

Dynamic Terminology Enhancement for Integrated ECG Resources

Alexandra Kokkinaki*, Ioanna Chouvarda and Nicos Maglaveras

Lab of Medical Informatics, The Medical School, Aristotle University of Thessaloniki, Thessaloniki, Greece

Public Biosignal Resources, containing signals and annotations, can serve an important role in biomedical research and education, provided that an appropriate framework allows for population, querying and data/metadata access of these resources. In this paper, we present Dynamic Terminology Enhancement Method (DTEM) to support enrichment and extensibility in a biosignal integration system called ROISES (Research Oriented Integration System for ECG signals), which integrates diversely encoded ECG signals and the corresponding annotation and metadata. The DTEM method manipulates the ontological structure of the system whose taxonomy is based on UMLS and ecgML, to dynamically enhance its terminology from the newly added sources. The method is performed in two successive steps. The first step investigates the suitability of the term to constitute a class in the ontologies hierarchy, by comparing it with UMLS Metathesaurus terms. To perform the comparison, the system engages Support Vector Machines and employs semantic and syntactic measures based on UMLS characteristics. The second step determines the most suitable point in the ontologies taxonomy to place the previously selected terms. The method constructs a vector whose elements constitute similarity indices of the terms UMLS ancestor hierarchy list compared to the required ontology, and adopts decision tree algorithms to decide upon the most suitable hierarchy to be subsumed by the ontology. DTEM was applied to ROISES to support the extensibility of its schema. In the first step, the testing dataset contained a set of medical terms originating from medication, medication categories and free text entries for medical history, not present in the global ontology. The procedure resulted in 98.6% sensitivity and 77.7% specificity and classified the suitable from the inapt classes. The second step trained a decision tree with a dataset of ancestor hierarchies related to the previously acquired terms. The testing procedure resulted in 100% sensitivity and 96.0% specificity.