Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference "Dialogue 2023"

June 14–16, 2023

Multimodal Discourse Trees in Forensic Linguistics

Boris Galitsky
Knowledge Trail Inc.
San Jose, CA, USA
bgalitsky@hotmail.com

Dmitry Ilvovsky
NRU HSE
Moscow, Russia
dilvovsky@hse.ru

Elizaveta Goncharova
NRU HSE, AIRI
Moscow, Russia
egoncharova@hse.ru

Abstract

We extend the concept of a discourse tree (DT) in the discourse representation of text towards data of various forms and natures. The communicative DT to include speech act theory, extended DT to ascend to the level of multiple documents, entity DT to track how discourse covers various entities were defined previously in computational linguistics, we now proceed to the next level of abstraction and formalize discourse of not only text and textual documents but also various kinds of accompanying data. We call such discourse representation Multimodal Discourse Trees (MMDTs). The rational for that is that the same rhetorical relations that hold between text fragments also hold between data values, sets and records, such as Reason, Cause, Enablement, Contrast, Temporal sequence. MMDTs are evaluated with respect to the accuracy of recognition of criminal cases when both text and data records are available. MMDTs are shown to contribute significantly to the recognition accuracy in cases where just keywords and syntactic signals are insufficient for classification and discourse-level information needs to be involved.

Keywords: natural language processing; discourse structure; multimodality
DOI: 10.28995/2075-7182-2023-22-79-87

Построение мультимодальных дискурсивных деревьев для анализа судебных документов

Б.А. Галицкий
Knowledge Trail Inc.
Сан-Хосе, Калифорния, США
bgalitsky@hotmail.com

Д.А. Ильвовский
НИУ ВШЭ
Москва, Россия
dilvovsky@hse.ru

Е.Ф. Гончарова
НИУ ВШЭ, AIRI
Москва, Россия
egoncharova@hse.ru

Аннотация

В работе исследуется концепция построения мультимодального дискурсивного дерева для структурированного представления текста, обогащенного дополнительной информацией из источников различной природы. В более ранних работах были введены понятия коммуникативных дискурсивных деревьев, расширенных с помощью теории речевых актов, а также расширенных дискурсивных деревьев, которые отражают структуру не одного текста, а набора связанных документов; в данной работе мы исследуем возможность расширения дискурсивной структуры за счет включения данных из дополнительных (нетекстовых) модальностей. Мы называем подобное дерево мультимодальным дискурсивным деревом и показываем, что отношения, которые можно установить между частями текста (дискурсивными единицами), также переносятся на данные, дополняющие текст, к которым можно отнести записи из баз данных (например, истории вебпоиска или финансовых операций и т.д.). Мы показываем, что построение мультимодального дискурсивного дерева помогает улучшить качество решения задач поиска на примере анализа судебных документов, которые в большинстве случаев сопровождаются информацией из различных дополнительных источников, по сравнению с поиском по ключевым словам или поиском по стандартному (текстовому) дискурсивному дереву.

Ключевые слова: обработка естественного языка; дискурсивные деревья; мультимодальность

1 Introduction

Discourse analysis plays important role in constructing a logical structure of thoughts expressed in text. Discourse trees are means to formalize textual discourse in a hierarchical manner, specifying rhetorical relations between phrases and sentences. Discourse trees (DTs) are a high-level representation compromise between complete logical representations like logical forms and informal, unstructured representations in the form of original text. Learning DTs has found a number of applications in content generation, summarization, machine translation and question answering (Amgoud et al., 2015; Joty et al., 2015; Joty et al., 2019). The limitation of DT's employment in a general data analysis task is that they are designed to represent the discourse of a text rather than a causal relationship between components of an abstract data item. In this paper, we will address this limitation and propose a solution to generalize DTs towards arbitrary data types with applications to health management and security. In previous works, the authors took DTs to the higher level of abstraction with the goal to form a unified structure for interactive knowledge discovery (Galitsky, 2020). The authors believed that a knowledge exploration should be driven by navigating a discourse tree built for the whole corpus of relevant content. They called such a tree as an extended discourse tree (EDT, (Galitsky, 2019)). It is a combination of discourse trees of individual paragraphs first across paragraphs in a document and then across documents. In this paper, we demonstrate application areas of a discourse representation with a higher level of abstraction and generality. We extend the concept of a discourse tree in the discourse representation of text towards data of various forms and natures. Having defined communicative DT (CDT) to include speech act theory, extended DT to ascend to the level of multiple documents (Ilvovsky et al., 2020) and entity DT to track how discourse covers various entities, we now proceed to the next level and discourse abstraction and formalize discourse of not only text and textual documents but also various kinds of accompanying data. The motivations here are that the same rhetorical relations that hold between text fragments also hold between data values, sets and records, such as *Reason*, *Cause*, *Enablement*, *Contrast*. We call DTs for text and other data forms Multimodal DTs (MMDTs) and apply them in the domains of forensic linguistics (Svartvik and Evans, 1968). Forensic linguistics examines language as it is used in cross-examination, evidence presentation, judge's direction, police cautions, police testimonies in court, summing up to a jury, interview techniques, the questioning process in court, and in other areas such as police interviews (Solan and Tiersma, 2005; Coulthard, 2014).

2 Multimodal Discourse Representation

In this work, we present the notion of Multimodal Discourse Tree (MMDT) that operates on the text level supported with the additional information derived from various sources, where the data is kept in more structural way rather than simple raw texts. Our objective is to recover chains of events from logs of transactions of various sorts including textual descriptions. We show a simple idea of merging various data sources in Figure 1. The trick is how to retain an original structure inherent to each source and merge it with the logical structure of text (an original story). We are motivated by the fact that any coherent text such as patients' complaints or description of the crime scene from the police report is structured so that we can derive and interpret the information.

Discourse analysis aims to reveal the logical structure of some coherent text. This structure shows how discourse units (text spans such as sentences or clauses) are connected and related to each other. In this work, we utilize the Rhetorical Structure Theory (Mann and Thompson, 1988) (RST) as a framework to derive this structure. RST divides a text into elementary discourse units (EDUs). It then forms a tree representation of a discourse called a discourse tree using rhetorical relations such as *Elaboration* and *Explanation* as edges, and EDUs as leaves. EDUs are linked by a rhetorical relation and are also distinguished based on their relative importance in conveying the author's message; nucleus is the central part, whereas satellite is the peripheral part.

In the multimodal setup, we propose to extend the original DT derived from plain text with additional information retrieved from the external sources, such as various logs (financial, call, driving, etc.). The discourse tree extended with this additional information is called MMDT.

Let us consider the motivation that lies behind MMDT construction. A user of some system is not

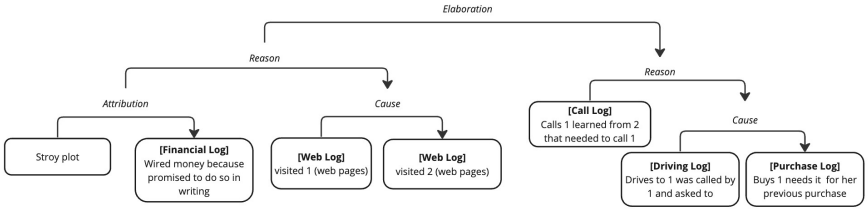

Figure 1: A scheme for a MMDT extended with additional data sources.

always aware of sharing the whole information about his needs, thus, his or her real agenda can be hidden from the system. By the analysis of the external sources of information and merging it with the initial user's request we can reconstruct the whole story and provide more relevant responses for user's request.

We consider various logs such an example of these external sources. It can be bank transactions logs, driving or call logs, web logs, and others. These sources of data can be parsed and represented as the EDUs in the MMDT as shown in Figure 1.

For example, let us imagine a situation, where user wants to make a money transfer, and bank manager is not sure whether this operation is fraud or not. By analysis of the log information kept for this user, the manager can know that this particular user promised to wire money in written form (known from the financial log), then the user visited some web pages from bank system (web logs) in order to make money transaction that he promised to. Thus, this transaction is not fraud and can be performed by the bank. This part of history of the user's log can be hierarchically linked and presented in the form of MMDT presented in Figure 1 (left branch). There we can see that an inner relation within a given data source are combined with interrelations between sources. The same relations hold within a source and between them. The overall logical structure of data is now independent of its nature. A numerical record for banking can be rhetorically connected with a numerical record for calling, which is in turn connected with that of for driving.

Let us now proceed with another example of crime scene description using the MMDT representation. We have a formal description of the crime scene described in the police report. We build the MMDT that can describe this situation within the supported facts represented as the additional modalities. The sources of the extra modalities are shown as the pictograms on the figures.

Let us split the police report about the crime scene into small chunks and build the MMDT supported with extra data for each of them. The MMDT for the first part of the original story is shown in Figure 2. There, the pictograms show the sources for the multimodal data, such as (driving logs, financial logs, etc.).

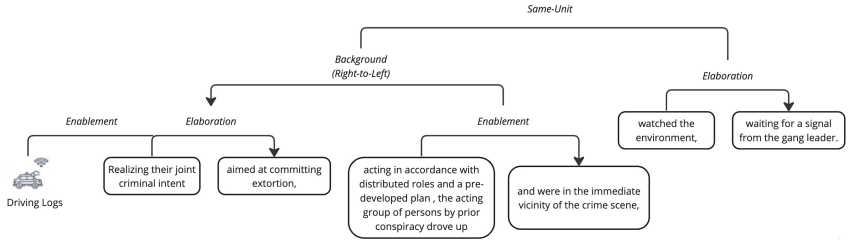

Figure 2: A MMDT for a preparation for a crime. The data record for driving is shown by a pictogram connected with the textual DT by *Enablement*.

Original story. Part 1.

Realizing their joint criminal intent aimed at committing extortion, acting in accordance with distributed roles and a predeveloped plan, the acting group of persons by prior conspiracy drove up and were in the immediate vicinity of the crime scene, watched the environment, waiting for a signal from the gang leader.

We proceed to the start of the extortion crime (Figure 3).

Original story. Part 2.

The victim, unaware of the impending crime against him, at about 5pm, arrived at the house number 162. He was awaited by Jones, acting by a group of persons in a prior conspiracy with Smith and Clark, who, under the pretext of taking out the garbage, left the house and went out to call Smith and Clark that the victim was now located indoors. Thus, Jones gave the signal to start committing the crime.

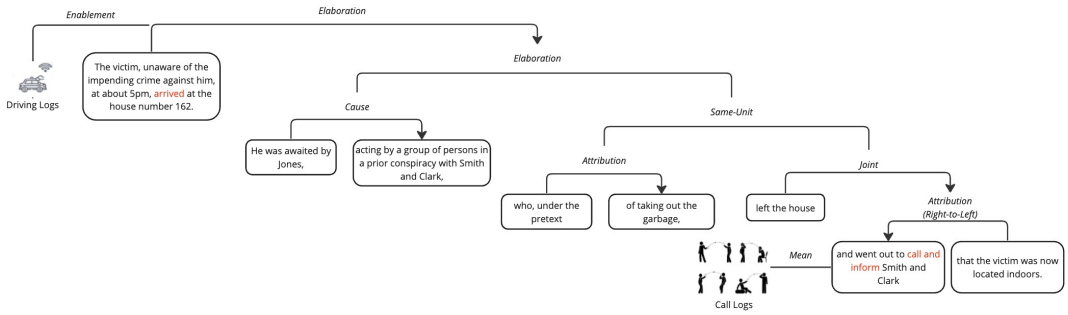

Figure 3: A MMDT for initiation of the crime extended with the multimodal data from the driving logs and call logs.

Rhetorical relations link text EDUs as well as discourse units with information chunks of other modalities: call logs and driving logs, connected with rhetorical relations of Means and Enablement. We now proceed with the crime description (Figures 4-5).

Original story. Part 3.

In the continuation of her joint criminal intent aimed at committing extortion, Jones returned to the house. She did not lock the front door with a key, in order for the gang to enter the house. Smith and Clark acted with her jointly and in agreement. Then they entered the house through the unlocked front door, yelling at the victim.

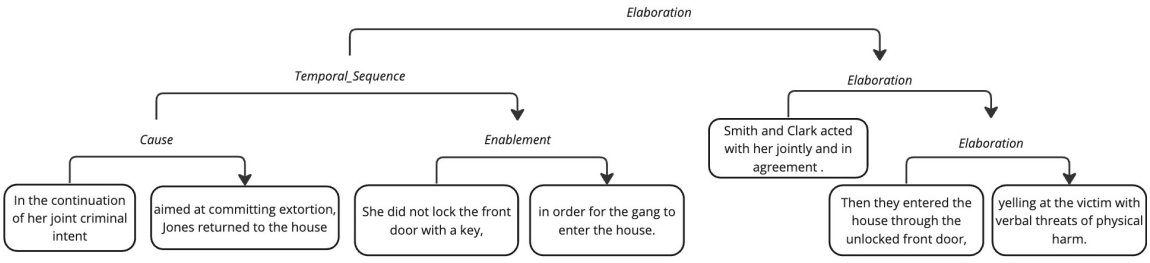

Figure 4: A MMDT for the start of the crime.

Original story. Part 4.

Smith pointed the knife to the victim's elbow and Gereyhanov pointed the handgun to the victim's back, threatening the victim with the murder, unless he does a money wire from his Chase account to Jones's Sberbank account. As the attackers needed more time to have the wire completed, they decided to move the victim to another house to continue money transfer. Jones made a call, making sure certain arrangement were made. Then the attackers pulled the victim out of the house and lead him to the car to drive 35 miles north.

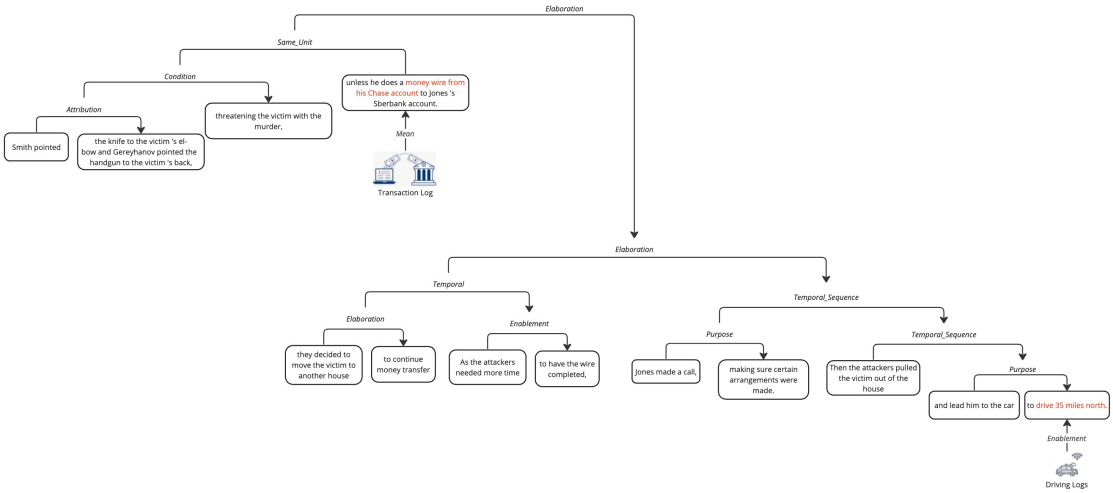

Figure 5: A MMDT for the main stage of the crime.

We now show a tree-like visualization of an arbitrary MMDT which can represent a crime scenario as well as a legal behavior one (Figure 6). This is an example of MMDT where discourse units are data elements such as phone calls, automated number plate recognition (ANPR) records, financial transactions, and texts.

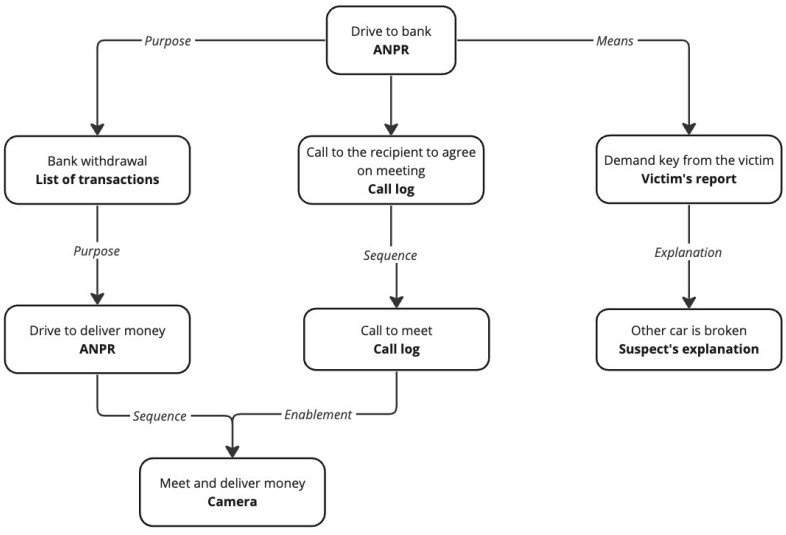

Figure 6: A Multimodal Discourse Tree. The source of additional information are in bold, while the corresponding discourse relations are in italics.

2.1 Multimodal Data Sources and References between them

In this section we analyze the use-case example of crime scene description introduced in the previous section. Via analysis of the MMDTs presented in Figures 2-6, we can observe the whole story in more structural way that can enable us to retrieve answers to the questions by navigating this discourse tree and automatically finding the correct part of the text where this answer can be found. By extending the discourse tree with the additional sources of data we provide extra logical links among the text parts.

Various data sources are not only connected via a discourse tree but are also linked with each other directly. It is hard to form a meaning from a single source, but once we can correspond event parameters from multiple sources and build a whole picture, the constructed event becomes meaningful. For example, if two cars follow each other with a short interval (as determined from the ANPR system), it means that their movement is coordinated.

Once the detective establishes that the gang member drove through a certain point one after another, she would look for confirmation from other sources. If people in one car can see another car, they do not need to call with the purpose of coordination or orientation; if they do make calls then they can have another communication purpose.

It is hard to prove that extortion occurs as the victim could possibly meet the demands of the attackers voluntarily, or the demands did not occur. Sources like calling logs, banking transfers and web logs can indicate whether extortion indeed occurred or not.

Once the extortion process starts, one would expect the victim to be deprived of communication means including phones and the Internet to avoid her calling for help. The call log can easily confirm or reject this expectation. If the frequency of calls of the victim is zero or much lower than that of the gang members, this is a confirmation of an extortion process. The web log can confirm the activity of the victim directly by showing how the victim logged into different accounts and made transfers. Corresponding weblog activities of the attackers who check the receipt of money would be informative as well. Moreover, victims' calls to the banker to perform a transaction that cannot be completed online can also be tracked and matched against the transactions themselves. IP addresses of bank requests can be matched against IP addresses of weblogs. Bank branch locations can be matched with ANPR locations (not used in this particular case).

When a financial transaction happens, a sender and a recipient need to call each other. Also, they likely drove together, or met at some location, as determined by ANRP and call log. Hence for two sources and events in each of them, there are frequently causal links between these events (shown as arrows in Figure 6).

The multimodal DT can be used as the additional source of information that allows us to answer the question based on the extended discourse tree and also ask questions w.r.t. the constructed DT. We now can enumerate multimodal DT-based questions that can be formalized and asked against a MMDT:

For a given individual, find people who visited at one point any location visited by a given person and transferred money to him or back

- Find all pairs of people who drove on different cars following each other within a kilometer of each other
- Find people who call each other and then meet
- Find people who call each other and then transfer money
- Find all people who were once in a location where a given person stayed/visited
- If A calls B who is in a branch in location L to check on account B?

All these questions are relevant for practical applications, where one can easily navigate through the connected textual corpora represented in the form of MMDT.

3 System architecture

We build a conventional CDT from text, convert into MMDT using available structured sources, and then put it into the index for classification and search. The steps of converting a DT into the MMDT are as follows:

1. Once we build an individual CDT for each portion of text, we build a single DT for the whole corpus.
2. As the DT is available, we start preparing accompanying data to incorporate it into DT to form the MMDT. Each data source is converted into a unified, canonical form with normalized named entities: time, date, location, person name, phone number, account number (if available). A scheme for multimodal data transformation is shown in Figure 7.
3. Iterate through each EDU of DT, identifying candidate phrases that can potentially be associated with accompanying data. Extract name entities with their types. Form a list of candidate EDUs for linking with data record.
4. For each candidate EDU, attempt to match entity values against those in data records.
5. In data records taken separately from DT, match records with each other and establish causal links, employing R-C reasoning framework.
6. Iterating through all causal links (and other link types), including internal in data records and external (DT - data records) links, confirm or reject each.
7. For confirmed causal links, insert respective edges in DT to obtain MMDT without relation labels.
8. Recognize types of rhetorical relations between DT and data records. Also, recognize rhetorical relations between data records.
9. Determine if the data record as EDU is connected with DT as nucleus or satellite.
10. Convert obtained labelled MMDT into a normalized MMDT.

As the additional multimodal data sources we use specific logs that provide us with the structured textual descriptions of the described event. For example, if a data record is linked to a pair of text EDUs connected with Elaboration, then Cause is inserted to strengthen the nucleus. We show the scheme for the normalization procedure that turns the data record into a regular EDU below in Figure 7.

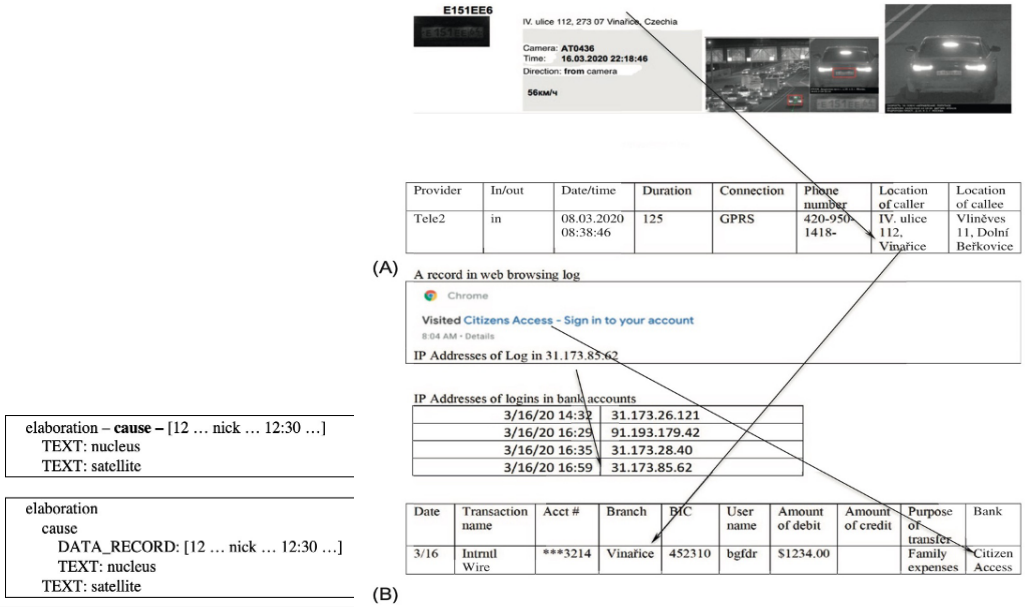

Figure 7: Data records in a call log, web browsing, IP Addresses of financial logins and banking transactions.

4 Evaluation

For the evaluation we considered one practical domain, where the MMDTs can be used for the analysis of texts with the supported structural information (such as log data). To evaluate the contribution of MMDT relative to DT for recognizing scenarios such as criminal cases, we classify them with respect to the felony category such as robbery, theft, abduction and extortion.

Recognition method	Keyword-based	DT	CDT	MMDT–brief description	MMDT - structured
Extortion vs Robbery	58.2	65.0	66.8	70.3	74.3
Robbery vs Theft	61.9	66.8	68.2	71.2	73.7
Extortion vs Theft	66.3	70.4	71.6	73.4	75.4
Abduction vs Extortion	69.1	74.2	76.0	78.5	80.3
Abduction vs Robbery	66.7	72.2	73.6	75.1	81.3
average	64.4	69.7	71.2	73.7	77.0
improvement		5.3	1.5	2.5	3.3

Table 1: Recognition accuracy for felony classes.

It should be noted that in such practical use-cases simple yet effective key-words based search is often applied that allows interpretable and fast search of relevant information. However, using keywords is usually insufficient, as the crime descriptions in court decisions are written in the same or similar keywords for all these crimes: property, car, guns, threats, violence. Discourse level considerations are required, and the more accurate and richer the representation is, the higher the expected recognition accuracy.

We form a dataset of criminal court decisions and attempt to automatically classify the entries in this dataset with respect to felony class. We mine the case site for criminal cases based on statute number and retain the description of corpus delicti to automatically relate it to the statute number. Case descriptions are mined from http://www.sudrf.ru and translated from Russian into English by Bing Translation API.

Due to the lack of complete data on criminal cases other than anonymized textual decision documents, evaluation of the contribution of MMDT is difficult. We build a hybrid dataset of genuine anonymized textual descriptions and attach the same randomized multi-source set of data records. We autogenerate a generic dataset of data records (GDDR) of phone calls, ANPR, weblog, and bank transactions. Having the names, dates, locations and other entities anonymized in both GDDR by the authors and in the public criminal dataset by the court authorities, we insert random entity value to associate actual criminal cases with randomized, hypothetical data records to obtain the complete criminal case data. We recognize one felony category against another, where there is a high similarity in how a crime in a given category is described. Each class there contains 500 documents with 3000 words on average.

Our baseline is keyword-based recognition and regular DTs (columns two and three). In the fourth column we include the phone, drive and money transfer data as a brief description rather than a complete data record and there are no inter–data record rhetorical relations. Finally, in the fifth column, more complete, structured multimodal information is included with built internal data record — data record rhetorical relations.

One can observe that DTs yield more than 5% recognition accuracy compared to keywords, and as we proceed to CDT we gain just 1.5% (Table 1). The next step of enhancement towards the ‘light’ MMDT delivers 2.5% while the ‘complete’ MMDT gives further 3.3%. The recognition rate does not vary significantly across GDDR with the felony class. The contribution of MMDT to an accurate representation of criminal case turns out to be significant and we expect this representation to not depend significantly on the machine learning method.

5 Discussions and Conclusion

In this paper, we took the discourse representation via trees to the next level of abstraction, going beyond textual data and enforcing rhetorical relations between arbitrary components of data items. This allowed us to treat computationally complex scenarios of inter-human interactions described in text and also as numerical and string vectors, once a causal relationship between the latter elements is established.

Complex scenarios of interactions such as GDDR also appear in such domains as security and health management, beyond criminalistics, where textual descriptions need to be merged with numerical values and the logical structure of these data sources must be analyzed together. In this work in progress, we consider the practical application of the MMDTs that can be used to organize complex texts derived from the various source in structural logically-organized format. In future research, we plan to extend the applicability of the introduced framework for more domains, showing where this approach to build the MMDT instead of simple DT or the analysis of plain texts can be preferable.

We computationally evaluated that complex scenarios of inter-human interactions described in plain words and also in data records can be adequately represented via MMDTs in the forensic analysis domain. MMDTs can be employed in other domains involving complex interactions between people or complex correlation between parameters such as customer and patient complaints, prediction of patients' behavior at pandemic times, control of a military unit and prediction of market behavior. These domains are hybrid in the sense that textual information is combined with numerical data and needs to be organized in a uniform way that is invariant with respect to the nature of features used in problem-solving. Statistical learning including deep learning families of approaches encodes all information numerically and certain meanings expressed in text are always lost. Even with a high recognition accuracy of statistical methods, explainability cannot be achieved because numerical representation cannot always be converted back into an interpretable form.

Conversely, MMDTs attempt to encode all available information via a graph with the focus on a high-level logical flow irrespectively of the learning machine which would be applied. Therefore, the MMDT – based approach fully supports explainability and avoids information loss under knowledge representation. MMDT can be naturally combined with additional characteristics of numerical data as well as syntactic and semantic representations.

Acknowledgements

The publication was supported by the grant for research centers in the field of AI provided by the Analytical Center for the Government of the Russian Federation (ACRF) in accordance with the agreement on the provision of subsidies (identifier of the agreement 000000D730321P5Q0002) and the agreement with HSE University No. 70-2021-00139.

References

Leila Amgoud, Philippe Besnard, and Anthony Hunter. 2015. Representing and reasoning about arguments mined from texts and dialogues. *// 13th European Conference on Symbolic and Quantitative Approaches with Uncertainty (ECSQARU 2015)*, volume 9161 of *Lecture Notes in Computer Science book series (LNCS)*, P 60–71, Compiègne, France, July.

Malcolm G. Coulthard. 2014. Whose text is it? on the linguistic investigation of authorship.

Boris Galitsky. 2019. *Discourse-Level Dialogue Management*. Springer International Publishing, Cham.

Boris Galitsky, 2020. *Recognizing Abstract Classes of Text Based on Discourse*, P 379–414. Springer International Publishing, Cham.

Dmitry Ilvovsky, Alexander Kirillovich, and Boris Galitsky. 2020. Controlling chat bot multi-document navigation with the extended discourse trees. *// Proceedings of the 4th International Conference on Computational Linguistics in Bulgaria (CLIB 2020)*, P 63–71, Sofia, Bulgaria, September. Department of Computational Linguistics, IBL – BAS.

Shafiq Joty, Giuseppe Carenini, and Raymond T. Ng. 2015. CODRA: A novel discriminative framework for rhetorical analysis. *Computational Linguistics*, 41(3):385–435, September.

Shafiq Joty, Giuseppe Carenini, Raymond Ng, and Gabriel Murray. 2019. Discourse analysis and its applications. *// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, P 12–17, Florence, Italy, July. Association for Computational Linguistics.

William Mann and Sandra Thompson. 1988. Rethorical structure theory: Toward a functional theory of text organization. *Text*, 8:243–281, 01.

Lawrence M. Solan and Peter M. Tiersma. 2005. *Speaking of Crime. The Language of Criminal Justice*. Chicago Series in Law and Society. University of Chicago Press.

J. Svartvik and T.J. Evans. 1968. *The Evans Statements: A Case for Forensic Linguistics*. Acta Universitatis Gothoburgensis. University of Göteborg.