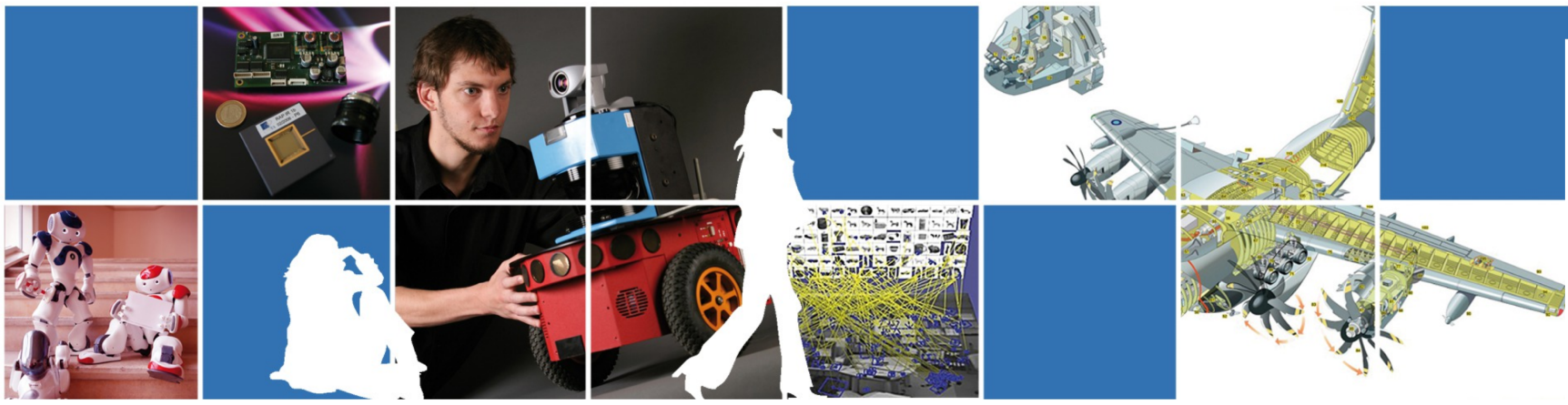

Couch potato or gym addict? Semantic lifestyle profiling with wearables and fuzzy knowledge graphs

Natalia Díaz-Rodríguez (*),
ENSTA-ParisTech, U2IS, Palaiseau, France
natalia.diaz@ensta-paristech.fr

Ignacio Huitzil and Fernando Bobillo,
University of Zaragoza, I3A, Spain
{ihuitzil, fbobillo}@unizar.es

Aki Härmä and Rim Helaoui,
Philips Research, Personal Health Dept.
Eindhoven, The Netherlands
{aki.harma, rim.helaoui}@philips.com

Umberto Straccia,
ISTI-CNR, Pisa, Italy. umberto.straccia@isti.cnr.it

(*) Work done during internship at Philips Research

Abstract

Automatic lifestyle profiling to categorize users according to their daily routine-based lifestyles is an unexplored area. Despite the current trends on having wearable devices that generate large amounts of heterogeneous data, figuring out the lifestyle patterns of people is not a trivial task. We present *Lifestyles-KG*, a knowledge graph (fuzzy ontology) for semantic reasoning from wearable sensors. It can serve as a pre-processing taxonomical step that can be integrated into further prediction techniques for intuitively categorizing fuzzy *lifestyle* concepts, treats or profiles. The ultimate aim is to help tasks such as long-term human behavior classification and consequently, improve virtual coaching or customize lifestyle recommendation and intervention programs from free form non-labelled sensor data.

Motivation

- Semantically interpretable models to better understand the statistics of individual lifestyle patterns of people and common-sense representation can enhance data-driven processes and improve accuracy and precision of recognition in human activities, while knowledge-driven human activity models such as our proposal can be upgraded through data-driven learning techniques.

Methods

- OWL 2 $\mathcal{EL}$: a fragment particularly useful in applications employing ontologies with large numbers of properties and/or classes.
- Fuzzy DL-Learner[a]: uses pFOIL-DL, a method for the automated induction of fuzzy $\mathcal{EL}(\mathbf{D})$ concept descriptions. Unlike FOIL (First-order Inductive Learning rule-based learning algorithm), in which the goodness of an induced rule is evaluated independently of previously learned rules, fuzzy DL-learner evaluates the quality of a candidate in terms of the *ensemble* of so far learned rules, and it does not face problems such as providing a mapping from fuzzy DLs to logic programming that is incomplete or its entailment can become undecidable.

[a] http://www.umbertostraccia.it/cs/software/FuzzyDL-Learner/index.html

Machine Learning workflow for fuzzy datatype learning

The long term goal is learning the rules, e.g. as general concept inclusions (GCIs); however, in this work the focus is onto:

- Design a crisp/fuzzy ontology by domain data scientists and cardiac disease specialists for *fuzzyDL* reasoner[a] to perform reasoning (e.g., queries such as max-satisfiability degrees of individuals belonging to a given concept or satisfaction degree of axioms in a Knowledge Base (KB) among others).
- Select a set of target atomic (lifestyle) concepts T for which one would like to learn classification rules.
  - If we have enough examples of instances of T, we can use fuzzy DL-Learner to obtain the rules. In particular, fuzzy DL-Learner provides fuzzy $\mathcal{EL}(\mathbf{D})$ GCIs axioms of the form "$C_1 \sqcap \ldots \sqcap C_n \sqsubseteq T$", where each C_i is either an atom A, or an existential restriction *(objectProperty some C)*, *(datatypeProperty some d)*, where d is a fuzzy membership function learned from data. Given the target concept name T, the *training set* $\mathcal{E}$ consists of crisp concept assertions of the form $a{:}T$ where a is an individual occurring in KB $\mathcal{K}$.
  - If we do not have enough examples of instances of T, we use clustering algorithms to learn fuzzy membership function values for each datatype that appears as a possible range of the attributes (data properties) for the instances of T. (This is our (unsupervised) case, and thus, we focus on learning the fuzzy datatypes; future work will add learned CGIs to the ontology).

[a] http://www.umbertostraccia.it/cs/software/fuzzyDL/fuzzyDL.html

Data

We learn fuzzy datatypes over 40 records of volunteers of middle age living in the Eindhoven area (The Netherlands). The basic unit in which a *day footprint* is divided into is a *day segment*, either by clock intervals, or divided into all places a user calls *Home*, *Work*, (frequented during commuting) *Key* or *Other* (sporadic) points.

Contributions

Lifestyles-KG, a (both fuzzyDL and crisp) Knowledge Graph that:

- Models lifestyles from heterogeneous sensor data from a large variety of wearable (sleep, HR, GSR) sensors, phone apps, scales, etc[a], and can learn his descriptive fuzzy $\mathcal{EL}(\mathbf{D})$ sub-concepts with fuzzy DL-Learner. In our unsupervised learning case, we learn fuzzy datatypes from data using (k-means, fuzzy c-means and mean-shift) clustering.
- Given a set of day (segments) records from digital/sensor traces, provides intuitive lifestyle treats categorizations.

% **Concrete concepts**
(define-primitive-concept CommuteToWork WeekdaySegment)
% **Axioms defining data properties' range**
(range walkedNStepsAtLunch *integer* 0 10000)
% **Fuzzy datatypes**
(define-fuzzy-concept MediterraneanLunchStartT trapezoidal (0, 10000, 690, 720, 780, 830)) % 11:30, 12, 13, 13:30 am
% **Fuzzy datatypes** for Day Segment *(Commuting) To Work*:
(define-fuzzy-concept LowCyclingRatioToWork left-shoulder (-10000, 10000, 0.0, 0.1))
% **Fuzzy lifestyle profile (LP) concept definitions (actual rules)**
(define-concept *MediterraneanLuncher* (g-and AtWork (some walkedNStepsAtLunch LowNSteps)(some hasLunchStartT MediterraneanLunchStartT)(some hasLunchDuration MediterraneanLunchDuration) (some spentNCalories LowCalorieCountMan)))
(define-concept *EarlyBird* (g-and NightAndMorning (some hasEndT VeryEarlyLeaveHomeTime)(some hasStartT EarlyArriveHomeTime)))
% **Input instances (individuals)**
(define-concept day1segment1 (g-and NightAndMorning (= hasStartT 0) (= hasEndT 519) (= walkedNSteps 586) (= walkedNStepsAtLunch 58) (= spentNCalories 211)
% **Queries**: max. satisfiability of digital footprint *day1segment1* matching a LP
(max-sat? (and GymAddict day1segment1))
(max-sat? (and DutchLuncher day1segment1))

Table 1: Lifestyles-KG: a lifestyle profiling ontology (excerpt of rules in *fuzzyDL*).

[a] https://github.com/NataliaDiaz/Ontologies

Conclusion and Future Work

This work presents *Lifestyles-KB* fuzzy ontology and focuses on learning fuzzy datatypes; future work should:

- Avoid design pitfalls (e.g. with OOPS! Linked Data Tool), apply use-case adapted tools towards coverage evaluation, e.g., CQOA (Competency Question-driven Ontology Authoring) tests, and integrate data from other (e.g., OpenSNP, Sussex-Huawei Locomotion) datasets.
- Automatize larger part of the KB design procedure (concepts, relations, cardinality, functionality, etc.), e.g, using deep nets for relational reasoning.