Considering Business Process Complexity Through the Lens of Textual Data

Aleksandra Revina
Faculty of Economics
Brandenburg University of Applied Sciences
Brandenburg an der Havel, Germany
e-mail: revina@th-brandenburg.de

***Abstract*— Organizations are challenged by a growing Business Process complexity as a result of new technologies and continuously generated data flows of various types and volumes. These are likely to result in the organizational performance decrease and loss of control. While textual data is reasonably considered one of the most typical data types in organizations, the mainstream of Business Process complexity research is driven by software complexity approaches, whereby textual data remains discarded. In our work, we explore the potential of textual data generated by the process participants from a linguistic perspective and suggest a textual data-based process complexity concept. We illustrate our proposition with the real-world IT ticket processing scenario and identify IT ticket complexity based on the IT ticket description texts provided by the customers. Our findings evidence sufficient prediction quality and positive influence of linguistic features on the prediction quality.**

Keywords-Business Process Management; Process Complexity; Natural Language; Text Understanding; Linguistics.

I. Introduction

The overused concept of process complexity and strategies to solve this complexity gain new interest due to the rapid penetration of new technologies into Business Process Management (BPM) and resulting dynamics. The focus of BPM as a discipline has traditionally been on the modeling of organizational processes. Hence, the research on complexity is also mainly driven by this perspective [1]. The major complexity approaches in BPM, i.e., complexities of process models, event logs, work- and control flows, have been derived from the software complexity based on the graph-theoretic measures suggested by McCabe in the 1970s [2]. The observation demonstrates a strong focus on the technical artifacts prevalent in the BPM community. In these approaches, the social component expressed through *the textual data massively generated by the business process (BP) participants in the process execution* remains dismissed.

With the maturity of Natural Language Processing (NLP), the attention of BPM research and practice shifts towards textual data. Despite that, the efforts dedicated to analyzing textual data have been mainly directed to support modeling activities [3].

Recently, the studies considering both perspectives of natural language generated by process workers and process logs generated by information systems have started to appear, paving the way to further research. Hence, Fan and Ilk [4] suggest a text analytics framework for automated communication pattern analysis using conversation logs and natural language. In [5], the authors propose enriching Information Technology (IT) ticket resolution logs with the topical phrases extracted from comments to capture the underlying process interactions.

On a positive note of such research directions and in line with “humanistic” enrichments of BPM in general and process mining in particular [6], we suggest inquiring into the potential of natural language generated in the processes and showcase our work from the unconventional viewpoint on process complexity based on natural language.

The remainder of this extended abstract is organized as follows. Section II reflects the approaches used in our work as well as research methodology. Section III provides an illustrative application case. Finally, Section IV discusses the results and gives a short outlook.

II. Overview of Approaches

In the present work, we propose to use textual data generated by BP participants as a valuable source of information allowing to identify the complexity of the process. Linguistics is the discipline concerned with the study of human language, its structure, usage, and history [7]. Traditionally, linguists define textual complexity in terms of the text readability, understandability, or comprehensiveness which is prevailingly based on the average sentence length and the proportion of complex words [8]. Whereas we partially make use of this traditional linguistic understanding, i.e., text understandability for the reader, we aim to extend and adapt it to the BPM context.

In doing so, we consider the three levels of text understanding typically differentiated by the linguists [9] which can potentially indicate the BP complexity: (i) *objective knowledge* (the who, what, where questions), (ii) *subjective knowledge* (opinion, sentiment), and (iii) *meta-knowledge* (what can be learned about the text other than its contents, mainly about its author).

Our research methodology is based on a *three-fold triangulation*. *First*, we follow literature review process to determine and structure our approach as well as define the NLP techniques to be used in the study. *Second*, expert knowledge of both researchers and practitioners is used to adapt these techniques to the BPM and application case context. *Third*, we perform experiments applying our approach to a real-world process to demonstrate its practical value and evaluate the results.

 ISBN: 978-1-61208-877-8

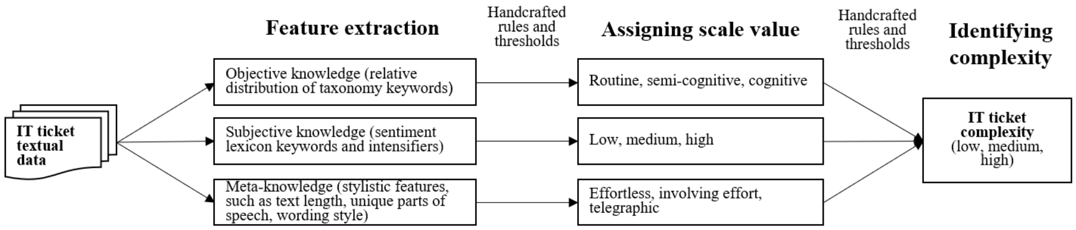

Figure 1. Experimentations

Hence, according to the three above mentioned knowledge types, we use common NLP techniques and adapt them to the BPM context. Thus, we propose domain-specific taxonomy and taxonomy keyword-based pattern matching algorithm to extract *objective knowledge* and estimate *cognitive (mental) efforts* including professional contextual experience of the BP worker necessary to understand the task / process at hand and successfully execute the task / process [10]. The taxonomy contains four basic elements of a BP text: (1) *Resources* (nouns indicating the specificity of BP elements), (2) *Techniques* (verbs of knowledge and information transformation activity affecting *Resources*), (3) *Capacities* (adjectives describing situation specificity of *Techniques*), and (4) *Choices* (adverbs determining the selection of the required set of *Techniques*). These four elements extracted from the BP text are then organized according to the three following levels: (i) routine, i.e., daily, activities, (ii) semi-cognitive, i.e., including some non-routine BP elements, activities, and (iii) cognitive activities demanding much mental effort and involving complex problem-solving [10]. These levels of routine, semi-cognitive, and cognitive also serve as a scale to formalize the objective knowledge extraction and measure the cognitive efforts.

To extract *subjective knowledge* and assess *attention efforts* needed to be paid to particular BP elements and BP as a whole, we develop a domain-specific business sentiment lexicon and lexicon keyword-based pattern matching algorithm extended by semantic and syntactic rules and formalize it on the ordinal scale of low, medium, high [11]. Business sentiment [11] is suggested as an instrument to measure those business-related emotions implied by the BP text author and indicating urgency or importance of the task / process at hand.

To obtain information regarding *meta-knowledge* and *reading efforts* (readability) needed to comprehend the text, we use stylistic patterns [12] expressed with a number of stylistic features, such as text length, unique parts of speech, and wording style [12]. We measure the readability on the ordinal scale of effortless, involving effort, and telegraphic, the latter indicating the texts written in the shortest possible way, as a rule, by professionals to professionals already knowing the specific professional jargon. Meta-knowledge is a naturally distinct type of knowledge. As defined above, objective and subjective knowledge is about extracting information contained in the text. Meta-knowledge indicates the latent information about text quality, i.e., information about the text author [9]. The text quality will likely be determined by the author's professionalism, competence, and stress level. Obviously, a well-written explanation of task / process facilitates timely and successful execution. On the contrary, poorly written explanation will complicate the work.

In the end, as shown on Fig. 1, it is suggested to aggregate the three extracted and formalized on the mentioned scales knowledge types to a BP complexity measured on the scale low, medium, high.

III. ILLUSTRATIVE APPLICATION CASE

To evaluate the suggested concepts, we use an IT ticket processing scenario of an IT Information Library (ITIL) [13] Change Management (CHM) department of a telecommunication company. As BP textual descriptions, we use customer requests for changes in IT infrastructure products or services of the company reaching CHM workers in a free text form, as a rule, per e-mail. For the evaluation purposes, we obtained two data sets with the textual customer requests comprising 28,157 and 4,625 entries correspondingly. Preprocessing and extraction of the knowledge types were conducted using Python 3.4. To obtain application case specific scale values for each of the knowledge types as well as IT ticket complexity (see Fig. 1), the handcrafted rules and thresholds were developed based on the qualitative (using the values manually assigned by the CHM workers) and quantitative (using historical ticket data in case of BP complexity identification) evaluation process implemented in the Microsoft Office Excel 2016 application. This rule-based approach evidenced the BP complexity prediction precision of approximately 65%.

IV. DISCUSSION AND CONCLUSION

In this study, to showcase those meaningful insights inherent in the BP textual data, we proposed that the three types of knowledge, i.e., linguistic levels of text understanding, can potentially indicate textual data-based BP complexity. We tested this proposition using the real-world ITIL IT ticket processing scenario and developed the rule-based approach demonstrating 65% precision [14]. In our further work, we experimented with the semi-supervised machine learning using the suggested linguistic features as IT ticket text representation and available 90 labels with the assigned IT ticket complexity to train the classifiers and predict the complexity [15]. Based on the comparative analysis of various classifiers and text representation approaches, i.e., TF-IDF in our case, we managed to confirm the positive influence of our linguistic features on the prediction quality. In future work, we aim to compare the textual data-based and process mining-based process view, specifically in terms of complexity identification, and derive and analyse discrepancies and commonalities between these two views.

References

[1] J. Cardoso, J. Mendling, G. Neumann, and H. Reijers, “A Discourse on Complexity of Process Models,” Business Process Management Workshops, BPM 2006, 2006, pp. 117–128.

[2] T. J. McCabe, “A Complexity Measure,” IEEE Transactions on Software Engineering, vol. SE-2, no. 4, pp. 308–320, 1976.

[3] H. van der Aa, J. Carmona, H. Leopold, J. Mendling, and L. Padró, “Challenges and Opportunities of Applying Natural Language Processing in Business Process Management,” The 27th International Conference on Computational Linguistics, 2018, pp. 2791–2801.

[4] S. Fan and N. Ilk, “A text analytics framework for automated communication pattern analysis,” Information and Management, vol. 57, no. 4, p. 103219, June 2020.

[5] M. Gupta, P. Agarwal, T. Tater, S. Dechu, and A. Serebrenik, “Analyzing Comments in Ticket Resolution to Capture Underlying Process Interactions,” Business Process Management Workshops, BPM 2020, 2020, pp. 219–231.

[6] W. Tang and M. Matzner, “Creating Humanistic Value with Process Mining for Improving Work Conditions - A Sociotechnical Perspective,” The 41st International Conference on Information Systems, ICIS 2020, 2020, pp. 1–9.

[7] V. A. Fromkin, Linguistics: An Introduction to Linguistic Theory. Wiley, 2002.

[8] T. Loughran and B. Mcdonald, “Measuring readability in financial disclosures,” Journal of Finance, vol. 69, no. 4, pp. 1643–1671, Aug. 2014,

[9] W. Daelemans, “Explanation in Computational Stylometry,” International Conference on Intelligent Text Processing and Computational Linguistics, 2013, pp. 451–462.

[10] N. Rizun, A. Revina, and V. Meister, “Method of Decision-Making Logic Discovery in the Business Process Textual Data,” The 22nd International Conference on Business Information Systems, 2019, vol. 353, pp. 70–84.

[11] N. Rizun and A. Revina, “Business Sentiment Analysis. Concept and Method for Perceived Anticipated Effort Identification,” The 28th International Conference on Information Systems Development, 2019, pp. 1–12.

[12] N. Rizun, A. Revina, and V. Meister, “Discovery of stylistic patterns in business process textual descriptions: IT ticket case,” The 33rd International Business Information Management Association Conference, IBIMA, 2019.

[13] AXELOS, ITIL® Foundation, ITIL 4 Edition. Norwich, UK: The Stationery Office, 2019.

[14] A. Revina and N. Rizun, “Multi-criteria knowledge-based recommender system for decision support in complex business processes,” ComplexRec Workshop Proceedings, Recommender Systems Conference RecSys, 2019, pp. 16–22.

[15] A. Revina, K. Buza, and V. G. Meister, “IT Ticket Classification: The Simpler, the Better,” IEEE Access, vol. 8, pp. 193380–193395, 2020.