Business Process Extraction From Documents With AI

Simona Fioretto[1]

[1]Università degli studi di Napoli Federico II, Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'informazione

Abstract

Keeping pace with technological advancements is a challenging task, particularly for Public Administration. The knowledge of working methods in Public Administration actually can speed up the digital transformation process and improve the integration of Information Technology; however, due to the coexistence of several different offices and working methods, discovering this information without the continuous help of experts in the field is far from trivial. The main goal of this paper is to explore and show one of the driving factors for Public Administration digitalization. Since extracting business processes in the public sector is time-consuming, the approach is inverted by using official documents as the source of information, by the consciousness that workflows are dictated by policies, guidelines, regulations, and norms. In order to accomplish this goal, literature proposes many techniques from the use of traditional Natural Language Processing to more advanced Neural Network solutions. This paper aims to give an overview of the problem by exploring the current state of the art, and to show potential future work directions.

Keywords

Business Process, Natural Language Processing, Digital transformation, Public Administration

1. Introduction

Today, public administrations are at the heart of digital transformation. The economic growth of a country depends to a large extent on the services provided by public administrations, which can be the facilitators or the blockers of economic development. With the advent of new technologies and the ability of existing companies, large and small, to be part of a global market, it is also necessary for the PA to adapt to the new economic context in which it finds itself. The digitalisation of the PA is one of the key factors for improving the services offered, both in terms of time, quality, transparency and agility. Known IT solutions have not yet been applied to PAs, which are much slower to follow this digitalisation path. Some authors propose the application of artificial intelligence techniques to the PA, for services such as virtual agents, the automation of some activities, the ability to optimise resources or even the possibility of finding the right task-user match. In order to use new IT systems or apply artificial intelligence techniques, it is necessary to know the working context of public administration offices. Identifying and analysing business processes is a prerequisite for digitalisation. Business Process Management proposes a series of steps that can lead to the improvement of processes, and in particular the first two consist of the discovery and modelling of a process, in order to identify all the

SEBD 2023: 31st Symposium on Advanced Database System, July 02–05, 2023, Galzignano Terme, Padua, Italy

simona.fioretto@unina.it (S. Fioretto)
0009-0006-8700-8188 (S. Fioretto)

activities that must be carried out necessarily and sequentially for the realisation of a process and therefore for the completion of a result. The realization of an IT system to support the improvement of the services offered must take into account the working methods of the public administration offices, which is why the identification phase of the business processes of each public administration office is of fundamental importance. Even within the same sector, different services are offered by a State to its citizens, requiring different ways of working and interacting. From the above, it is clear that business process identification is not an immediate activity in a PA, but must be tailored for each office under consideration. With regard to the digitalization of the PAs of a country, it would be ideal to have a clear idea of the business processes of all PAs, but this is not possible. In fact, identifying these processes is a complex task from many points of view:

- requires the help of experts of the domain;
- it is a process that requires a lot of time for each office analyzed;
- it is subject to and influenced by the perception of the users involved;
- it is intrinsic to errors.

To overcome this problem, it is necessary to perform the extraction of processes in an objective and faster way. The solution could lie in an extraction approach that is no longer based on a bottom-up approach, which foresees the reconstruction of the process starting from its execution and, therefore from the executing user, but is based on the recognition that public administrations, just like companies (although with less rigidity), base their working methods on the procedures imposed by public documents which cannot be changed. In this way, it will be possible to rely on the official information in the procedures, regulations, standards, guidelines, etc., not only to overcome the problems related to obtaining the process on the ground, but also to update the work procedures more quickly when there are new regulatory orders that need to be taken into account. Based on the activities identified in this way, artificial intelligence systems can then be used to enable innovative, effective and efficient management of activities and, with the support of process mining [1], it could be possible to perform a conformance check between the process extracted from documents and the executed process extracted from event logs. The problem, which is the information extraction from documents (unstructured text) is not new and was addressed in the literature by various methodologies. It belongs to natural language processing field. Even if the application of NLP methodologies is growing rapidly, the same cannot be said about the analyzed problem, which is in fact still in an initial exploration phase. The aim of this research is to explore the methods used to perform this task through a literature review, identify possible opportunities among the new methods and perform the application of some methods for a set of documents, with the aim of completely extracting a process starting from a document and performing this task in less time.

2. Background and Problem statement

The topic of digital transformation in public administration was already discussed several times. While on the one hand information technologies and artificial intelligence are becoming increasingly innovative, the same cannot be said for IT applications in public administration.

There is a mismatch between the integration of IT services in the private sector and in the public sector; as highlighted by [2] the challenges presented to public organizations are connected to other challenges such as conservative views, risk aversion to IT and innovation, increased citizen's expectations, lack of digital literacy and leadership in public organizations.

2.1. AI applications in public administration

There are several heterogeneous AI applications that support some of the features required by the PAs, and some of them have been highlighted and compared by [3]. It can be seen that it is a young research field that lacks the description of the related applications and challenges, and in the attempt to systematize it, 5 research categories are found, namely AI in public service dealing with service-oriented applications and new challenges, applications focusing on workflow improvement, AI predictive models, data management, decision and knowledge management, AI-influenced work and social environment, public policy and law related to AI, AI ethics, AI policy in public service. Among these areas, the one that has a more concrete impact as a driving factor is the one related to AI - public services, and in particular the applications of virtual agents and automation of workflows that allow to improve the effectiveness, efficiency and quality of services provided by PAs. [3] also offers papers that implement these applications, and the most important are [4] and [5]. [4] manages the workflow of processing immigration forms on digitized documents using form processing software that performs several key assessment and decision support functions with automatic data extraction and integration of an AI modules, including workflow case assignment, automatic assessment, follow-up action generation, precedent case retrieval, and learning of current practices. [5] which recently deployed AI engines in the social insurance field, starting from standardized workflows, to help in optimization of the Chinese government services, select civil servants for the next processing step in the workflow using a task and civil servant optimization approach.

Despite showing important results, both proposed services use already standardized workflow for applying AI applications; the considered workflows are related only to the case study taken into consideration, and not to the PA in general. It is clear that a crucial part of ensuring the introduction of new information technologies is the identification of working procedures, an activity that can not be carried out individually if the goal is to create applications that are able to support the whole PA sector.

2.2. Business process extraction

The integration of software for public administration, as well as resources optimization, derive first of all from the identification of working procedures: once the tasks and the actors are identified, it will be much easier to implement the right software function, and to apply algorithms that solve optimization problems. The identification of working procedures could be approached using process mining highlighted by [1] as an instrument to discover, monitor and improve real processes (i.e., not assumed processes) by extracting knowledge from event logs readily available in today's (information) systems. Process mining is thus a kind of bottom-up discovery of processes from event logs; however in public administration it could be challenging to obtain event logs in order to identify business processes due to both:

- non-integration of IT tools through all the phases of the process;
- non-availability of public operational data coming from information systems;

It can be said, in a certain way, that the extraction of business processes can be done in two ways: one goes from the execution of the process to its recognition following a bottom-up path, and one starts from rules that define the necessary actions to be performed in order to obtain a result and follows a top-down path. Since the first way is the most uncertain due to human interaction and an alternative objective way such as process mining requires research for the application in the chosen domain due to the previous highlighted challenges, in this context the extraction of business processes using documents is explored. Additionally, the application of business process extraction could also be matched with process mining to check whether the execution of the process is responding to the requirements expressed in official documents.

[6] proposes a new kind of extraction, moving from event logs to policies, calling it Policy-based Process Mining (PBPM), to automatically extract process information from policy documents in text; the approach uses several text mining techniques applied to business policy texts in order to discover process-related policies and extract process components. This field of research is still quite new. An important role and step forward was done by [7] who extracts the process from a text without making assumptions on text structure, furthermore makes a proper anaphora resolution to identify concepts that are referenced using pronouns and articles. Since NLP techniques are constantly and rapidly evolving, it is not an easy task to deal with this topic, because the contributions and approaches are also changing. Most recent approaches use AI methods such as neural networks. [8] use in-context learning giving a prompt to GPT-3 model obtaining good results on participants' information and still unpromising results on follows relationship. [9] extract business process models from texts on emergency plans and instructions, using four layers of neural networks to extract emergency elements and introduce an Adversarial Training Layer for nested entities.

3. Discussion and Conclusion

In this section, based on the previous background analysis, a possible future work is examined and a conclusion is discussed.

3.0.1. Future work

Achieving the solution to the problem of digital transformation for public administration a first attempt can be seen in the identification of working procedures by information extraction from texts. In this topic many approaches are used, which can be distinguished by two variables:

1. *techniques*: most of the recent work uses neural networks, whereas the earlier work used more traditional NLP techniques; this is obviously because the current research can benefit from the power of current neural networks in solving NLP tasks;
2. *word model*: in some studies, the process is derived directly from the text, while in other studies a word model is created and then the process is derived from it;

Even though this challenge has already been faced several times, to the best of my currently knowledge, these are some of the points still open which can affect the future work direction:

- *model notation*: which model notation for process representation best fit the needs of public administration;
- *text structure*: the structure of the text is clearly the most important factor influencing the results;
- *documents relations*: identifying relationships between multiple documents;
- *through text relations*: difficulty in identifying the relationships between tasks and relationships between task and user;
- *gateway identification*: difficulty in identifying the kind of relationships between tasks (which gateway construct they follow? under what condition?);

In order to find possible answers to the open questions, the aim of this research is to break the problem down into several smaller problems. A possible approach to deal with the problem can be structured as follow:

- Given an unstructured text containing both process-related and non-process-related sentences, solve a classification problem to extract only process-related sentences;
- Starting from process-related sentences, solving a new classification problem to identify process constructs;
- Investigate new models and principles for control-flow based models;
- Find a rule to identify the relationships between tasks and between task and user;
- Dealing with the process described in several documents;
- Integration or modification of current processes with new policy requirements.

3.1. Conclusion

In conclusion, this paper highlights the importance of discovering or extracting process models from documents in order to support the digital transformation of public administrations. Discovering or extracting process models from policies is a difficult task, but could lead to huge improvements in supporting the digital transformation of public administrations. The support provided by a simpler and more immediate identification of the activities and roles of a process is linked to many stages of public administration management: simpler design of information and IT systems, analysis of current procedures and comparison and integration with new ones. Furthermore, this technique could be seen in a more general way as a support for process mining, both to check whether the actual execution of the process complies with the policy, and to evaluate the efficiency of the actual execution of the process, i.e. to check which of the processes is the more efficient. As seen, although this subject has been studied and analysed, given the number of disruptive new technologies in the field of Artificial Intelligence and, in particular, Natural Language Processing, there is still much work to be done to improve the phases of process extraction and reconstruction to improve some of the phases of extraction and reconstruction of the process. To conclude it is possible to state that the cohesive application of both process mining and policy based process mining could help in the identification and consequent check of business process execution in public administration domain. Further research in this area can help to improve the integration of Information Technology and support the ongoing digital transformation of Public Administration.

Acknowledgments

This publication article was produced while attending the PhD programme in Information and Communication Technology at the University of Naples, Cycle XXXVIII, with the support of a scholarship financed by the NRRP Ministerial Decree no. 351 of 9th April 2022 for Public Administration
Work supported from the project ”MOD-UPP” - Macroarea 4 - project PON_MDG_1.4.1_17-PON GOV grant.

References

[1] W. Van Der Aalst, A. Adriansyah, A. K. A. De Medeiros, F. Arcieri, T. Baier, T. Blickle, J. C. Bose, P. Van Den Brand, R. Brandtjen, J. Buijs, et al., Process mining manifesto, in: Business Process Management Workshops: BPM 2011 International Workshops, Clermont-Ferrand, France, August 29, 2011, Revised Selected Papers, Part I 9, Springer, 2012, pp. 169–194.

[2] J. M. Goh, A. E. Arenas, It value creation in public sector: how it-enabled capabilities mitigate tradeoffs in public organisations, European Journal of Information Systems 29 (2020) 25–43.

[3] B. W. Wirtz, J. C. Weyerer, C. Geyer, Artificial intelligence and the public sector—applications and challenges, International Journal of Public Administration 42 (2019) 596–615.

[4] A. H. W. Chun, An ai framework for the automatic assessment of e-government forms, AI Magazine 29 (2008) 52–52.

[5] Y. Zheng, H. Yu, L. Cui, C. Miao, C. Leung, Q. Yang, Smarths: An ai platform for improving government service provision, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 32, 2018.

[6] J. Li, H. J. Wang, Z. Zhang, J. L. Zhao, A policy-based process mining framework: mining business policy texts for discovering process models, Information Systems and E-Business Management 8 (2010) 169–188.

[7] F. Friedrich, J. Mendling, F. Puhlmann, Process model generation from natural language text, in: Advanced Information Systems Engineering: 23rd International Conference, CAiSE 2011, London, UK, June 20-24, 2011. Proceedings 23, Springer, 2011, pp. 482–496.

[8] P. Bellan, M. Dragoni, C. Ghidini, Extracting business process entities and relations from text using pre-trained language models and in-context learning, in: Enterprise Design, Operations, and Computing: 26th International Conference, EDOC 2022, Bozen-Bolzano, Italy, October 3–7, 2022, Proceedings, Springer, 2022, pp. 182–199.

[9] G. Zhu, R. Yang, E. Q. Wu, R. Law, Extraction of emergency elements and business process model of urban rail transit plans, IEEE Transactions on Computational Social Systems (2023).