Extracting Decision Model Components from Natural Language Text for Automated Business Decision Modelling

Vedavyas Etikala[1,2][0000−0002−5184−3812]

[1] Leuven Institute for Research on Information Systems (LIRIS), KU Leuven
{vedavyas.etikala}@kuleuven.be
[2] Leuven.AI - KU Leuven Institute for AI, B-3000 Leuven, Belgium

Abstract. The decision model in the DMN (Decision Model and Notation) standard is a declarative representation of decision knowledge, which is favored across industry and academia to represent operational decisions. Many current modeling approaches rely on a) a human modeler, which is a costly, time-consuming approach and it struggles to keep up with domain changes, and b) a lot of data logs, to apply automated modeling, which is not feasible for all domains due to unavailability of data. Furthermore, natural language is a standard and convenient way to document decision knowledge in organizations such as rules, policies, and regulations. Despite such vast availability, decision knowledge extraction from the text is relatively new in this domain. This research investigates state-of-the-art NLP techniques, Rule-based approaches, and ML-based approaches in relevant domains. We provide a general framework, Text2DMN, to automatically convert the decision descriptions to the Decision Models. Using this approach, we aim to support decision modelers by reducing the cost and time of the modeling process. This approach also allows improving the quality of models generated, guided by domain expert knowledge as heuristics. We also discuss some of the challenges of this research.

Keywords: Decision Model and Notation · Decision Logic · Decision Tables · Natural Language Processing

1 Introduction

Decision Model and Notation (DMN) is a decision modeling standard designed by the Object Management Group [3]. DMN supports E2E decision management in business organizations with easy communication between all stakeholders of an enterprise [1, 4]. Manually constructing these models, however, takes time and effort. Recent research to come up with automated techniques from various knowledge sources such as event logs and process models [2, 20, 21] to

optimize these modeling efforts. But these approaches are only applicable when such knowledge sources are available. Organizations popularly store their decision knowledge such as requirements and logic in text formats such as documents [5], guidelines or manuals. Natural Language Processing (NLP) is a field of AI that studies various techniques that enable systems to process and understand human language. NLP is a popular research tool for information extraction applications [7, 8]. Therefore, we present novel research exploring how NLP can automatically apply in the DMN domain to generate decision models from existing documentation. The main idea is that from a syntactic and grammatical structure of a sentence, the components of a decision model can be derived (i.e., rules, concepts, dependencies, and constraints) to assemble this information as the decision model in the DMN standard. By pursuing this research goal, we look into related works in DMN and NLP to bring together best practices to extract DMN models from the text. Therefore, we have the following research questions to begin our research with:

- Can we automatically derive DMN decision models from textual descriptions?
- Is it possible to improve the time and cost efficiency of modeling with NLP?
- Are manually modeled DMNs comparable with automatically generated DMNs from text?

Research plan: to answer the above research questions our research is planned in the following steps:

- Perform a literature study to gain good understanding of the techniques and current methods used in the NLP domain as well as technical experience in working with existing tools in information extraction And propose a theoretical framework to solve the identified problem.
- Know the characteristics of decision texts by analyzing the linguistic patterns syntactically and semantically and by mapping text to their corresponding model components. This step helps to understand the various text patterns and challenges in applying NLP.
- Effectively handle these patterns by designing appropriate modeling heuristics inspired from DMN modelling guidelines. These rules will be build into a prototype tool to assess the framework.
- Experiment with a prototype tool to test the proposed solution to extract decision model component from text.
- Further improve and evaluate the tool against human modellers to determine the performance of the framework and quality of the models generated.

This paper is structured as follows. First, the literature study explains the concepts of DMN, NLP and related information extraction in concept modeling context. Next, the research approach explains which problem the program tries to solve. It also explains what tools were used and finally sheds a light on how the program was developed. In the current results section, the program performance is explained, the results are discussed, compared to previous research. In the challenges section, recommendations for future research are given. The last section summarizes the paper with the conclusion.

2 Related Work

In the relavent works of NLP in DMN domain, the automatic extraction of decision dependencies from paragraphs and generation of DRDs is investigated [4] and In [19] decision rules and decision tables are derived automatically from a single sentence, not including decision dependencies.

Prior work in related domains processes and rules can be largely grouped into two types of solutions a rule-based approach and a machine learning approach. *Rule based:* There are many applications of Rule-based NLP for information extraction in the domains such as UML, business rules, and even process models [22]. Rule-based aka pattern-based approaches use domain understanding as heuristics to determine the concepts and relationships, but it is not always the case that domain understanding is available. but if available, the extraction power of rule-based systems is efficient, and the output is explainable too. [11]

ML Based: when there is a large amount of annotated data, supervised machine learning could be used for information extraction. In this method, the ML model learns the patterns from the data. ML-based extraction system will predict the patterns encountered. [13, 15]

Hybrid: some approaches in the literature have opted for the hybrid method for extracting information from text. Here limitations of both rule-based and ML-based NLP are addressed. [11] [12] [14]

3 Challenges in Decision Model Extraction from Text

Based on the study of related research works, decision model extraction from natural language text faces a few issues which can be categorised into challenges that are due to natural language and modelling challenges.

Example: Health risk level

*The health risk level of a patient should be assessed from the obesity level, waist circumference and the sex of the patient. Furthermore, the degree of obesity should be determined from the BMI value and sex of the patient. Patient's height and weight are considered to calculate his BMI value. If the weight of the patient given in kgs and height of patient given in meters, then the BMI value is weight/(height*height).*[a]

[a] taken from https://www.nhlbi.nih.gov/files/docs/guidelines/prctgdc.pdf

Natural language challenges Decision descriptions in policies and guidelines are quite different from the spoken text or informal text. These descriptions come with bullet lists, or logic occurs at various parts of the text with references. However, we can assume that the ambiguity of the decision text is less compared to any social media text. But still, the text with various structures can lead to multiple interpretations. Therefore finding the best preprocessing steps is trivial.

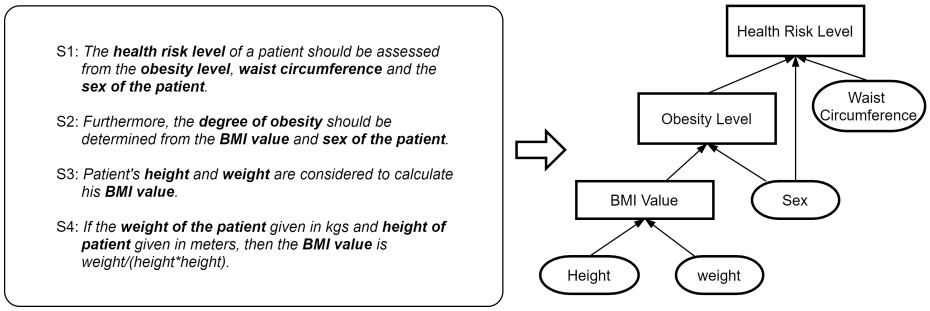

Fig. 1: The DRD of the health risk level description

Decision Modelling Challenges A decision model is an interconnected requirement diagram. Extracted concepts must be semantically grouped and pragmatically placed in DRD to avoid loops, duplicates, and unconnected components. The directions of dependencies are essential for assembling the model. Additionally, there are challenges related to model evaluation on how to measure the quality of models. sample DRD generated for the health risk level can be seen in figure 1.

4 Proposed Solution

The challenges discussed in the previous section point out the complexity of extracting DMN models from decision descriptions. We earlier proposed a solution called Text2dec in[20] as a theoretical and practical level solution to extract decision dependencies. The proposed solution was implemented as a python project and combined with different NLP tools into a pipeline. The Text2dec[20] approach is divided into three distinct phases: text selection, text processing, and model generation. This paper further conceptually extends Text2dec to extracting decision logic as Text2DMN, as shown in the figure below.

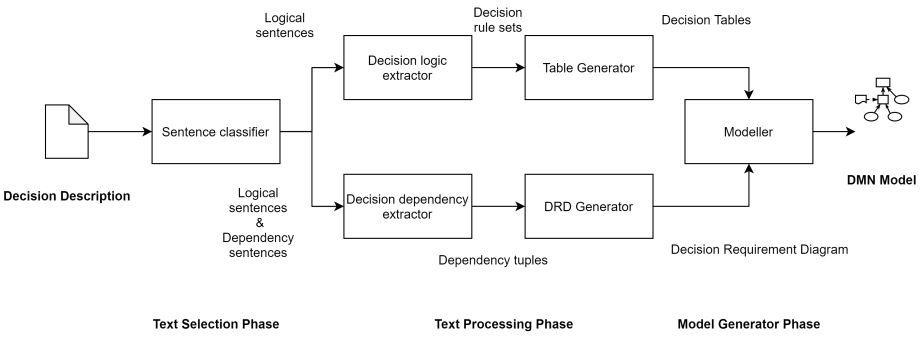

Fig. 2: Text2DMN Framework.

This framework takes a decision text and performs concepts, logic, and dependency identification. We used patterns corresponding to sentences of interest are used in the development of the system. The three key phases, text selection, text processing, and model generation, are being described below to give an idea of how the extraction works.

Prepossessing text and selecting text: The first step of the sentence level analysis phase is the preprocessing of the text. In this step, each sentence is analyzed. Then the execution proceeds with the Spacy tool so that the syntactic and grammatical structure of the sentence is acquired. In other words, POS tagging is performed, and then parse tree-based dependencies are extracted. In this step, noun-based phrases are used for concept identification. Coreferences and anaphoras are resolved using neuralcoref API.

Text processing and information extraction: There are two parallel steps in this phase, as shown in figure 1, and the first step is for Logic extraction. For the logic identification, we used conditional markers such as "if", "then", "whenever," or "unless" are used to detect logical rules. The second step of this stage is dependency identification, here verb phrases such as "determines", "decided from," "depends on" are used to determine decision dependency tuples, where a tuple is (base concept, verb, derived concept). In this step, the previously extracted concepts are matched to base concepts and derived concepts.

Decision Model Generation: The last phase of the transformation approach can also be divided into two parallel steps, decision table builder and DRD builder, as depicted in figure 1. In the first step, tables are built from the identified rules. In the second step, detected dependencies are transformed into DRD and consequently completing the full DMN model.

5 Current Results

A general quantitative approach of information extraction (IS) is used to evaluate the Text2DMN framework's effectiveness. Precision and recall were used as evaluation metrics per each component of the decision model identifiable in the text. The decision logic extractor and decision dependency extractor were evaluated separately. Comprehensive test data set of 20 example texts are tailored and labeled to evaluate the components—data set containing different decision descriptions and manually created models covering various domains. Note that since this methodology is the first time in the DMN domain, there is no available baseline data to test. The overall F_1-score was determined based on the comparison between the desired information and what was automatically extracted by the program. Preliminary results for dependency extractor is shown in table. 1 and decision logic extractor is shown in table. 2.

With an overall F1-score of 0.769, we say that the developed rule-based technique is helpful to extract decision model components, precisely logic, and dependencies, from simple texts that are not ambiguous and straightforward. But real-world texts are seldom simple.

Information	Precision	Recall	F_1-score
dependency tuple	0.955	0.913	0.934
dependency direction	0.818	0.783	0.800
Overall	0.887	0.848	0.867

Table 1: Performance of dependency extractor

Information	Precision	Recall	F_1-score
If variables	0.786	0.723	0.753
Then variables	0.657	0.616	0.636
Else variables	0.605	0.535	0.568
If values	0.718	0.661	0.688
Then values	0.581	0.545	0.562
Else values	0.421	0.372	0.395
If comparison operators	0.748	0.688	0.717
Then comparison operators	0.962	0.902	0.931
Else comparison operators	0.842	0.744	0.790
Overall performance	0.702	0.643	0.671

Table 2: Performance of logic extractor

6 Further Research Plan and Potential Contributions.

The next step in the research plan is to build a tool based on the Text2DMN framework that supports the extraction of decision model components from real-world textual descriptions. Since the literature [19, 20] suggests that complete automation is an ambitious task for real-world texts, considering limitations of current NLP tools, we make semi-automated with little user intervention, at least at the early stages of text selection.

6.1 Contributions:

1. First time application of natural language processing for complete decision modeling in terms of approach and methodology. Natural language processing has been regarded highly in the domains of conceptual modeling because it helps to handle rapid information changes in the organization's policies, rules, or guidelines, and it can also facilitate communication of domain knowledge stored in models through chatbots (similar to [23]). Therefore, automation of decision modeling using NLP will consequently add significant value to decision knowledge management. The tool implemented in python using state-of-the-art open-source linguistic tools such as Spacy, WordNet, and NLTK will improve quality and performance with time, as these tools constantly improve with each breakthrough in NLP.
2. Custom-built co-reference resolution and anaphora resolution components, which require an understanding of the structure of decision text, which in

itself is a massive challenge as there are many ways in natural language to say the same thing. We use the hugging face neuralcoref co-reference resolution system, which can be trained to fit the problem. Which we believe would improve the efficiency component extraction stage.
3. Proposing evaluation metrics for the generated models, and we plan to use semantic and syntactic similarities as a qualitative evaluation metric and to use the complexity metrics of DMN such as a number of decisions, input nodes, and links as a quantitative metric to measure generated to the one modeled from an expert.
4. Conformance checking by aligning decision texts with the models built manually or from data logs.

7 Conclusion

In this work, we explain the current modeling approaches and problems with these approaches. To tackle these problems, we aim to automate the DMN component extraction from textual descriptions. The main idea is that from the syntactic and grammatical structure of a sentence, the decision model components can be derived (i.e., concepts, dependencies, rules, and constraints). The result would be a decision knowledge model represented as DMN - a popular decision modeling standard. We discuss some related work in the area of model extraction and provide a general framework called Text2DMN. We argue that the evaluation of the model extraction process, the results of automated modeling should be close to the original ones that domain experts manually model. Our preliminary results show that the rule-based approach can perform reasonably well on short example descriptions, but they struggle with large real texts. Therefore we plan to incorporate ML techniques and develop hybrid approaches.

Acknowledgements This research is supervised by Prof. Dr. Jan Vanthienen and Dr. Johannes De Smedt.

References

1. Vanthienen, J.: Decisions, advice and explanation: an overview and research agenda. A Research Agenda for Knowledge Management and Analytics (2021)
2. Etikala, V., Vanthienen, J.: Overview of decision model generation methods. In preparation (2021)
3. OMG: Decision model and notation 1.0 (2015), https://www.omg.org/spec/DMN/1.0/
4. Figl, K. et al. "What we know and what we do not know about DMN." Enterp. Model. Inf. Syst. Archit. Int. J. Concept. Model. 13 (2018): 2:1-16.
5. Froelich, J., Ananyan, S.: Decision support via text mining. In: Handbook on Decision Support Systems (2008)
6. Vanthienen, J., Mues, C., Aerts, A.: An illustration of verification and validation in the modelling phase of kbs development. Data & Knowledge Engineering27(3),337–352 (1998)

7. Jurafsky & Martin (2019). Speech and Language Processing 3d edition.
8. Manning cd, schutze h. foundations of statistical natural language processing. themit press; 2000.
9. Bazhenova, E., Weske, M.: Deriving decision models from process models by enhanced decision mining. In: International conference on business process management. pp. 444457. Springer (2016)
10. De Smedt, J., Hasic, F., vanden Broucke, S.K., Vanthienen, J.: Holistic discovery of decision models from process execution data. Knowledge-Based Systems 183, 104866 (2019)
11. Chiticariu, L., Li, Y., Reiss, F. (2013). Rule-Based Information Extraction is Dead! Long Live Rule-Based Information Extraction Systems! EMNLP.
12. M. Dragoni, S. Villata, W. Rizzi, and G. Governatori, "Combining NLP Approaches for Rule Extraction from Legal Documents," in 1st Workshop on MIning and REasoning with Legal texts (MIREL 2016), (Sophia Antipolis, France), Dec. 2016.
13. Bajwa, I.S., Lee, M.G., Bordbar, B. (2011). SBVR Business Rules Generation from Natural Language Specification. AAAI Spring Symposium: AI for Business Agility.
14. Hassanpour, S., O'Connor, M., Das, A.: A framework for the automatic extraction of rules from online text. pp. 266–280 (2011). https://doi.org/10.1007/978-3-642-22546-8 21
15. Danenas, P., Skersys, T., Butleris, R.: Natural language processing-enhanced extraction of sbvr business vocabularies and business rules from uml use case diagrams. Data & Knowledge Engineering 128, 101822 (2020)
16. Friedrich, F., Mendling, J., Puhlmann, F.: Process model generation from natural language text. In: International Conference on Advanced Information Systems Engineering. pp. 482496. Springer (2011)
17. Sanchez-Ferreres, J., Burattin, A., Carmona, J., Montali, M., Padro, L.: Formal reasoning on natural language descriptions of processes. In: International Conference on Business Process Management. pp. 86101. Springer (2019)
18. van der Aa, H., Di Ciccio, C., Leopold, H., Reijers, H.A.: Extracting declarative process models from natural language. In: International Conference on Advanced Information Systems Engineering. pp. 365382. Springer (2019)
19. Arco, L., Napoles, G., Vanhoenshoven, F., Lara, A.L., Casas, G., Vanhoof, K.: Natural language techniques supporting decision modelers. Data Mining and Knowledge Discovery 35(1), 290320 (2021)
20. Etikala, V., Van Veldhoven, Z., Vanthienen, J.: Text2dec: Extracting decision dependencies from natural language text for automated dmn decision modelling. In: International Conference on Business Process Management. pp. 367379. Springer (2020)
21. Robeer, M., Lucassen, G., Werf, J.V., Dalpiaz, F., & Brinkkemper, S. (2016). Automated Extraction of Conceptual Models from User Stories via NLP. 2016 IEEE 24th International Requirements Engineering Conference (RE), 196-205.
22. Honkisz, K., Kluza, K., Wisniewski, P.: A concept for generating business process models from natural language description. In: International Conference on Knowledge Science, Engineering and Management. pp. 91103. Springer (2018)
23. Lopez, A., S'anchez-Ferreres, J., Carmona, J., Padr o, L.: From process model to chatbot. In: International Conference on Advanced Information Systems Engineering. pp. 383–398. Springer (2019)