Tabular Constraint Learning

Name1 Surname1 and **Name2 Surname2** and **Name3 Surname3** [1]

Abstract. abstract

1 Introduction

SERGEY bullet points for luc to start introduction
Key question:
Can we discover or reconstruct structural relations in flat tabular data? [in a general way that allows declarative specification of constraints to discover]

Motivation:

- File generated from model, model got lost - ¿ reconstruct
- Constraint programming is hard - is Excel hard?
- Avoid manual analysis, provide selection of constraints
- Error checking
- Completion, gain speed and insights (Complicated constraints, also complicated to verify, too much output)

Novelty:

- Unsupervised setting (contrary to flashfill, etc)
- Numeric, different constraints (contrary to single textual function solution in flashfill, etc)
- Data format (2D) – data is no longer in rows like a classic ML or DM settings
- Declarative, general / modular, stacking of constraint problems

SERGEY we need structure here
Approach

- Notation
- Algorithm (select constraints, find assignments, find solutions)

Experimental questions

- How accurate are we? (Accuracy / recall)
- How fast are we and which factors affect the runtime (how)?
- How general is our approach, what limitations are there?

2 Related Work

SERGEY key bullet points for Luc and possibly Samuel and me to make related work section
SERGEY ECAI reference style file ignores their guideline and their guideline ignores what is written in the guidelines! flashfill, flashextract, flashmeta [**? ? ?**]

- their supervised vs our unsupervised approach
- they look for a single "smallest" solution, we enumerate them all
- they are looking for a function, we solve constraint satisfaction problems
- we do not assume classic row based data layout, we work in the tabular setting

sketch [**?**]

- look for a constant that would fill in the gap in a program
- tailored for programming languages
- similar to model checking
- looks for a single solution
- similar to constraint satisfaction and sat, where one is interested in a single assignment that works for any potential input

tabular [**?**]

- language based on the excel tables that specify probabilistic models
- a system for probabilistic inference and similarity mostly in the usage of excel
- probabilistic constraint satisfaction (?) and graphical models
- single solution again

modelseeker [**?**] SERGEY Samuel, Luc, probably you would need elaborate here more in details

- not designed for excel-like data representation (type consistency, groups, etc)
- not designed for excel-like constraints (lookups, conditional ifs, etc)
- does not support user extensions (?)

claudien [**?**] SERGEY Samuel, Luc, you would need to help with this one

[1] KU Leuven, Belgium, email: firstname.lastname@kuleuven