Workshop on Configuration

Vienna, Aug. 29th-30th, 2013

Recommender Systems for Configuration Knowledge Engineering

Alexander Felfernig, Stefan Reiterer, Martin Stettinger,
Florian Reinfrank, Michael Jeran, and Gerald Ninaus

Institute for Software Technology
Graz University of Technology, Austria
{*firstname.lastname*}@*ist.tugraz.at*

Overview

- Comments on Existing KE Practices
- Recommendation Technologies for KE (Scenarios)
- Results of First Empirical Studies
- Ongoing & Future Work

Comments on Existing KE Practices

- Model-based Knowledge Representations
- Graphical Development Environments
- Automated Testing & Debugging
- Our Goal: Recommendation Technologies for KE

Example: Collaborative Filtering

user	c_1	c_2	c_3	c_4	c_5	c_6
1	4	2	3	5	1	6
2	3	2	5	6	1	4
3	1	3	2	4	6	5
4	3	2	4	5	1	6
current	?	2	?	?	1	?

nearst neighbors

current user

Recommendation: users who inspected c5 and c2, also inspected c1.

k-means Clustering of Constraints

$c_i \in C$	c_1	c_2	c_3	c_4	c_5	c_6	c_7
c_1	1.0	-	-	-	-	-	-
c_2	0.33	1.0	-	-	-	-	-
c_3	0.16	0.33	1.0	-	-	-	-
c_4	0.16	0.5	0.16	1.0	-	-	-
c_5	0.1	0.25	0.1	0.37	1.0	-	-
c_6	0.0	0.0	0.0	0.0	0.12	1.0	-
c_7	0.0	0.33	0.33	0.16	0.12	0.16	1.0

iteration	c_1	c_2	c_3	c_4	c_5	c_6	c_7
1	$1(cs)$	1	1	2	$2(cs)$	2	2
2	1	$1(cs)$	1	1	$2(cs)$	2	1

$$sim(c_a, c_b) = \frac{\sum_{v \in V} co\text{–}occurrence(v, c_a, c_b)}{|V|}$$

k groups generated by k-means clustering: each group has a centroid which is the constraint most similar to all others.

Constraint Clustering: Empirical Study

Grouping approach	kba_1: *SOL*	kba_2: *CON*
Similar variables	21.43%	42.86%
Similar operators	30.77%	53.85%
Random	38.46%	76.92%

Table 5: Error rates for completing the tasks *find a solution (SOL)* and *find a conflict (CON)* depending on clustering approach (variable-based, operator-based, or random).

The best results were achieved with variable similarity-based clustering (N = 40 subjects).

Understandability of Constraints

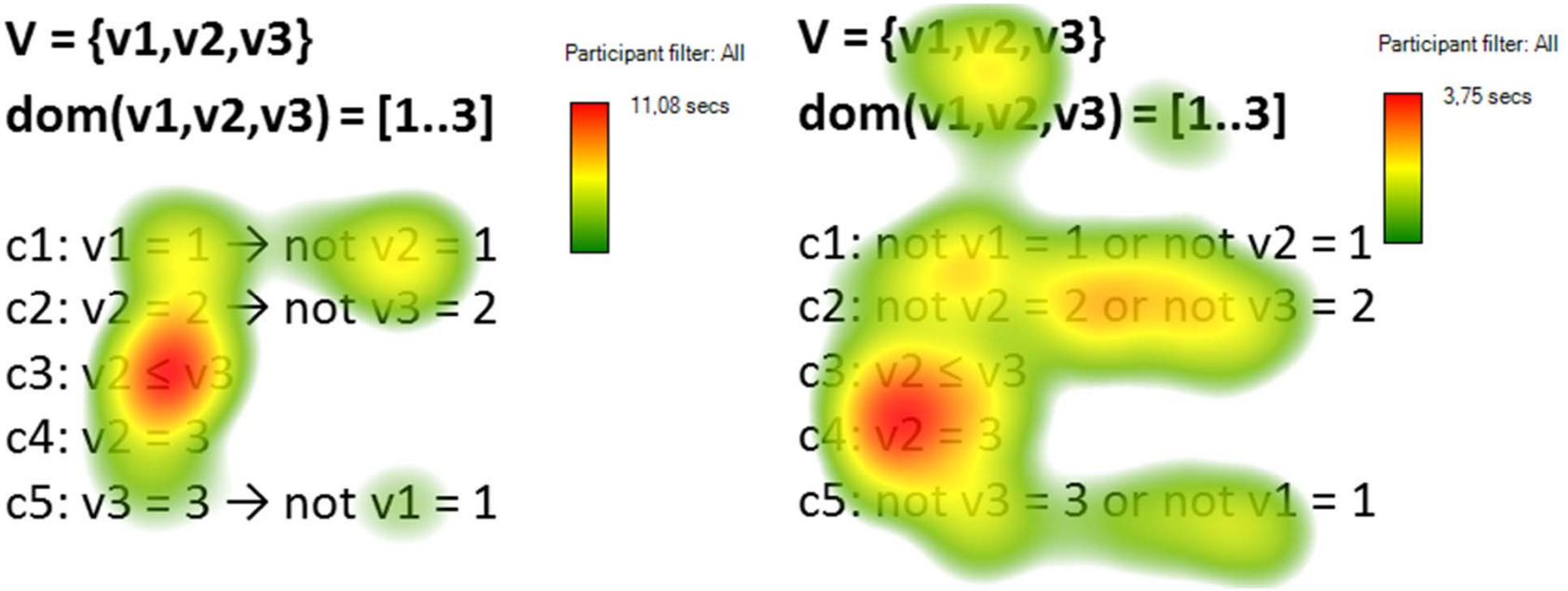
kbb_1: SOL	errors	kbb_2: SOL	errors
$X \rightarrow Y$	21.43%	$X \rightarrow \neg Y$	14.29%
$\neg X \vee Y$	50.0%	$\neg X \vee \neg Y$	34.62%
$\neg Y \rightarrow \neg X$	96.43%	$Y \rightarrow \neg X$	50.0%
$\neg(X \wedge \neg Y)$	73.08%	$\neg(X \wedge Y)$	42.31%
$Y \leftarrow X$	25.0%	$\neg Y \leftarrow X$	16.67%

- Knowledge bases with different representations of „requires“ and „incompatibility“ relations.
- “Preferred” representations can serve as a basis for recommending constraint refactorings.

Ongoing & Future Work

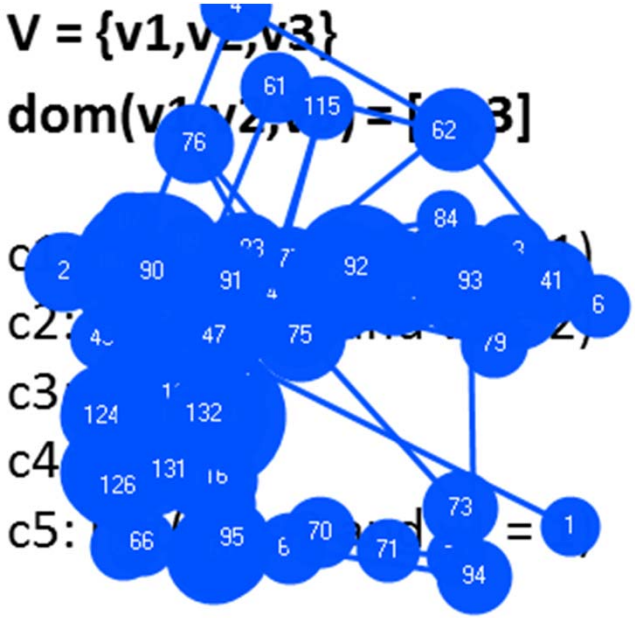
- Evaluation of further recommendation algorithms
- Misunderstandings when formalizing natural language domain descriptions (e.g., “A is only compatible with B”)
- Misinterpretations of formal sentences (e.g., translation of feature model into text-based representation)
- Exploitation of eye tracking data …

Future Work: Heatmap Visualization



- Overview of areas, knowledge engineers looked at.
- Can be used, for example, for constraint ranking.

Future Work: GazePlot Visualization



Minimal Diagnoses
- Preferred
- Based on „confusion patterns“

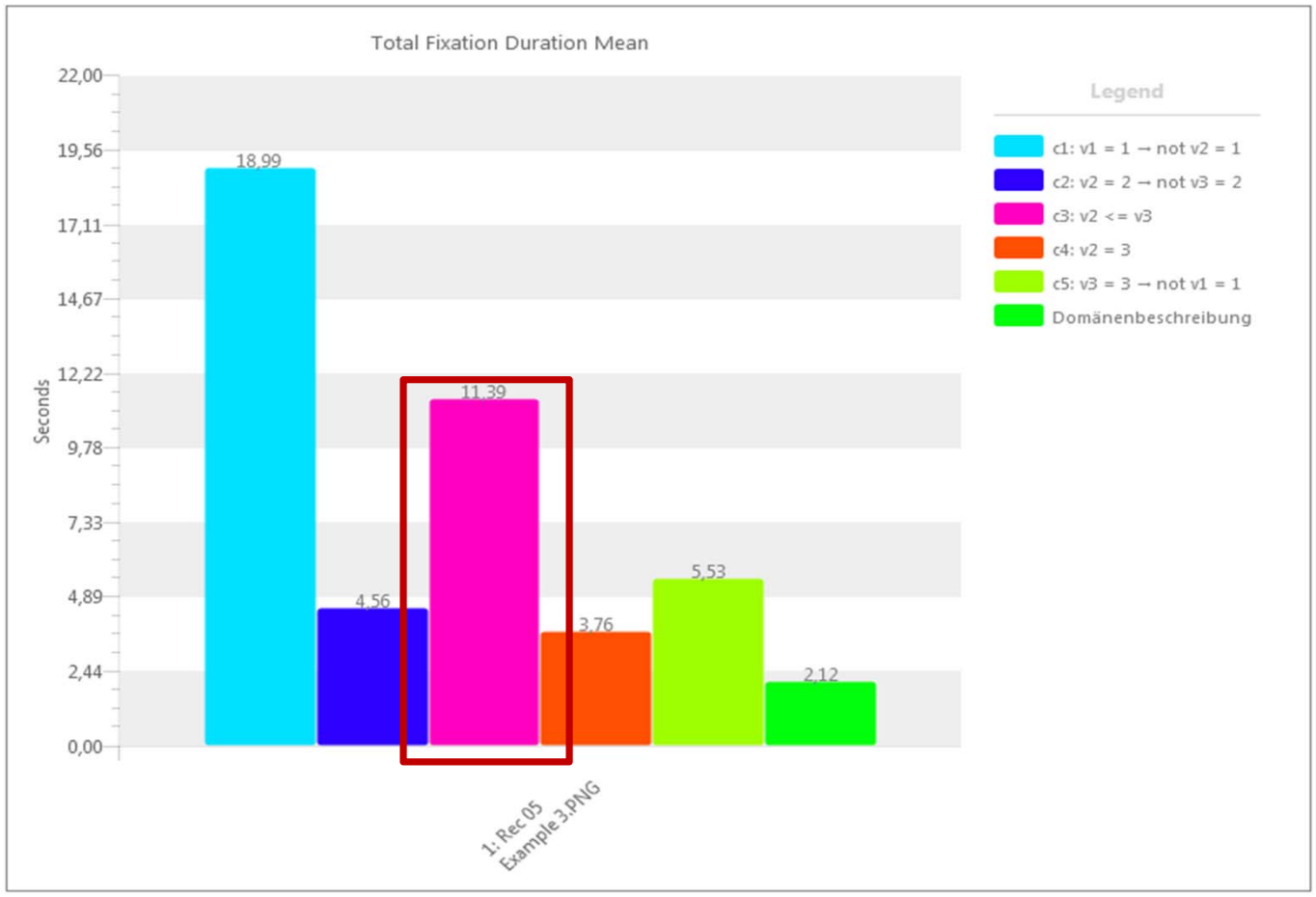
- Sequence of points, knowledge engineers looked at.
- Can be used for the detection of “confusion patterns”.

Conclusions

- Recommendation technologies can improve the quality of knowledge engineering support
- Strongly interdisciplinary research field (CS and cognitive psychology)
- Plenty of open issues, empirical research is needed
- Future work: estimation of cognitive complexity by using orthogonal data sources such as eye tracking

Thank You!

Future Work: Fixation Durations



Time used to investigate different constraints.