International Journal of Computer Science and Engineering Communications,
Vol.3, Issue.4, Page.1229-1234, (2015)
www.scientistlink.com

Extraction of Business Contracts and Temporal Constraints in Business Events

[1] Angelin Jeba Linta.L.T and [2].P.Subhashini
[1] PG Scholar, Department of Computer Science, St. Peter's University, Chennai, India. jebalinta@gmail.com
[2] Assistant Professor, Department of Computer Science, St. Peter's University, Chennai, India. Subhait2k@rediffmail.com

ABSTRACT

The key to successful contract management is the presence of competent people on both the mine owner's and the contractor's teams. When competent people are present on a project, problems will nearly always be resolved; the work will be well planned by the mine owner and well executed by the contractor Projects may survive inadequacies on the mine owner's side but no1: in the contractor's team.A competent contractor can often compensate for deficiencies on the other side. Unfortunately, disaster will often strike if the contractor's team does not know what its doing. The first casualty when competence is lacking is trust and cooperation between the parties. This is because each party will be blaming the other for all the problems that will inevitably be starting to trouble the project. Contracts are legally binding descriptions of business service engagements. In particular, we consider business events as elements of a service engagement. Business events such as purchase, delivery, bill payment, bank interest accrual not only correspond to essential processes but are also inherently temporally constrained. Identifying and understanding the events and their temporal relationships can help a business partner determine what to deliver and what to expect from others as it participates in the service engagement specified by a contract.

Keywords: Cloud Computing, Service engagements, Contract mining, Business events.

1. Introduction

Modern business service engagements are becoming Increasingly more numerous and more complex. To consider service engagements in the broad sense. Thus to include not just traditional examples of service engagements, such as customer relationship management or business process outsourcing, but also other business interactions, such as manufacturing and software licensing. Because service engagements are specified via business contracts, the expansion of the importance of service engagements in modern business is seen in the increasing number of contracts. For example, InfoSys reports1 that 60% to 80% of business transactions are governed by contracts and that an average Fortune 2000 company manages 20,000 to 40,000 active contracts at any given time. The above business trend exposes some new broad challenges in service computing. The first challenge is how, during enactment, a contractual party can understand a contract so as to determine its actions (and design its IT systems) to support its participation in the service engagement. Specifically, would it be able to guide the development of its business processes and monitor its interactions? That is, would the party be able to deliver its part of a service engagement and determine what to expect from its partners in that service engagement? The second challenge is how, during negotiating a service engagement, a party can examine and draft contracts in a manner that incorporates the general practices of the relevant domain. The problem of specifying, adopting, and enacting a service engagement is exacerbated by the fact that contracts are expressed in natural language. Further, often the people who negotiate and those who implement a contract have different skill sets. Accordingly, we are pursuing a research program that seeks to break the problem down into chunks that are amenable to computational analysis. In previous work [1], we tackled a part of the second of the above challenges by mining a repository of contracts to determine the possible business exceptions identified in different domains. To develop an approach that addresses both of the above challenges. This approach is based on the idea of business events—including business-related actions and activities such as purchase, delivery, bill payment, bank interest accrual, licensing, and dispute resolution. Business events indicate the essential processes involved in a service engagement as well as the risks and exceptions to consider. Moreover, the events are naturally temporally constrained, indicating the conditions on their occurrence. The violation of a temporal constraint is often an important factor in contractual breach and the resulting complications. For these reasons, identifying and understanding business events and their temporal relationships in a service engagement can help a business partner in successfully enacting a contract: that is, determining both what to deliver (to others) and what to expect (from others). Understanding business events and their temporal relationships can can also potentially help it decide whether to enter into an engagement in the first place. Note that real-life service engagements are complex interactions with many nuances: so do not claim to have addressed all of the nuances just by identifying events and temporal constraints from contracts, though what to do identify provide a necessary underpinning for more elaborate future analyses.

EXTRACTION METHODS

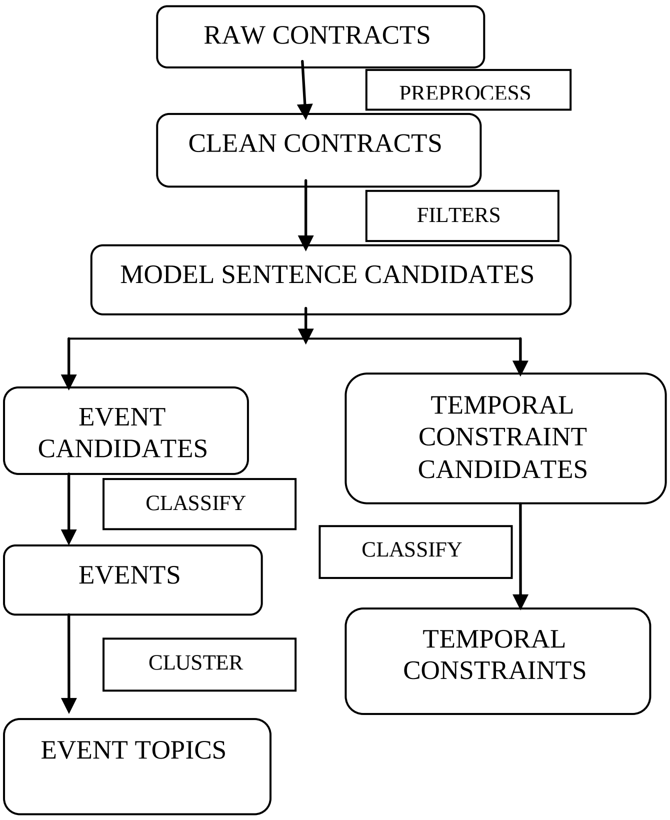
The flow of approach as a hybrid of surface patterns, linguistic parsing, and machine learning techniques. Contract Miner, first, takes raw online contracts as input, removes noise such as HTML tags and segments the contracts into sentence collections. Second, it filters out sentences such as definitions and postal addresses that obviously do not contain business events and temporal constraints. Third, it parses and prunes the remaining sentences to generate candidate events and temporal constraints. Fourth, it applies machine learning on local and contextual features to separately identify true events and temporal constraints from the candidates. Fifth, it applies topic modeling to extract hidden event topics.

We divide our approach into three major tasks:

1) Business events extraction
2) Business event topics discovery
3) Temporal constraints extraction

The three tasks are intimately related.

A contract usually is drafted and signed before the relevant business service engagement occurs; that is, a contract refers to future behaviors. In contrast, the events in other domains usually are descriptive of natural or social phenomena or scientific facts. Event extraction methods rely heavily on patterns. Such methods typically work well in a specific area,for example, natural disaster events [2]. But they suffer from poor portability. Traditional temporal information extraction approaches prove inadequate for extracting temporal constraints from service engagement contracts. Unlike in domains such as news, where the challenge is figuring out temporal orderings [5], in service engagement contracts, time is often explicitly mentioned in prepositional phrases (PPs). However, a challenge is to tease apart the temporal constraints from the other kinds of information that PPs can express, such as space or the intention of an actor. Understanding business events and their temporal relationships can also potentially help it decide whether to enter into an engagement in the first place. Note that real-life service engagements by identifying events and temporal constraints from contracts are complex.



2. TASK 1: BUSINESS EVENT EXTRACTION

A Typical Service Engagement Contract Contains Parts Such As Header, Definition, Body, And Sign Off. At The Core Of A Contract Are The Clauses Specifying Mutual Expectations Expressed As Normative Relationships Such As Commitments, Powers, Authorizations, Prohibitions, And Sanctions Of The Participating parties [3]. normative relationships express business relationships among the parties to a service engagement and these normative relationships are built on top of business events. in english grammar, these normative expressions are often associated with modal verbs such as “shall,” “may,” and “must” [6]. we use modal verbs as signals to signify the occurrence of business events. signal words are widely used in information. After the initial cleanup, Algorithm 1 selects contract sentences that include the signal words as event candidates, parses each candidate sentence to induce the grammar tree, then prunes the grammar tree, and finally builds a feature vector for each candidate using the features extracted from the grammar tree.theUsing the Stanford Parser [7], to parse each event candidate sentence to produce its grammar tree that associates each token with a part-of-speech tag.

Algorithm 1 Business events extraction.
Require: Contract corpus C
1: for all contract c in C do
2: for all sentence s in c that contains a signalword do
3: Parse sentence s to induce grammar tree t
4: Prune tree t to obtain event candidate e
5: Build feature vector f for the event candidate
6: end for
7: end for
8: build classification model with the training data composed of entries in the form of (e; f; Boolean)

3 TASK 2: EVENT TERM CLUSTERING

Business events in service engagements naturally fall into categories such as product delivery, payment, and natural hazards. Automatically discovering the event categories can help better organize events in different service engagement domains. Further, it would help complete the full knowledge discovery cycle by beginning from raw text. Classification and clustering are widely applied to categorize text. Classification methods [12] are supervised, so a training dataset needs to be built manually before hand that predefines the categories. business events found in contracts cut across numerous service engagement domains, with potentially different categories across domains. For example, in licensing contracts, the event categories may be of patent infringement, financial payment, and product licensing. And, in leasing contracts, the event categories may be of property management, rent payment, and eviction. We seek a method that can apply to the services domain where the categories may have not been seen, so classification would not be applicable here. Clustering methods [13] do not need predefined classes and are unsupervised.

Algorithm 3 Event Term Extraction.
Require: Event candidate sentence grammar tree: t
1: **for all** subtree sub in t with a signal word as root **do**
2: **if** the preceding sibling of sub is ps AND ps is NP **then**
3: Return ps as the subject of st
4: **else**
5: **if** the preceding uncle of sub is pu and pu is NP **then**
6: Return pu as the subject of event candidate st
7: **end if**
8: **end if**
9: **end for**

4 TASK 3: TEMPORAL CONSTRAINTS EXTRACTION

Service contracts involve temporal information of various forms. The temporal expression format also varies. Some temporal information is expressed explicitly as dates, for example, "Feb. 3th, 2010" and "10-01-1949 In service engagements, the most relevant temporal information pertains to the constraints that the participants need to observe. For example, a business workflow usually follows a temporal order, and the successful fulfillment of a service engagement greatly depends on the timely completion of those business processes. Such temporal relations among the business events are usually expressed explicitly for the purpose of clarity and emphasis. Temporal constraints in contracts are mostly expressed in prepositional phrases (PP). A prepositional phrase comprises a preposition and noun phrases or clauses. Prepositional phrases function as adverbs in a sentence, and express "where," "how," and "when." Some prepositions indicate temporal boundaries for the completion of a task. For example, "before," "after," "within," "during," "upon," "at," "until," and "between" generally convey the temporal constraints on business events. In our approach, as illustrated in Algorithm , apply similar early steps as in event extraction: clean up the contract text, filter with signal words,and parse the sentences using linguistic tools. To extract the prepositional phrases labeled as "PP" by the Stanford Parser [7]. Because a PP may express a wide range of meanings such as "when," "where," "how," and "why," to treat prepositional phrases as temporal constraint candidates, and employ a classification model to decide if each candidate is a true temporal constraint. Prepositional phrases serve multiple functions in a sentence. For example, prepositional phrases below followed by "at" may indicate "when", "where," or "how" and only the first expresses a temporal constraint.

Algorithm 4 Temporal constraints extraction.
Require: Contract corpus C
1: for all contract ci in C do
2: for all sentence s in ci that contains signal word do
3: Parse sentence s to induce grammar tree
4: Extract the PPs from the grammar tree as temporal constraint candidates
5: Build a feature vector for each temporal constraint candidate
6: end for
7: end for
8: Build a classification model with the training data composed of entries in the form of (PP; Boolean)

To formulate the problem as a text classification task: given a prepositional phrase p, to assign either class label t (temporal constraint) or n (not a temporal constraint) to p. The above problem faces unusual challenges. Traditional text classification tasks generally consider passages from news articles and technical that are long enough to build a useful feature vector. The task is classifying short phrases not exceeding twenty words in most cases. The temporal property of prepositional phrases has been studied in extracting temporal information [15]. However, the ambiguity of prepositional phrases has not been explored. To disambiguate a whether prepositional phrase signifies a temporal or another kind of property. In this task, to apply well-known classification techniques—KNN, Na¨ıve Bayes, and Logistic Regression—to classify the PPs into two classes: temporal and not temporal. In summary, the temporal constraint extraction task is decomposed into two stages: finding PPs and classifying PPs. Linguistic parsing using the Stanford Parser produces PPs and the classification methods detect the temporal PPs. Since the temporal extraction approach is supervised classification, so manually annotated 1,000 prepositional phrases from manufacturing contracts from the One contract repository—the same one can be used above for business events. The annotated prepositional phrases serve as the ground truth. Examples of the positive training set are there. To adopt the bag-of-words model for the features of PPs. For each classification approach, to perform a ten-fold cross validation. To compare the temporal constraints extracted by the system with the ground truth to compute the true and false positives and negatives. With such data, to calculate the precision, recall, and F-measure averaged over ten folds. Using Lingpipe,12 we build a classification model on the training set and evaluate its performance. The detail of each classification method's output below.

As a probabilistic text classification approach, Na¨ıveBayes assumes that the word mutually independent [10]. The experiment involves three settings: no preprocessing, removing stop words only, and removing stop words and stem tokens. The first setting produces Assuming a parametric form for the distribution (Y jX), Logistic Regression learns a mapping from an input vector to a continuous output [10]. Using Logistic Regression, to obtain four sets of results with different selections of features . there also obtain the coefficients associated with each stemmed token. The two ends of the spectrum of the token's association with Temporal properties. As expected, tokens such as "time," "date," "duration", and "day" fall near the highly temporal end, whereas tokens such as behalf,""purpose," "expense," and "exhibition" fall near the non temporal end. In temporal expressions, prepositions are often used together with tokens from the temporal end such as "by year xxx" and "before the month of xxx." However, in nontemporal expressions, prepositions are often used in the expression such as "on behalf of," "for the purpose of," and "at the expense of" to convey nontemporal properties.

SERVICE ENGAGEMENT

Recognizing that service engagements pervade the modern economy, Purvis and Long [27] take an interactionist rather than an objectivist perspective as the underlying principle for modeling real-world businesses. They place multi agent concepts such as norms and institutions at the center of service modeling. Purvis and Long's ideas are naturally cohesive with our approach because business events are the fundamental elements of normative relationships. Therefore, extracting events helps ground the relationships that characterize service engagements. work accords well with conceptual models for service-oriented applications in open environment, In these settings, contracts provide a natural basis for capturing how a service engagement is constructed and enacted. Chopra et al. [30] present an approach for modeling service engagements via commitment protocols to improve the flexibility and expressiveness of engagements. The approach can help elicit the business events and constraints that ground such protocols. Kohlborn et al. [31] study 30 extant service identification approaches and propose a consolidated approach to identify and analyze business services. However, in this work, the process of abstracting and identifying service engagements is manual. Therefore, significant human effort is needed to build the abstract representations of a service engagement. The supervised approach for extracting business events and temporal constraints facilitates service engagement analysis and provides the necessary foundations for automated service engagement identification, and addresses challenges posed in a open contractual environment. Service components analysis facilitates service requirements analysis in business domains. Vitharana et al. [32] propose the knowledge-based component repository (KBCR) to aid service requirement analysis.Similar to their approach, Contract Miner studies a repository of contracts describing service engagements.

In contrast to KBCR, which focuses on formally represented services, Contract Miner studies a contracts repository represented in unstructured text. Further, Contract Miner discover topics of different contract domains in an unsupervised fashion, thereby potentially facilitating the creation of a repository such as KBCR Evaluation demonstrates the effectiveness of machine learning methods for mining business events and temporal constraints. Supervised information extraction from service contracts faces unusual challenges. First, a contract is a legal artifact, and often exhibits more complicated nested structure and longer sentences than ordinary English text. Section and clause headings often cause the sentence boundary detector to break. The length of the sentences challenges the Stanford Parser to output the grammar tree. Second, an event is a subtle semantic unit that challenges automatic extraction. We define events as activities that capture essential business processes. Whereas other event extraction settings involve sentence selection, the events occur at the subsentence level. Pruning helps reduce redundancy in a long legal sentence to capture the most important phrase that expresses an event. The extra processing enhances clarity but may lose information in some cases. Third, building a gold standard dataset is time consuming. Due to the lack of benchmark datasets relating to contracts, we built our own training corpus for event and temporal classification. Evaluation of the event topics is time consuming because there is no gold standard data available. Business events such as payment and service delivery that bear implicit time requirements may lack temporal constraints. The resulting service engagement may fail. For example, disputes could occur when contracting parties default or fail to deliver services in a timely manner. The tool, Contract Miner, captures the essential elements of a contract and thus provides a basis for future work on commitment-based contract analysis[3]. We now define business events and temporal constraints in the setting of text mining contracts for service engagements.

.

KNN

The K-nearest neighbor (KNN) approach labels an instance with the class that is the majority of all its neighbors [10]. Two important factors in KNN are the number of neighbors, k, and the distance function. We adopt the commonly used Euclidean distance to measure the proximity of trained instances. With different neighbor thresholds, where $k = 5$ yields the the best results.

5 Task 4:ANNOTATOR

The text classification tasks to consider are not time critical. Applications such as annotator can process the documents offline and then provide users with highlighted information. To illustrate the use of the trained model, to built a temporal annotator using the model to trained on top of the GATE framework [16]. The quoted text below illustrates the annotation result on a purchasing agreement between Redhook Ale Brewery Incorporated (“Redhook”) and Anheuser-Busch Incorporated. The underlined text is the business event and the italic text is the temporal constraint discovered by the model. In the event that the orders and deliveries of Packaging Materials made by Supplier to Redhook have failed in respects material to Redhook’s Portsmouth operations to comply with the terms of the Supply Agreement and Redhook determines (such determination to be made in good faith and on a commercially reasonable basis) that such failures are likely to continue, Redhook may terminate the purchase and sale obligations of Redhook and ABI under this Agreement upon 30 days written notice to ABI and Supplier.

CONCLUSION

The contracts are specifications of service engagement. Business events and temporal constraints are crucial to enacting a service engagement, therefore extracting them is essential for each party to an engagement to ensure it is being enacted correctly. Business events and constraints can be automatically analyzed to determine whether a potential service engagement is well-formed. Moreover, each party can check if the engagement is acceptable given its individual goals. Importantly, the techniques work on real-life contracts and can thus facilitate service engagements that arise in practice. The classification-based extraction yields F-measures in the high 80% range and vocabulary clustering yields 85% match with the gold standard.

The plan to extend the tool suite. It would be interesting to discover the dependency relationships across business events, e.g., if one event is a prerequisite of another. In the case of manufacturing, a down payment may be a prerequisite for product delivery and instalment payments for continued product supply. Interlocked events form a network of business activities and lay the foundation for effective service engagements as a basis for successful commerce.

“Knowledge Based Business Service Engagement Identification (KBS)”. It would be discover the dependency relationships across business events, e.g., if one event is a prerequisite of another. In the case of manufacturing, a down payment may be a prerequisite for product delivery and instalment payments for continued product supply. Interlocked events form a network of business activities and lay the foundation for effective service engagements as a basis for successful commerce. It is also worth studying the types of dependencies because these are associated with different (normative) business relationships. In particular, this relationship is categorized as normative relationships, such as commitments, permissions, and prohibitions.

REFERENCES

[1] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan,"A framework and graphical development environment for robust NLP tools and applications," in Proceedings of 40th Annual Meeting of the Association for Computational Linguistics, ser.ACL.Philadelphia: Association for Computational Linguistics, 2002,
pp. 168–175.
[2]M. Curtotti and E. McCreath, "Corpus based classification of text in Australian contracts," in Proceedings of the Australasian Language Technology Association Workshop, Melbourne, Australia, 2010, pp. 18–26.
[3] M. De Marneffe, B. MacCartney, and C. Manning, "Generating typed dependency parses from phrase structure parses," in Proceedings of the 5th International Conference on Language Resources and Evaluation. Genoa: European Language Resources Association (ELRA), 2006, pp. 449–454
[4] J. Finkel, T. Grenager, and C. Manning, "Incorporating nonlocal information into information extraction systems by Gibbs sampling," in Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Ann Arbor, Michigan:
[5] X. Gao, M. P. Singh, and P. Mehra, "Mining business contracts for service exceptions," IEEE Transactions on Services Computing, vol. 5, no. 3, pp. 333–344, Jul. 2012.
[6] H. H. Malik, V. S. Bhardwaj, and H. Fiorletta, "Accurate information extraction for quantitative financial events," in Proceedings of the 20th ACM International Conference on Information and Knowledge Management. Glasgow: ACM, 2011, pp.
[7] M. Marcus, M. Marcinkiewicz, and B. Santorini, "Building a large annotated corpus of English: The Penn Treebank," Computational Linguistics, vol. 19, no. 2, pp. 313–330, 1993.
.[8] M. Pasca, "Answering definition questions via temporallyanchored text snippets," in Proceedings of the 3rd International Joint Conference on Natural Language Processing, Hyderabad, January 2008, pp. 411–417.
[9] P. Radha Krishna and K. Karlapalem, "Electronic contracts," IEEE Internet Computing, vol. 12, no. 4, pp. 60–68, 2008.
[10]M. P. Singh, "Norms as a basis for governing sociotechnical systems," ACM Transactions on Intelligent Systems and Technology (TIST), pp. 1–21, 2013, to appear; available at http: //www.csc.ncsu.edu/faculty/mpsingh/
[11] H. Tanev, J. Piskorski, and M. Atkinson, "Real-time news event extraction for global crisis monitoring," in Proceedings of the 13th International Conference on Natural Language and Information Systems: Applications of Natural Language to Information Systems, ser. NLDB. London: Springer-Verlag, 2008, pp. 207–218.
[12] R. Quirk, S. Greenbaum, G. Leech, and J. Svartvik, A Grammar of Contemporary English. Harlow, Essex: Longman, 1984.
[13]K. Vidyasankar, P. Radha Krishna, and K. Karlapalem, "Study of execution centric payment issues in e-contracts," in IEEE International Conference on Services Computing, vol. 2. IEEE, 2008, pp. 135–142.
[14]W. van der Aalst, "Service mining: Using process mining to discover, check, and improve service behavior," IEEE Transactions on Services Computing, p. 1, 2012
[15] I. Witten, E. Frank, and M. Hall, Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, 2011.