Non-Temporal Orderings for Extensional Concept Drift

Albert Meroño-Peñuela[1,2] Stefan Schlobach[1]

[1]Department of Computer Science, VU University Amsterdam, NL

[2]Data Archiving and Networked Services, KNAW, NL

October 22th, 2013
First International Workshop on Semantic Statistics
ISWC 2013

As the world changes continuously, concepts also change their meaning over time. We call this *concept drift*

We call this *concept drift*

- Smooth transitions
- Radical transitions (*concept shift*)

Concept drift takes place *over time*. **Time** is the intuitive **ordering** of the data.

Dutch historical census data:

1795 1830 1840 1849 1859 1869 1879 1889 1899 1909 1919 1920
1930 1947 1956 1960 1971

Concept drift takes place *over time*. **Time** is the intuitive **ordering** of the data.

Dutch historical census data:

~~1795 1830 1840 1849 1859 1869 1879 1889 1899 1909 1919 1920 1930 1947 1956 1960 1971~~

But no time series in the *challenge* data!

Australia: 2011
France: 2010

1. Resignation
2. Discard time for this experiment

No time, no party?

Consider time as just one of all possible dimensions that can be used as data **orderings**.

Meaningful orderings (i.e. intrinsically dynamic):

- GDP per capita
- Stock markets
- Population density
- (Il)literacy

Can we detect drastic extensional drifts over orderings other than time?

E.g. Is there drift in the concept of *youth unemployment*[1] in Australia/France as a *selected ordering* grows?

Australia: GDP per capita per state
France: Population density per *departement*

[1] 15-24 years old considered standard

```
PREFIX qb: <http://purl.org/linked-data/cube#>
PREFIX d2s: <http://www.data2semantics.org/core/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX ns: <row_property_URI>

SELECT ?d1label ... ?dnlabel ?p1label ... ?pmlabel ?population
FROM <named_graph_URI>
WHERE {
 ?cell d2s:isObservation [ a qb:Observation ;
                           qb:DimensionProperty ?d1 ... ?dn ;
                           ns:property1 ?p1 ;
                           ...
                           ns:propertym ?pm ;
                           qb:MeasureProperty ?population ] .
 OPTIONAL {
 ?cell d2s:isObservation [ns:propertyk ?pk ] .
 ?pk skos:prefLabel ?pklabel . }
 ...
 OPTIONAL {
 ?cell d2s:isObservation [qb:DimensionProperty ?dl ] .
 ?dl skos:prefLabel ?dllabel . }
 ?pt skos:broader ?pu .
 ?pu skos:broader ?pv .
 d1 ... dn skos:prefLabel ?d1label ... ?dnlabel .
 p1 ... pm skos:prefLabel ?p1label ... ?pmlabel .
 FILTER (?d1 IN (v1, ..., vr)) ...
 FILTER (?dn IN (w1, ..., ws))
}
```

SPARQL template for unfolding RDF Data Cubes

Case-Study
Querying

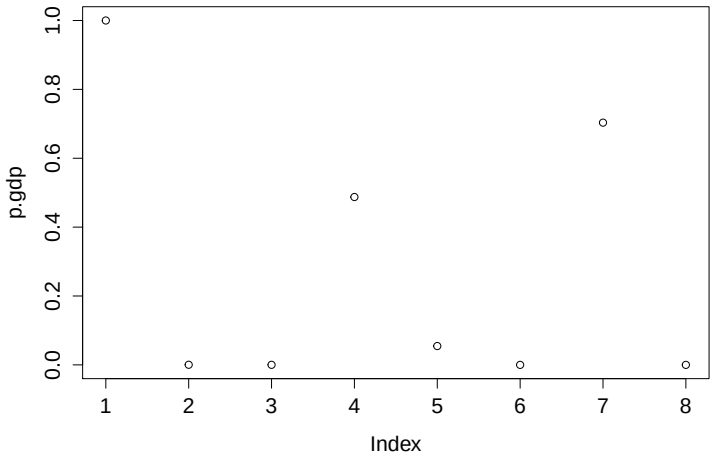
```
PREFIX qb: <http://purl.org/linked-data/cube#>
PREFIX d2s: <http://www.data2semantics.org/core/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX ns: <row_property_URI>

SELECT ?d1label ... ?dnlabel ?p1label ... ?pmlabel ?population
FROM <named_graph_URI>
WHERE {
 ?cell d2s:isObservation [ a qb:Observation ;
                           qb:DimensionProperty ?d1 ... ?dn ;
                           ns:property1 ?p1 ;
                           ...
                           ns:propertym ?pm ;
                           qb:MeasureProperty ?population ] .
 OPTIONAL {
 ?cell d2s:isObservation [ns:propertyk ?pk ] .
 ?pk skos:prefLabel ?pklabel . }
 ...
 OPTIONAL {
 ?cell d2s:isObservation [qb:DimensionProperty ?dl ] .
 ?dl skos:prefLabel ?dllabel . }
 ?pt skos:broader ?pu .
 ?pu skos:broader ?pv .
 d1 ... dn skos:prefLabel ?d1label ... ?dnlabel .
 p1 ... pm skos:prefLabel ?p1label ... ?pmlabel .
 FILTER (?d1 IN (v1, ..., vr)) ...
 FILTER (?dn IN (w1, ..., ws))
}
```

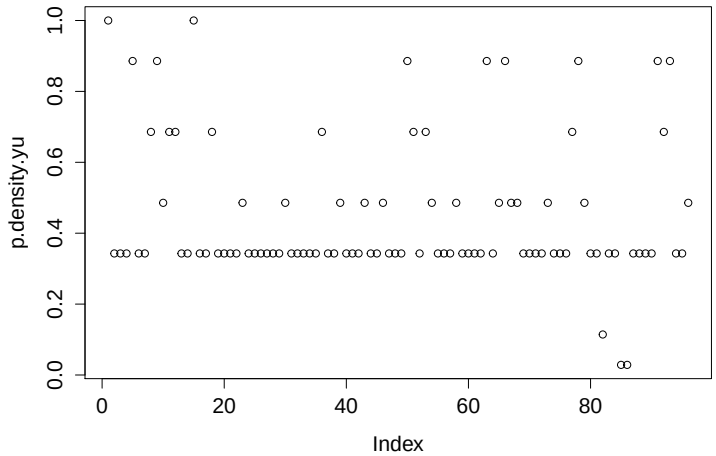
Data sources: local endpoint, DBPedia (for the orderings)

Age range	Gender	Location	Occupation	Population	**GDP**
15-19 years	Male	New South Wales	Unemployed, Total	17456	**57828**

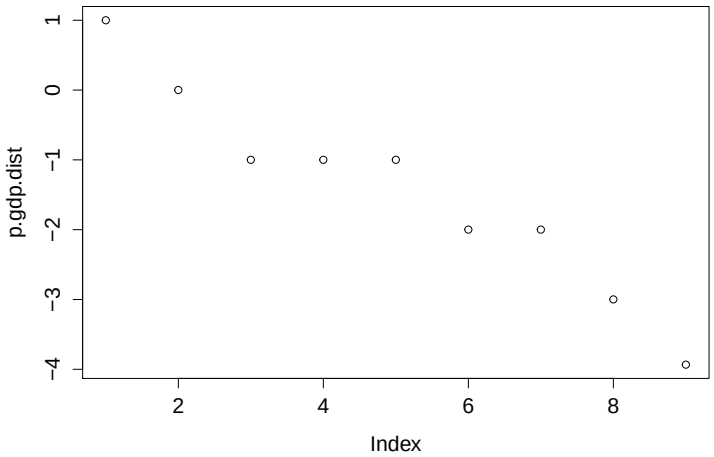
Age range	Gender	Location	Occupation	Population	**Density**
15 to 19 years	Women	Tarn	Unemployed	1.345e03	**64.17**



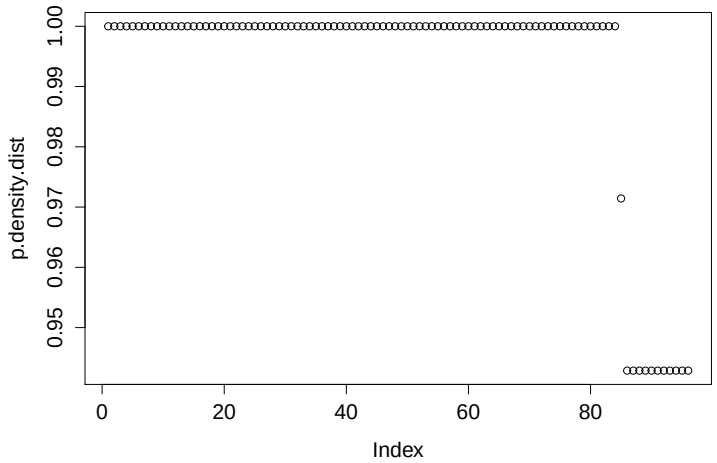
Drift of youth unemployment in Australian states using GDP per capita as ordering



Drift of youth unemployment in French *departements* using population density as ordering



Drift of youth unemployment in Australian states using GDP per capita as ordering (smooth function)



Drift of youth unemployment in French *departements* using population density as ordering (smooth function)

Conclusions

- Concept drift: concepts change over time
- No time, but **party**
- Meaningful orderings as valuable as time to study concept drift (smooth transition assumption)
- Takes full advantage of Semantic Web data

Thank you
Questions, suggestions?

@albertmeronyo
http://www.cedar-project.nl
http://www.data2semantics.org