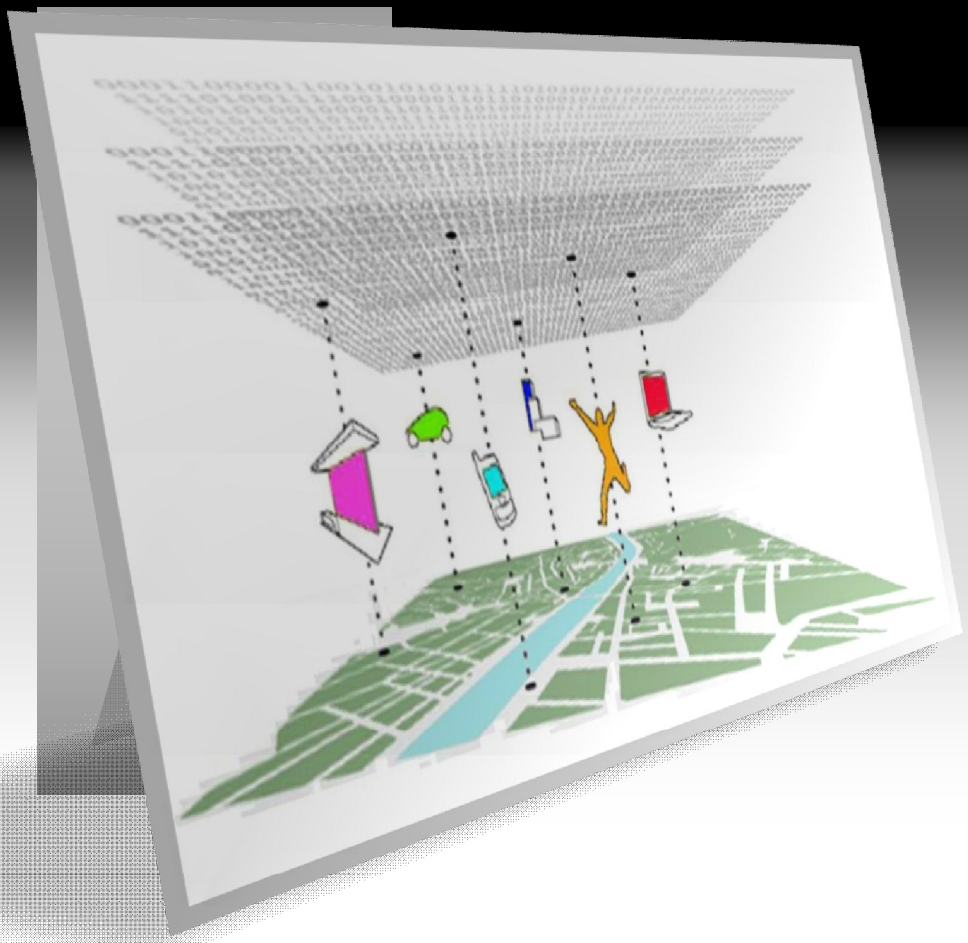
Classification of Whereabouts Patterns from Large-Scale Mobility Data



Laura Ferrari,
Marco Mamei

WOA 2010
Rimini, Settembre 2010

Agents and Pervasive Group

University of Modena and Reggio Emilia

WHEREABOUTS PATTERNS FROM MOBILITY DATA

HOME 0:00 – 8:00

HOME 18:00 – 24:00

WORK 8:00 – 18:00

WHEREABOUTS PATTERNS FROM MOBILITY DATA

CINEMA 22:00 – 24:00

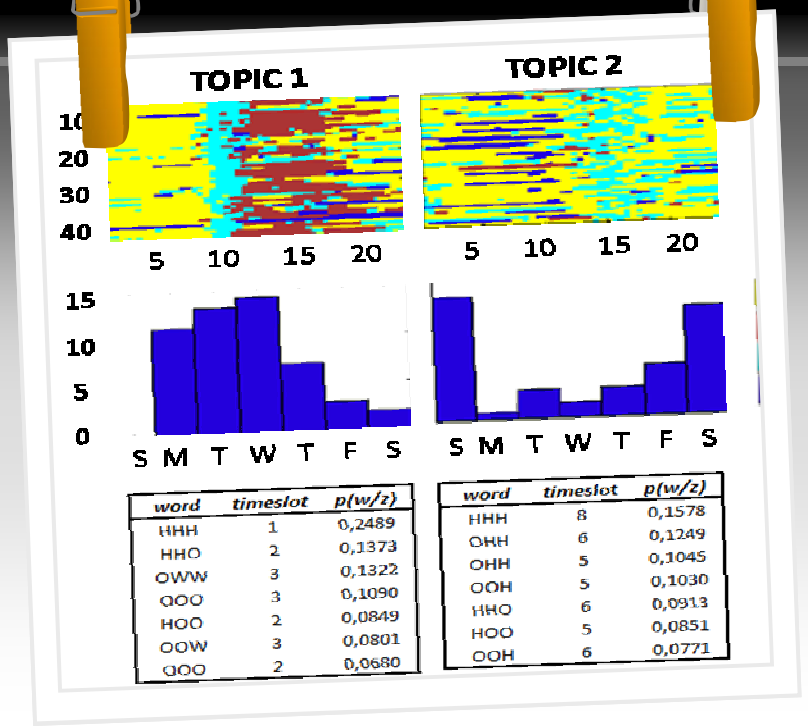
DISCO 0:00 – 2:00

HOME 2:00 – 18:00

PUB 18:00 – 22:00

EXTRACTING WHEREABOUTS PATTERNS

TOPIC MODELS

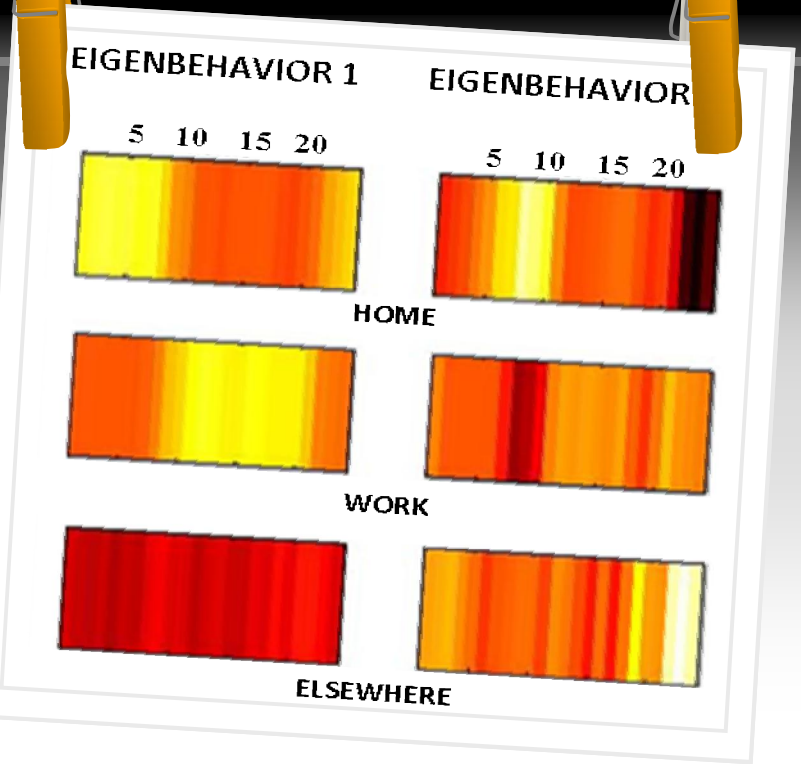


word	timeslot	p(w/z)
HHH	1	0,2489
HHO	2	0,1373
OWW	3	0,1322
OOO	3	0,1090
HOO	2	0,0849
OOW	3	0,0801
OOO	2	0,0680

word	timeslot	p(w/z)
HHH	8	0,1578
OHH	6	0,1249
OHH	5	0,1045
OOH	5	0,1030
HHO	6	0,0913
HOO	5	0,0851
OOH	6	0,0771

(Farrahi et al., 2009)

EIGENBEHAVIORS



(Eagle et al., 2009)

TOPIC MODELS FOR EXTRACTING PATTERNS

SEQUENCE OF VISITED PLACES

USERID	BEGIN	END	CELLID
22	27/08/2004 14:00	27/08/2004 16:00	102
22	27/08/2004 16:30	27/08/2004 17:30	122
...	...	...	...

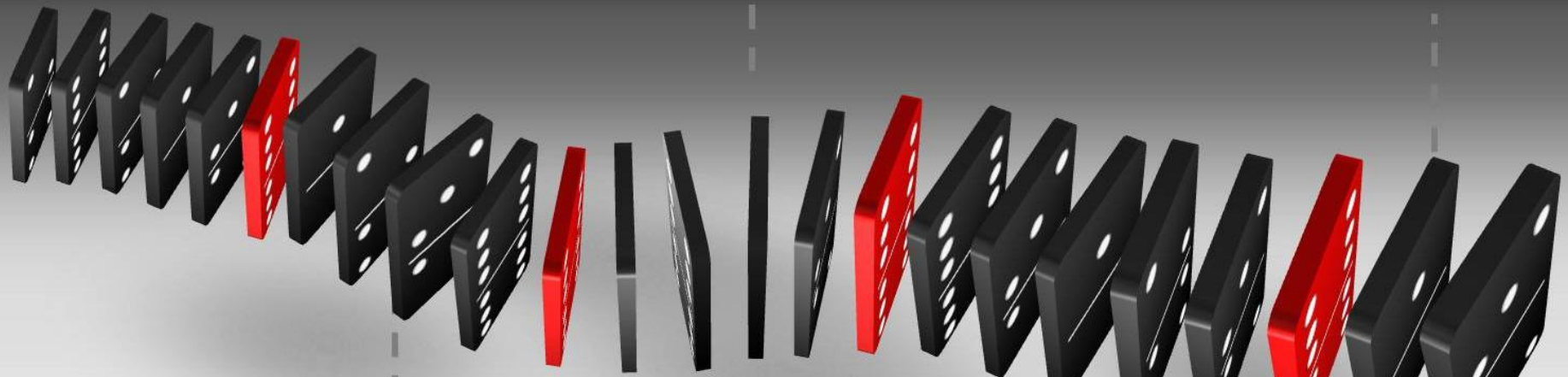
USERID	CELLID	LABEL
22	102	HOME
22	121	WORK
...	...	...

LATENT DIRICHLET ALLOCATION

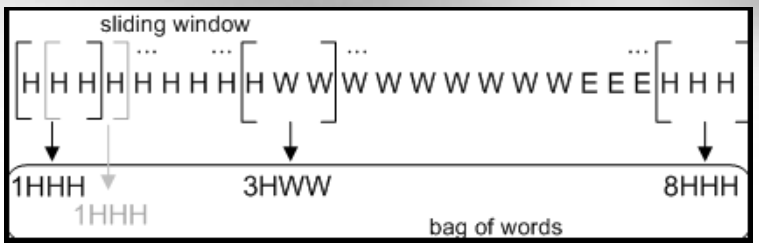
WORD	p(w/z)
HHH-1	0.2489
HHO-2	0.1379
OWW-3	
OOO-3	
HOO-2	

WORD	p(w/z)
HHH-6	0.1578
OHH-6	0.1249
OHH-5	0.1045
OOH-5	0.1030
HOO-5	0.0851

TOPIC MODELS



BAG OF WORDS



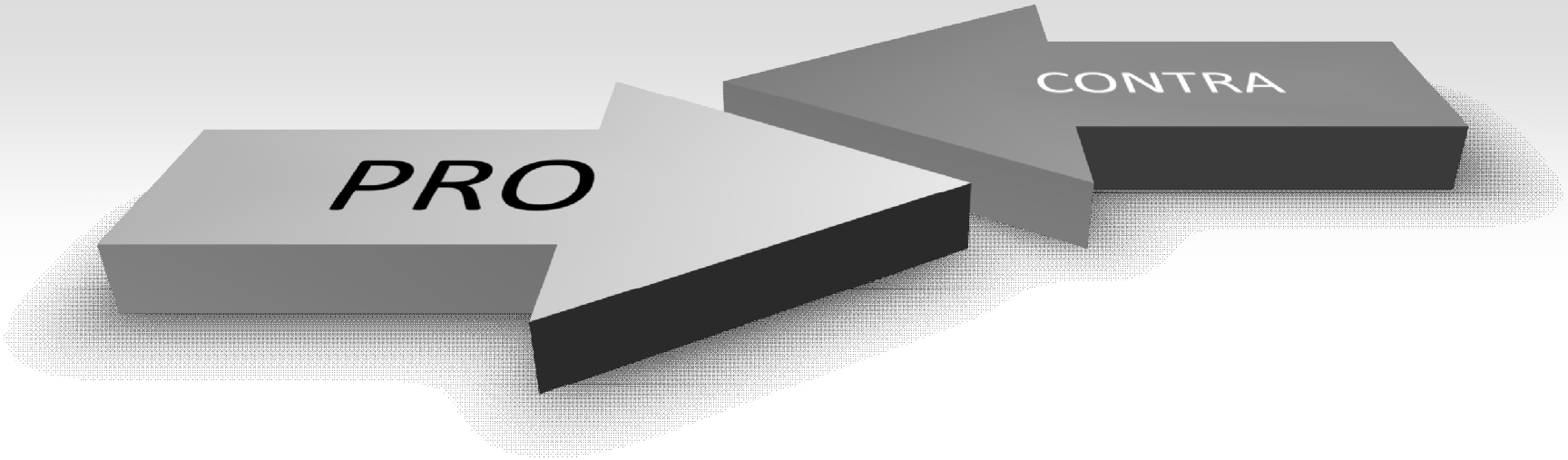
PROBLEM IDENTIFICATION

PRO

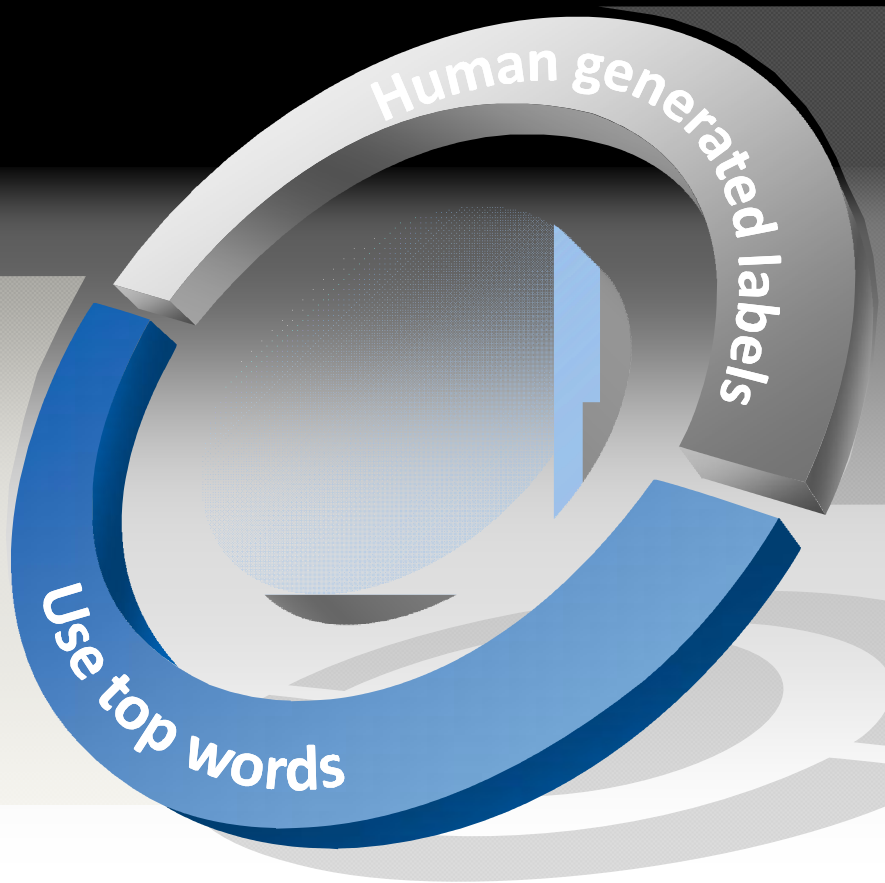
- Topical pattern analysis
- Summarization
- Subtopic discovery

CONTRA

- Predefined number of topics
- Hard to interpret



PROBLEM IDENTIFICATION



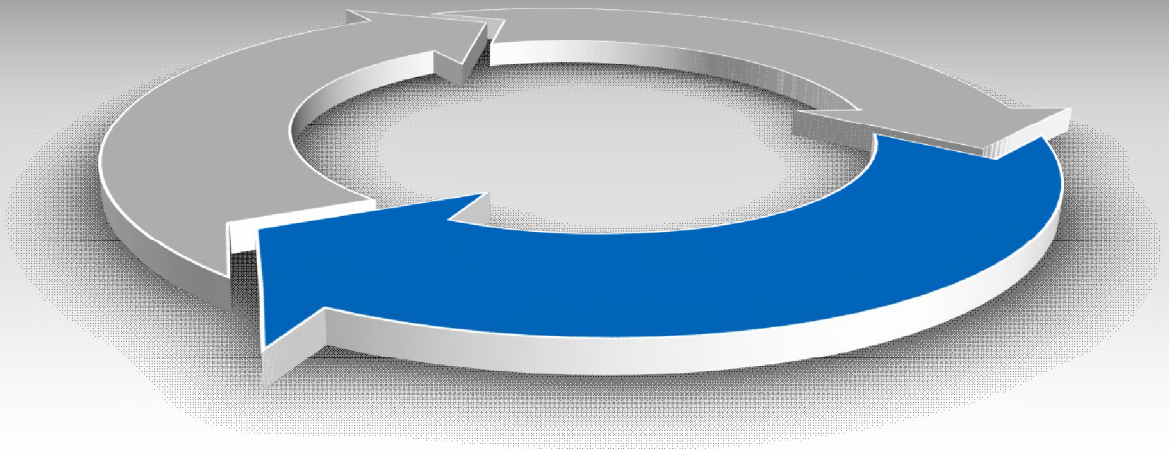
Make sense, but cannot scale up.

Automatic but hard to make sense

RESEARCH QUESTIONS

can we identify patterns from mobility data?

can we automatically generate understandable labels for topics?



can we automatically attach labels to such behavioral patterns?

APPLICATIONS OF LABELING PATTERNS

create an entry in the user blog

communicate
compact routines
affecting city-life

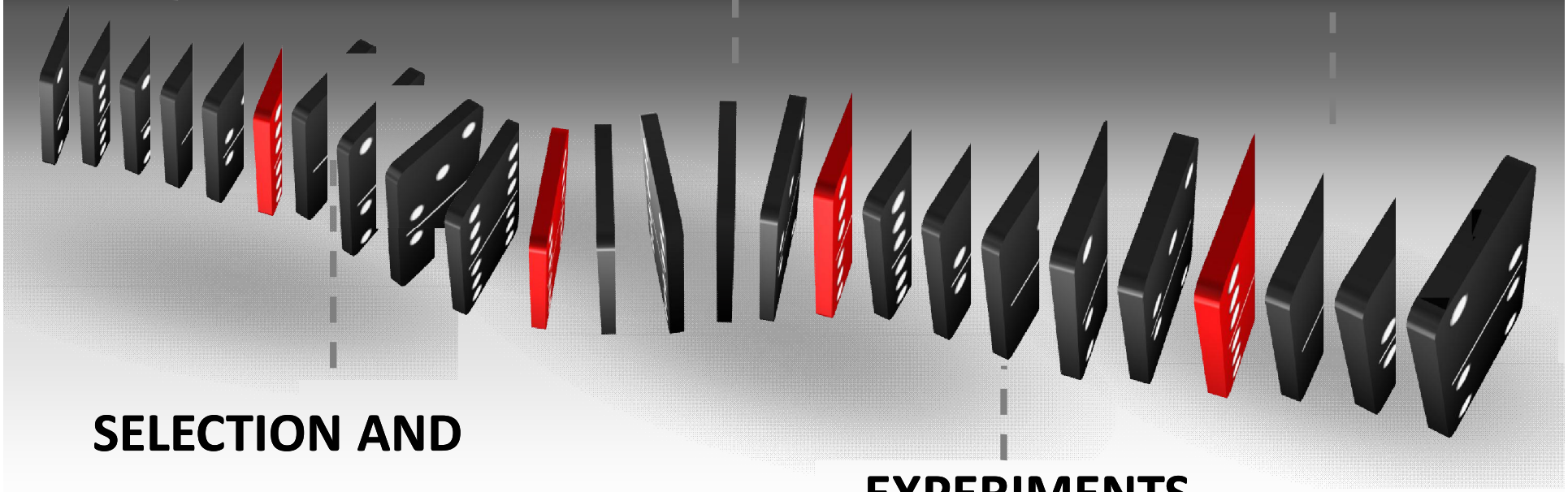
make patterns readily understandable
and usable in applications

REST OF THE PRESENTATION

OUR METHOD

LABELS RANKING: A PROBABILISTIC APPROACH

SUMMARY



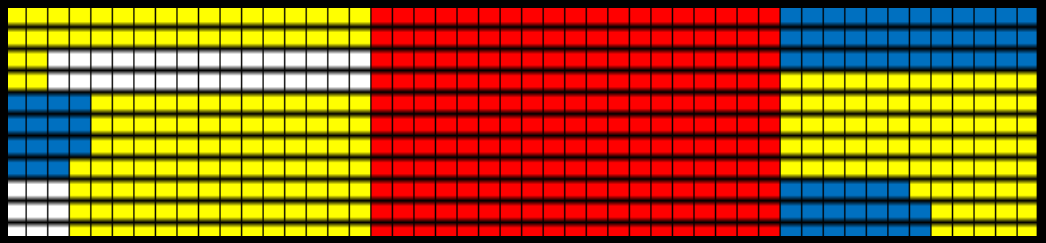
SELECTION AND REPRESENTATION OF LABELS

EXPERIMENTS

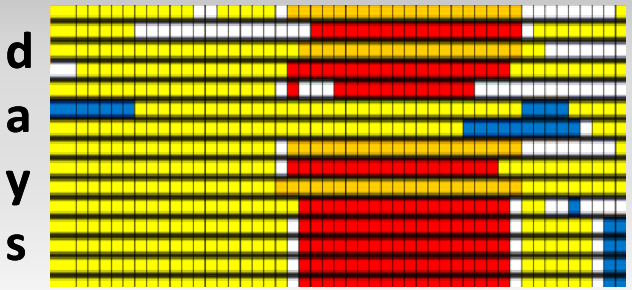
OUR METHOD

CANDIDATE LABELS POOL (E.G. "WORK 9-18", "HOME 12-14", ETC.)

LABEL PATTERN: e.g. "WORK 9-18"



REPRESENTATIONS



HHH-1	*0.1599*
HHH-2	*0.0752*
WWW-4	*0.0660*
WWW-5	*0.0372*
HHH-7	*0.0311*
EEE-5	*0.0310*
NNN-8	
HNN-8	
é	
é	

WWW-4	*0.5598*
WWW-5	*0.4978*
HHH-1	*0.0060*
NNN-2	*0.0072*
EEE-7	*0.0011*
EEE-8	*0.0010*
	0.0003
	0.0003
	0.0001

KULLBACK-LEIBLER DIVERGENCE

USER-GENERATED BEHAVIORAL PATTERNS

MULTINOMIAL WORD DISTRIBUTIONS

LABELS RANKING: A PROBABILISTIC APPROACH

user document *d* → a multinomial word distributio

a candidate label *l* → a multinomial word distribution

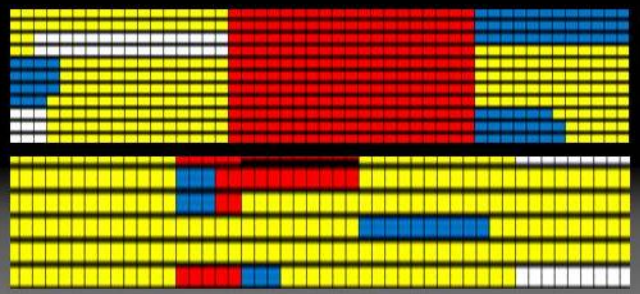
score and rank l by KL-divergence of these two multinomial word distributions

EXPERIMENTS

MOCKUP DATASET

SELECTED LABELS (E.G. "WORK 9-18", "HOME 0-12"

LABELS PATTERN



MULTINOMIAL WORD DISTRIBUTIONS

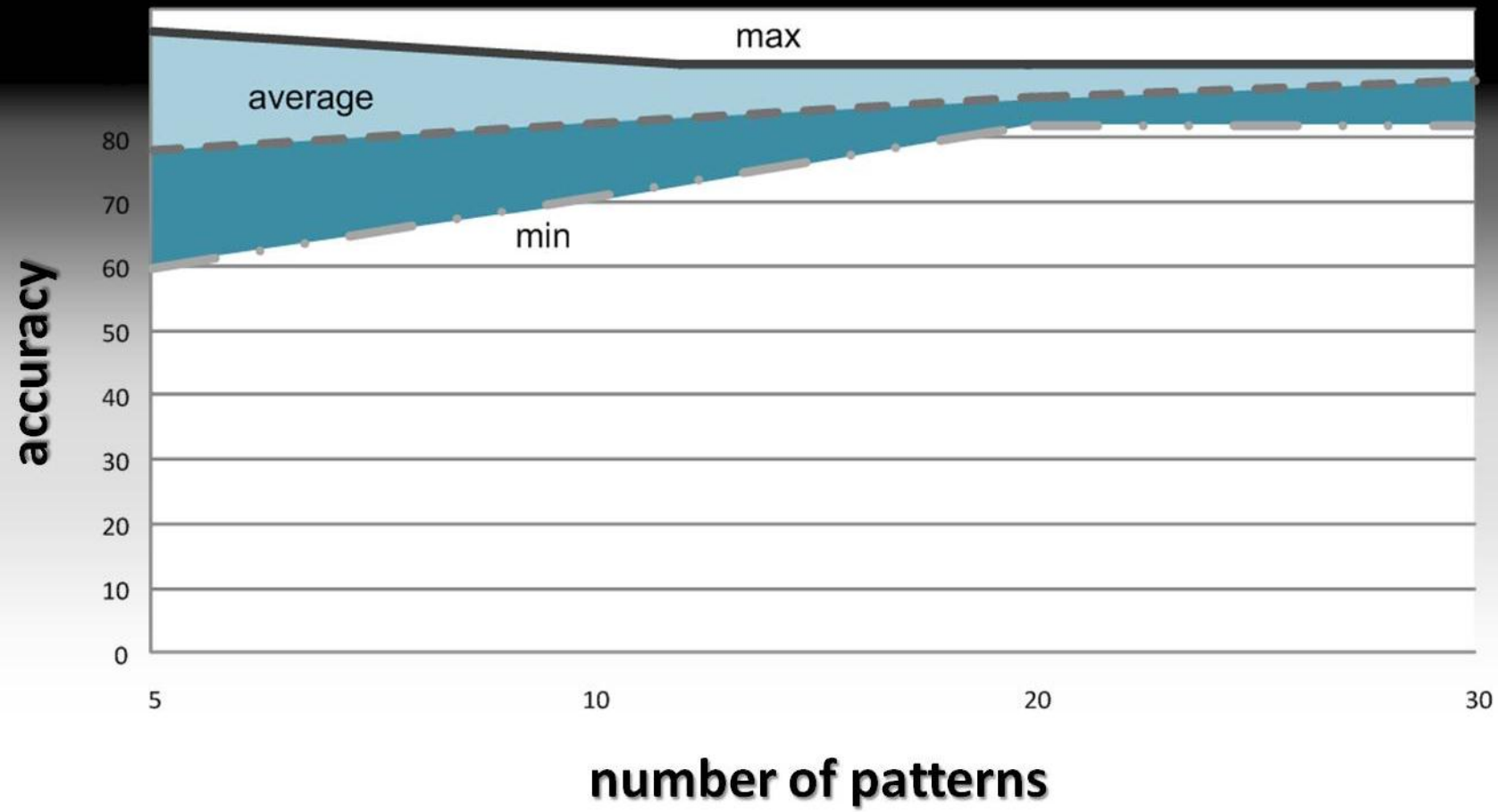
CANDIDATE LABELS POOL

CANDIDATE MULTINOMIAL DISTRIBUTIONS

KULLBACK-LEIBLER DIVERGENCE

EXPERIMENTS

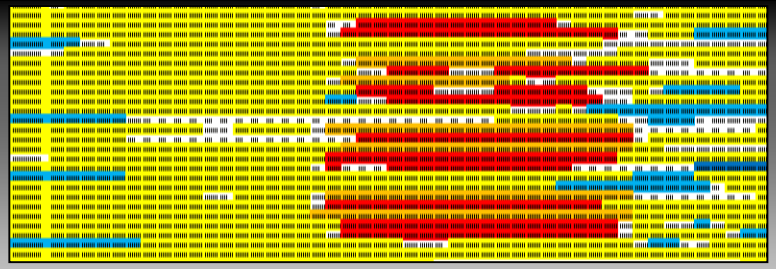
MOCKUP DATASET



EXPERIMENTS

REALITY MINING DATASET: 36 INDIVIDUALS, 121 DAYS

USER-GENERATED DAYS



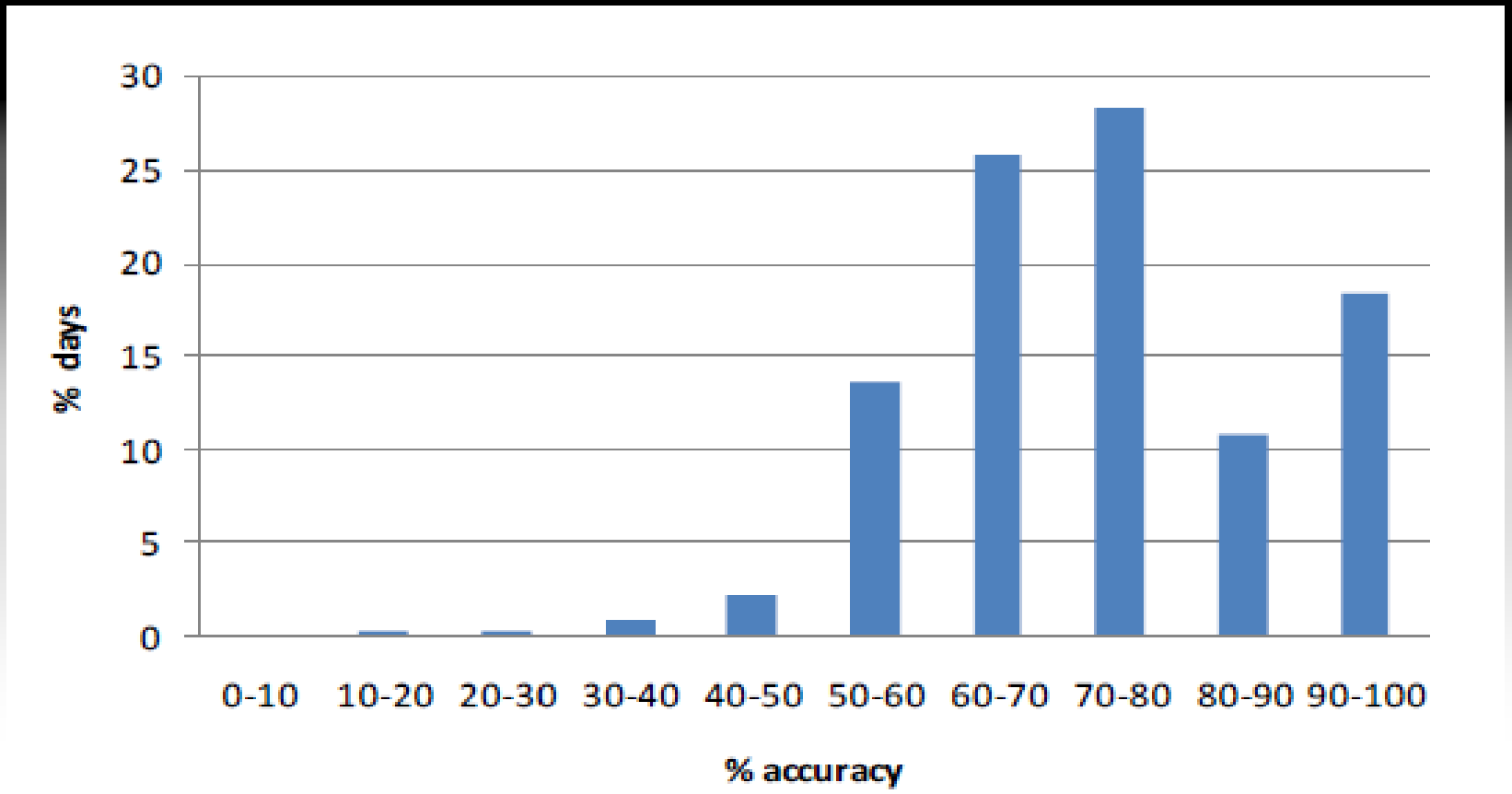
MULTINOMIAL DISTRIBUTIONS

CLASSIFICATION

DAYS RECONSTRUCTION

EXPERIMENTS

REALITY MINING DATASET: 36 INDIVIDUALS, 121 DAYS



SUMMARY

- Automatic classification of behavioral patterns
- A probabilistic approach to attach patterns with meaningful labels
- Effective when the candidate labels not represent short time-frame
- Future Work:

1. A robust way of generating candidate labels
2. A large-scale evaluation
3. A web-based visualization mechanisms to inspect and communicate whereabouts behaviors in an effective way

Thank you

laura.ferrari@unimore.it