Describing Trends and Patterns in Business Process Mining Data with Natural Language

Master thesis - Kickoff presentation

Siddhesh Kandarkar, Garching , 02.03.2020

Chair of Software Engineering for Business Information Systems (sebis)
Faculty of Informatics
Technische Universität München
wwwmatthes.in.tum.de

Partner

Outline

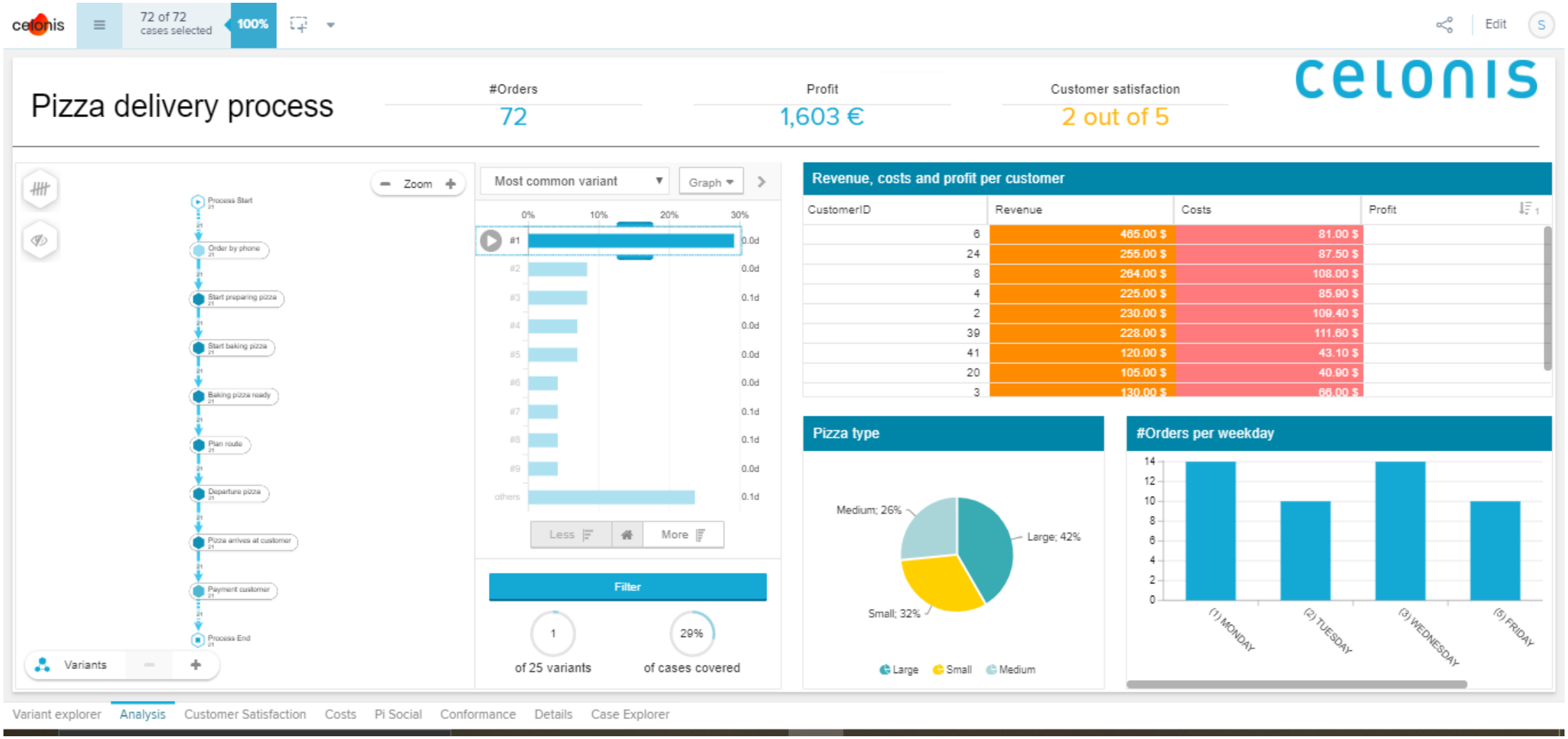
1. Introduction
   - Problem area
   - Celonis business dashboard
2. Motivation
3. Research questions
4. Approach
5. NLG
6. Evaluation techniques
7. Timeline

Introduction

Problem area

- Celonis IBC is a process mining software used to provide business solutions to clients.
- Business dashboard consists of number of visual representations of business data (e.g. Pie chart, histogram, tables etc.)
- Sometimes, it is difficult and time-consuming for non-technical business users to understand and interpret these visual representations.

Celonis business dashboard



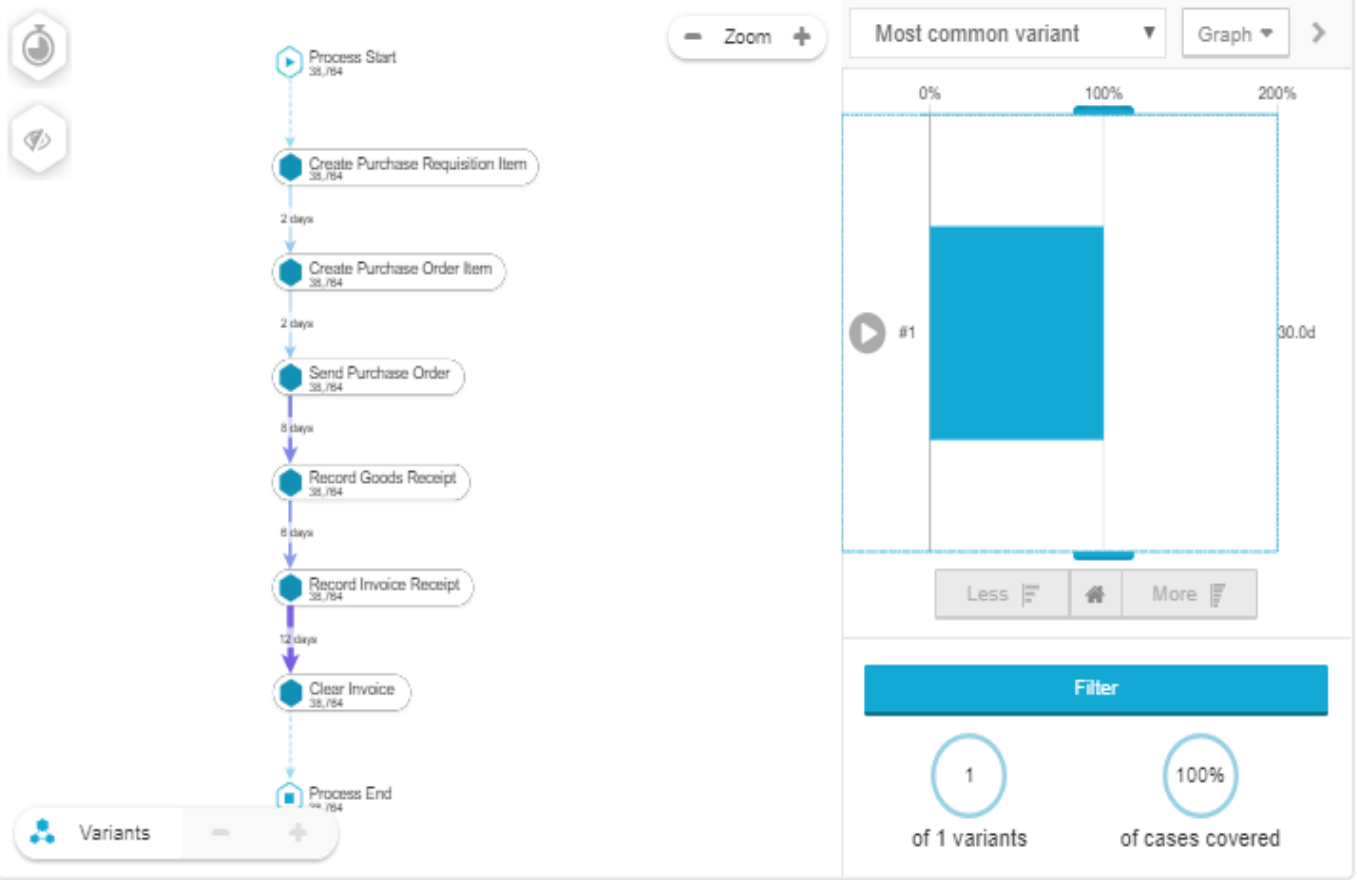
celonis | 38.8k of 1.12M cases selected | 3% | Creation Date 2016-07 | Variants Create Purchase Requisition Item... | Variants Create Purchase Requisition Item... | Variants Create Purchase Requisition Item... | Variants Create Purchase Requisition Item... | Creation Date 2016-07 | Edit

Purchase-to-Pay Overview

# Purchase Order Items	Net Order Value
38.8k	37.3M €

celonis

Process Explorer



Purchase Order Items and Value by Month

Purchase Order Value by C...

Open Action Engine

Purchase Order Items by Vendor

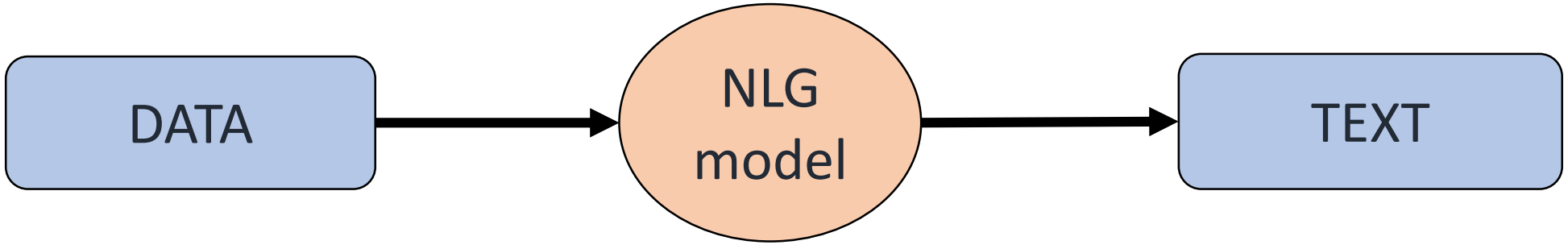
Select drilldo...

Vendor	# Purchas...	Net Value
Unisono AG	4,068	90,413 €
Sapsota Company Limited	3,612	32,085 €
IDES Consumer Products	2,648	35,745 €
SCT Inc.	2,444	195,016 €
C.E.B. BARCELONA	2,260	24,566,539 €
PAQ Deutschland GmbH	2,252	828,682 €
Allfresh Inc.	1,908	105,018 €
SKF Americas	1,908	319,260 €
MOBILE Inc.	1,500	47,114 €

Process | Overview | Automation | Rework | Benchmark | Conformance | Details | Case Explorer | Process Explorer

Motivation

Map business data into an easily understandable and meaningful textual representation



Time series data (timestamp – numerical value)

timestamp	raw_value	percentage	dpo	percentage2
2019-12-21	61	0.61	32	0.642
2019-12-20	70	0.7	29	0.711
2019-12-17	56	0.56	29	0.704
2019-12-16	76	0.76	29	0.709
2019-12-15	68	0.68	29	0.702

- In the year 2019, the average raw value was 36.6
- Raw value increased every year from 2015 to 2018, while dpo remained steady.

Research questions

- Which state-of-the-art NLG techniques / tools are best suited for Data to text generation for a Time series data ?
- What are the various steps involved in developing an NLG prototype ?
- How can we gain meaningful insights from contextual data (involving multiple KPI's) ?
- How to represent the output text to the end user ?
- Which could be the best possible evaluation technique for our prototype ? And what should be the evaluation baseline ?

Approach

1. Survey state-of-the-art NLG techniques for Data-to-Text Generation
2. Design and prototypically implement an approach for Data-to-Text Generation
3. Evaluate on real world data with non-expert business users

NLG (Natural Language Generation)

- NLG is a software process that transforms structured data into natural language.
  - NLP is computers *reading language*
  - NLU is computers *understanding language*
  - NLG is computers *writing language*

- Data to text generation can be divided into 2 parts :
  - Planning
  - Realization

NLG models and tools

- NLG models:
  - Markov chain
  - RNN
  - LSTM
  - Transformer

- NLG tools:
  - Commercial : Quil, Wordsmith, AX Semantics, Arria
  - Open source : SimpleNLG, NaturalOWL

Evaluation techniques

- Task-based evaluation
- Human ratings based
- Metrics based

Timeline

Tasks	February	March	April	May	June	July	August
Problem scope and definition							
Kick off presentation							
Literature review							
Development							
Evaluation and Feedback							
Writing							
Final presentation							

Thank you

© sebis

B.E.

Siddhesh Kandarkar

Technische Universität München
Faculty of Informatics
Chair of Software Engineering for Business Information Systems

Boltzmannstraße 3
85748 Garching bei München

Email: siddhesh.kandarkar@tum.de