TECHNICAL MEETING OF THEMATIC PARTNERSHIP "TRACEABILITY AND BIG DATA"

Smart Specialization Platform S3P Agrifood

TOPIC 4. CROSS-CUTTING TOPIC: OPEN DATA, INTEROPERABILITY, DATA GOVERNANCE AND INFORMATION SECURITY, CYBER SECURITY.

Pilot Projects

- DATA SETS PRODUCTION AND MANAGEMENT
  - (Big) data analysis with temporal component
  - Making EU RASFF **alert system** (Food and Feed Safety Alert) more friendly and easy-to-use for food chain (including restaurants, fast foods, ecc.)
  - Statistical matching to allow **different sources data integration**, but also to enable mutual and integrated access to genetic resources and related information
- BIG DATA INTEGRATION and LINKED OPEN DATA
  - Develop public databank of **markers** composition of food product (i.e. olive oil). Improving exchange of data along the food value chain, improving the access of crop scientists, farmers and food processors to consumer feedback on products. To improve traceability in the food chain and reduces information costs.

TITLE: Temporal Data Analysis (with application to temporal traceability)

OBJECTIVES: Developing and experimenting new temporal analysis techniques, and proving that using temporal information may increase the reliability of learned models

RESULTS: new temporal decision models that extend and improve the reliability of existing ones by taking into account dynamic aspects of the information in an explicit fashion.

Data Analysis

- Consider a collection of documental tracing data such as a smart label. By nature, the **history** of the traced product is **temporal** information.
- Its history contains the movements of the product along the timeline, from its birth to its death.
- The **events** that characterizes the product may be *punctual* (‘arrived at distribution station 1’) or *extended* (‘being exposed to too high temperature’ for so long)
- **Analyzing** such data presents a problem: how do we represent the evolution of the products and all its meaningful events?

Data Analysis Now

- Model learners *today* may be used in a completely **atemporal way**, that is: temporal information may be simply discarded and products identified with the collection of their **static data** and **categorical pseudo-temporal data** (i.e., ‘the product has passed through more than 3 distribution points’, ‘the product has been exposed to high temperature in more than one occasion’).
- A **good data scientist** may use **frequencies, pattern,** and, in general, may be able to **flatten** temporal information in a **static description,** allowing available learners to take some temporal information into account.

Temporal Data Analysis

- Within this project we aim to produce model learners that take into account temporal information via: better and improved temporal information representation and suitable extension of learning processes
- We use: extension of Decision Tree Learner with Pattern Decision and Interval Logic Rules (fuzzy) Extraction
- We model temporal information via pattern description languages and interval logic languages, allowing us to express universal as well as existential information: ‘Whenever the product has been exposed to high temperature during more than 3 minutes, the distribution station has discarded it, with probability higher than 90%’

TITLE: CEREALAB

OBJECTIVES: Data Integration and Open Data Access

RESULTS: An information system for breeders

Another success story: CEREALAB

- The DBGROUP, in collaboration with BIOGEST-SITEIA successfully applied Data Integration and LOD in the agri-food domain: the MOMIS system was used to integrate genotypic and phenotypic data, for the development of the CEREALAB database (www.cerealab.org), a tool realized for marker-assisted selection of wheat, barley and rice; it helps cereal breeders to choose molecular markers associated to economically important phenotypic traits.
- A methodology was also implemented for facilitating resource providers in publishing public data into the Linked Open Data (LOD) cloud, and for helping consumers to efficiently access and query them; we applied this methodology to the CEREALAB database

For more information
guido.sciavicco@unife.it