Technische Hochschule Brandenburg University of Applied Sciences
Fachbereich Wirtschaft

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

Faculty of Economics and Management, Technical University of Berlin, Berlin, Germany

revina@tu-berlin.de, revina@th-brandenburg.de

The 16th International Multi-Conference on Computing in the Global Information Technology, ICCGI 2021

Special Track on STIS: Socio-technical Information Systems - Design and Application of Socially-aware IT

18-22 July 2021, Nice, France

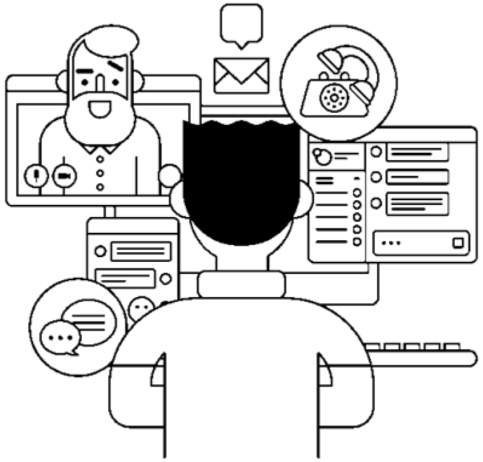
Biographical note

Aleksandra Revina obtained her M.Sc. in Business Administration with the focus on Business Process and Knowledge Management and Engineering from Brandenburg University of Applied Sciences in 2015. Afterwards, she worked as a research scientist and project manager at the Deutsche Telekom Innovation Laboratories in Berlin. In early 2018, she started her industrial Ph.D. at Technical University of Berlin and Brandenburg University of Sciences (cooperative procedure). She is currently working as an academic and research staff at the Brandenburg University of Sciences in various projects, i.a. digitalization. Her research interests include diverse methods and tools for business process analysis and automation from such subject fields as Business Informatics, Business Process Management, Text Analytics, Linguistics, Process Mining with the goal to develop efficient decision support for process workers.

Agenda

1. Introduction and Motivation
2. Overview of Approaches
3. Illustrative Application
4. Approach Extension and Discussion
5. Future Work

Introduction and Motivation

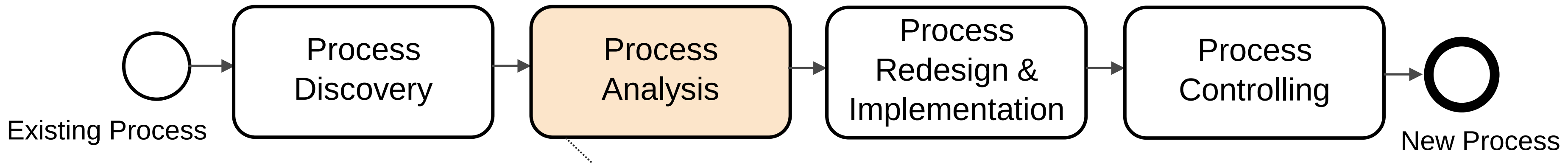


- New technologies, applications, and continuously generated data flows of various types and volumes
- Increase of sources with decision relevant information → difficulty to search
- Demand for (new) methods to analyse these data

Introduction and Motivation

Digitization I New technologies I Increasing Data & Complexity I BPM needs new methods

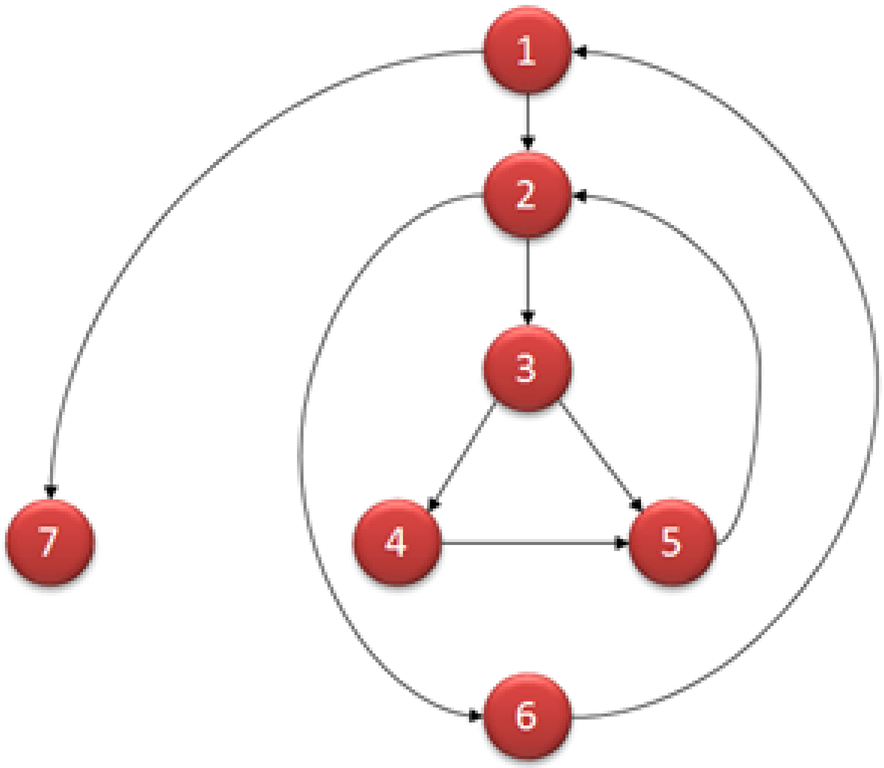
BPM Lifecycle



- Process modelling
- Problem diagnosis
- Root cause analysis
- **Process complexity**

- BPM has traditionally been focused on the modeling of organizational processes.
- BPM research on complexity is also mainly driven by this perspective. 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

Introduction and Motivation



- McCabe cyclomatic complexity
- V(G) = E - N + 2, where E - Number of edges, N - Number of Nodes
- Focus on programming language, software

https://www.guru99.com/cyclomatic-complexity.html

Introduction and Motivation

In the current BPM approaches on complexity, **textual data, which is making up to 80% of enterprise data and massively generated by the process participants in the process execution**, **remains out of scope**.

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**.

Agenda

1. **Introduction and Motivation**
2. **Overview of Approaches**
3. **Illustrative Application**
4. **Approach Extension and Discussion**
5. **Future Work**

Overview of Approaches

Three levels of text understanding in Linguistics

- **Objective knowledge** (answering the who, what, where, when?)
    - Rizun, N., Revina, A., and Meister, V.G., 2019. Method of Decision-Making Logic Discovery in the Business Process Textual Data. In: 22nd International Conference on Business Information Systems, BIS 2019. Sevilla: Springer. pp. 70-84.
- **Subjective knowledge** (who has which opinion about what?)
    - Rizun, N. and Revina, A., 2019. Business Sentiment. Concept and Method for Perceived Anticipated Effort Identification. In: 28th International Conference of Information Systems Development. Toulon
- **Metaknowledge** (what can we extract about the text apart from its contents, mainly about its author?)
    - Rizun, N., Revina, A., and Meister, V.G., 2019. Discovery of Stylistic Patterns in Business Process Textual Descriptions: IT Ticket Case. In: 33rd International Business Information Management Association Conference (IBIMA). Granada: Web of Science.

Taxonomies

Business Sentiment

Stylistic Patterns/ Stylometry

Overview of Approaches

Adapting to BPM context

- **Objective knowledge** (answering the who, what, where, when?)
  - BPM goal: Determine the business process (BP) activity 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
  - BPM adapted approach: **domain-specific taxonomy and taxonomy keyword-based pattern matching algorithm**. 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)
    Organized according to the three following levels, also serving as a scale values:
    (i) *routine*, i.e., daily, activities,
    (ii) *semi-cognitive*, i.e., including some non-routine BP elements, activities,
    (iii) *cognitive activities* demanding much mental effort and involving complex problem-solving.

Overview of Approaches

Domain specific taxonomy example

DECISION-MAKING LOGIC LEVELS		
Routine	***Semi-cognitive***	***Cognitive***
CONCEPTUAL ASPECTS		
RESOURCES		
22 %	8%	2%
user, task, user request, interface, tool, network, firewall	team, leader, project, colleague, production	management, CAB, measure, server farm
time, application, product, name, ID	description, environment, requirement, solution, problem	risk
server, database, file, location, dataset	requestor, case, rule, outage, power-supply	impact, approval
TECHNIQUES		
16%	6%	2%
send, note, deploy, document, decommission	check, assign, increase, create, modify	approve, delegate, define
follow, start, stop, monitor, run	implement, deploy, require, classify, process	propose
cancel, delete, activate, finish, mount	perform, support, plan, verify, migrate	freeze
CAPACITIES		
12%	9%	5%
additional, attached, online, virtual, same	separate, specific, technical, minor, successful	major, high, big, small, strong
new, old, preinstalled, fixed, ready	available, necessary, important, significant, successful	possible, desired, related, different, multiple
actual, full, current, valid, same	temporary, normal, previous, similar, standard	random, randomized, expected
CHOICES		
11%	5%	2%
automatically, manually, internally, instead, there	normally, well, shortly, enough, recently	approximately, properly
current, still, now, often, daily	newly, immediately, later, urgently	soon
consequently, completely, never, simultaneously, accordingly	successfully, however, usually, temporarily, previously	randomly, likely, maybe

Overview of Approaches

Adapting to BPM context

- **Subjective knowledge** (who has which opinion about what?)
  - BPM goal: assess *attention efforts* needed to be paid to particular BP elements and BP as a whole.
  - BPM adapted approach: **domain-specific business sentiment lexicon and lexicon keyword-based pattern matching algorithm** extended by semantic and syntactic rules and formalized on the ordinal scale of **low, medium, high**.

    Business sentiment 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.

Overview of Approaches

Domain specific business sentiment lexicon example

Token, [words, "idioms"]	Value, [-2…2]	*VADER valence*
Tickets based		
"disaster recovery"	0	
disaster	0	-3.1
recovery	0	
affected	0	-0.6
rejected	-2	-2.3
stop	0	-1.2
disable	0	
offline	-1	-0.5
dump	0	-1.6
alarm	0	-1.4
warning	0	-1.4
"set alarms warnings"	0	
risk	-1	-1.1
"poison attack vulnerability"	0	
poison	0	-2.5
attack	0	-2.1
vulnerability	0	-0.9
error	0	-1.7
prevent	0	0.1
drop	0	-1.1
cancel	0	-1
problem	0	-1.7

VALENCE RULES	VADER	Business Sentiment
Scoring rules	[-4; +4]	[-2; +2]
Semantic rules		
Typical business ethics words (e.g., “please”, “dear”, “thank you”)	strongly positive	decreased to 0
Words denoting complex IT problem solving (e.g., “incident”, “emergency”, “downtime”)	strongly negative	slightly increased to -0.5/-1
Typical daily work of IT ticket domain words (e.g., “problem”, “failed”, “adequate”)	positive/ negative	categorized as neutral with 0 valence as they belong to daily work
Typical positive words (e.g., “well”, “successful”, “happy”)	strongly positive	slightly decreased to +0.5
Syntactic rules (intensifiers)		
Capitalizations	additional +0.733/-0.74	additional +/- 0.5 alone standing intensifier - 0.1
“!”, “*”, “=”, “-”, “#”	-	
Negation	regular negation words	“no”, “not”

Overview of Approaches

Adapting to BPM context

- **Metaknowledge** (what can we extract about the text apart from its contents, mainly about its author?)
  - BPM goal: identify *reading efforts (readability)* needed to comprehend the text.
  - BPM adapted approach: **stylistic patterns** expressed with a number of **stylistic features, such as text length, unique parts of speech (BP elements), and wording style.**

    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 indicates the latent information about text quality, information about the text author, i.e., author of a BP task. 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.

Overview of Approaches

In the end, it is suggested to aggregate the three knowledge types, indicating **cognitive, attention, and reading efforts,** to a **textual data-based BP complexity.**

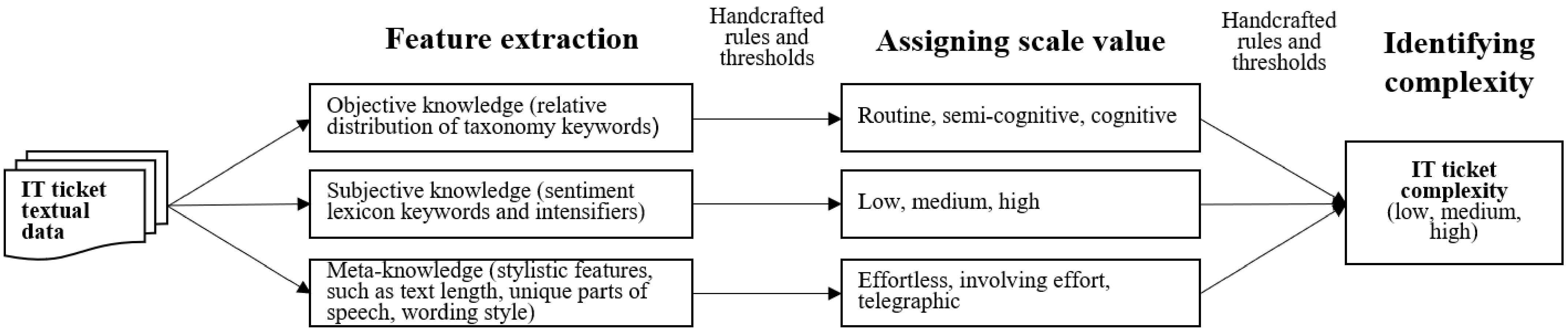
Agenda

1. **Introduction and Motivation**
2. **Overview of Approaches**
3. **Illustrative Application**
4. **Discussion and Approach Extension**
5. **Future Work**

Illustrative application

- **IT ticket processing scenario of an IT Information Library (ITIL) 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 email**.
- Two data sets with the **textual customer requests** comprising **28,157 and 4,625 entries**.
- 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, **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.

Illustrative Application



Agenda

Approach Extension and Discussion

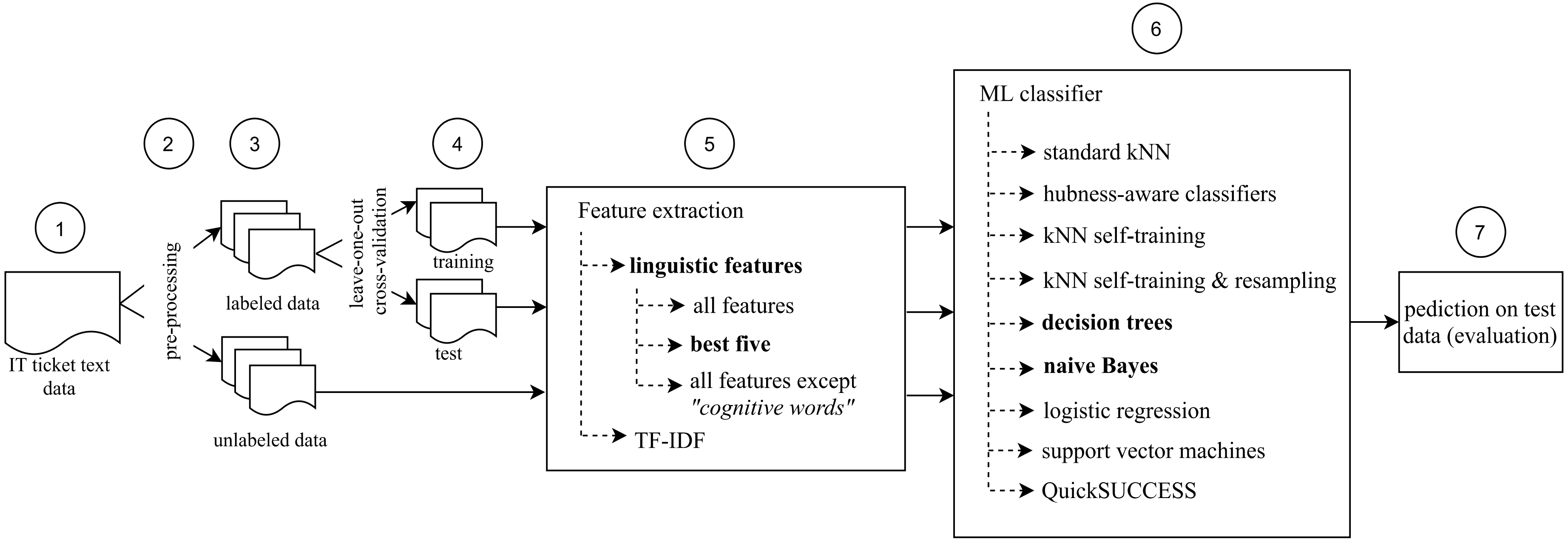
- Rule-based approach demonstrated approximately 65% precision. To address the limitations of the rule-based approach, we tested **ML-based text classification.**
- We compared the performance of **TF-IDF** and **linguistic features-based text representations** designed for ticket complexity prediction using **90 expert labeled tickets** with assigned complexity type.
- We apply various classifiers, including **kNN, its enhanced versions, decision trees, naïve Bayes, logistic regression, support vector machines**, as well as **semi-supervised techniques** to predict the ticket class label of low, medium, or high complexity.
- As our study shows, **linguistic representation not only proves to be highly explainable but also demonstrates a substantial prediction quality increase over TF-IDF**.
- Furthermore, our experiments evidence the importance of feature selection. We indicate that even **simple algorithms can deliver high-quality prediction when using appropriate linguistic features**.

Approach Extension and Discussion

Ticket example: "Refresh service registry on the XYZ-ZZ YYY server. See attachment for details."		
Aspects	**Description**	**Linguistic feature**
objective knowledge aspect	relative occurrence of words according to the taxonomy of routine, semi-cognitive and cognitive terms	routine = 0.8
		semi-cognitive = 0.2
		cognitive = 0
subjective knowledge aspect	relative occurrence of words with positive, neutral, and negative sentiment	negative = 0
		neutral = 1
		positive = 0
meta- knowledge aspect	word count	12
	occurrence of nouns in all words	0.5
	occurrence of unique nouns in all nouns	1
	occurrence of verbs in all words	0.17
	occurrence of unique verbs in all verbs	1
	occurrence of adjectives in all words	0.07
	occurrence of unique adjectives in all adjectives	1
	occurrence of adverbs in all words	0
	occurrence of unique adverbs in all adverbs	0
	wording style [80]	0 (no repeating words)

The best predictive feature

Approach Extension and Discussion



Approach Extension and Discussion

Evaluation Results Using The Five Best Performing Linguistic Features

Algorithm	Accuracy	Average precision	Average recall	F-score
Data1: five best linguistic features				
standard kNN	0.667	0.589	0.585	0.587
kNN self-training	0.667	0.589	0.585	0.587
kNN self-training & resampling	0.667	0.589	0.585	0.587
decision trees	1.000	1.000	1.000	1.000
naïve Bayes	1.000	1.000	1.000	1.000
logistic regression	0.633	0.611	0.580	0.595
SVM	0.967	0.976	0.970	0.973
QuickSUCCESS	0.733	0.814	0.636	0.714
Data2: five best linguistic features				
standard kNN	0.817	0.711	0.670	0.690
kNN self-training	0.817	0.711	0.670	0.690
kNN self-training & resampling	0.700	0.233	0.333	0.275
decision trees	1.000	1.000	1.000	1.000
naïve Bayes	1.000	1.000	1.000	1.000
logistic regression	0.850	0.554	0.606	0.579
SVM	0.983	0.958	0.970	0.962
QuickSUCCESS	0.867	0.791	0.693	0.739

Approach Extension and Discussion

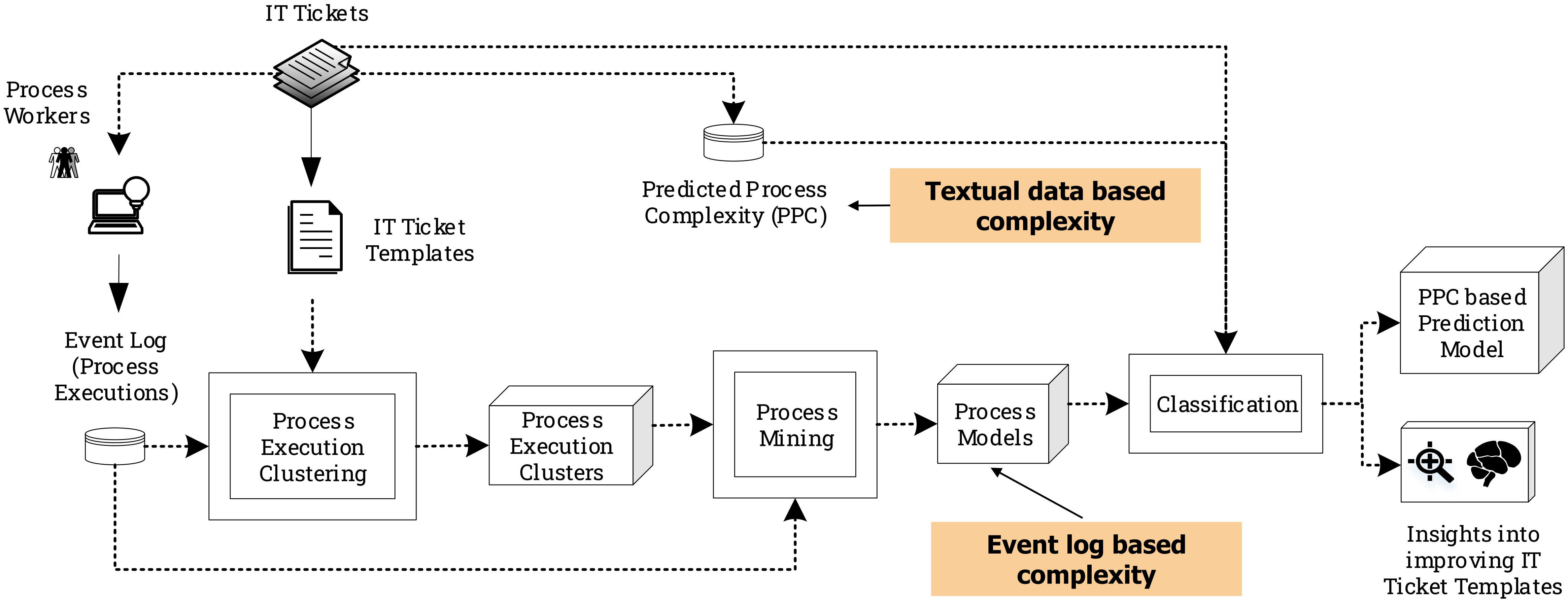
- **Novel textual data-based BP complexity concept** based on **the three linguistic levels of text understanding**.
- **Real world industrial data set is used to test and evaluate the approach**.
- The approach is based on the **common NLP techniques**, which can be relatively easily implemented.
- The **difficulty and limitation** lie in the preparatory work of vocabularies' compilation and establishment of threshold rules.
- However, first, our vocabularies have been compiled for **IT Service Management area**, which is a rather **broad application domain**. Second, the **ML approach based on our linguistic features** showed to **be efficient already with a simple classifier**.
- Our BP complexity concept can be referred as **explainable**. Process workers can **trace back** the suggested level of complexity. This is especially important in the context of erroneous classifications and the **Explainable Artificial Intelligence (XAI)** paradigm.

Agenda

1. **Introduction and Motivation**
2. **Overview of Approaches**
3. **Illustrative Application**
4. **Approach Extension and Discussion**
5. **Future Work**

Future Work

Process Mining: compare event log based and textual data based complexities identified in the IT ticket executions.



Considering Business Process Complexity Through the Lens of Textual Data

Thank you for your attention.

Questions, ideas, cooperation opportunities: Aleksandra Revina,

revina@th-brandenburg.de