's cookin'? interpreting cooking videos using text, speech and vision. *CoRR*, abs/1503.01558.

Paramita Mirza and Anne-Lyse Minard. 2014. FBK-HLT-time: a complete Italian Temporal Processing system for EVENTI-EVALITA 2014. In *Proceedings of the Fourth International Workshop EVALITA 2014*.

Michaela Regneri, Marcus Rohrbach, Dominikus Wetzel, Stefan Thater, Bernt Schiele, and Manfred Pinkal. 2013. Grounding action descriptions in videos. *TACL*, 1:25–36.

Eric Ringger, Peter McClanahan, Robbie Haertel, George Busby, Marc Carmen, James Carroll, Kevin Seppi, and Deryle Lonsdale. 2007. Active learning for part-of-speech tagging: Accelerating corpus annotation. In *Proceedings of the Linguistic Annotation Workshop*, LAW '07, pages 101–108, Stroudsburg, PA, USA. Association for Computational Linguistics.

Dan Shen, Jie Zhang, Jian Su, Guodong Zhou, and Chew-Lim Tan. 2004. Multi-criteria-based active learning for named entity recognition. In *Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics*, ACL '04, Stroudsburg, PA, USA. Association for Computational Linguistics.

Young Chol Song, Iftekhar Naim, Abdullah Al Mamun, Kaustubh Kulkarni, Parag Singla, Jiebo Luo, Daniel Gildea, and Henry A. Kautz. 2016. Unsupervised alignment of actions in video with text descriptions. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*, pages 2025–2031.

Manuela Speranza and Anne-Lyse Minard. 2015. Cross-language projection of multilayer semantic annotation in the NewsReader Wikinews Italian Corpus (WItaC). In *Proceedings of the Second Italian Conference on Computational Linguistics CLiC-it 2015*.

Weiyi Sun, Anna Rumshisky, and Özlem Uzuner. 2013. Evaluating temporal relations in clinical text: 2012 i2b2 challenge. *JAMIA*, 20(5):806–813.

Naushad UzZaman, Hector Llorens, James F. Allen, Leon Derczynski, Marc Verhagen, and James Pustejovsky. 2012. Tempeval-3: Evaluating events, time expressions, and temporal relations. *CoRR*, abs/1206.5333.

Guido van Oorschot, Marieke van Erp, and Chris Dijkshoorn. 2012. Automatic extraction of soccer game events from twitter. In *Proceedings of the Workhop on Detection, Representation, and Exploitation of Events in the Semantic Web (DeRiVE 2012)*, volume 902, pages 21–30, Boston, USA, 11.

Marc Verhagen, Roser Saurí, Tommaso Caselli, and James Pustejovsky. 2010. Semeval-2010 task 13: Tempeval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, SemEval '10, pages 57–62, Stroudsburg, PA, USA. Association for Computational Linguistics.

Changsheng Xu, Yifan Zhang, Guangyu Zhu, Yong Rui, Hanqing Lu, and Qingming Huang. 2008. Using webcast text for semantic event detection in broadcast sports video. *IEEE Trans. Multimedia*, 10(7):1342–1355.