

Biological event extraction using SVM and composite kernel function

Maha Amami*, Aymen Elkhelifi**, Rim Faiz***

*LARODEC, ISG de Tunis, 2000, Le Bardo, Tunisie

MahaAmami@isg.rnu.tn,

**LaLIC, Université Paris-Sorbonne, 28, rue Serpente, 75006 Paris, France

Aymen.Elkhelifi@paris4.sorbonne.fr

***LARODEC, IHEC de Carthage, 2016, Carthage Présidence, Tunisie

Rim.Faiz@ihec.rnu.tn

Abstract

With an overwhelming of experimental and computational results in molecular biology, there is an increasing interest to provide tools that will automatically extract structured biological information recorded in freely available text. Extraction of named entities such as protein, gene or disease names and of simple relations of these entities, such as statements of protein-protein interactions has gained certain success, and now the new focus research has been moving to higher level of information extraction such as co-reference resolution and event extraction. It is precisely the last of these tasks which will be focused in this paper. The biological event template allows detailed representations of complex natural language statements, which is specified by a trigger and arguments labeled by semantic roles.

In this paper, we have developed a biological event extraction approach which uses Support Vector Machines (SVM) and a suitable composite kernel function to identify triggers and to assign the corresponding arguments. Also, we make use of a number of features based on both syntactic and contextual information which were automatically learned from the training data.

We implemented our event extraction system using the state-of-the-art of NLP tools. We achieved competitive results compared to the BioNLP'09 Shared task benchmark.

Résumé

Étant donné l'importance des résultats scientifiques publiés dans le domaine de la biologie moléculaire, il est nécessaire de mettre en place des outils aidant les scientifiques à extraire des informations précises et structurées. L'extraction des entités nommées telles que les protéines, les gènes et les relations simples telles que les interactions entre les protéines, les associations entre les gènes et les maladies sont des tâches quasiment résolues.

Désormais, les recherches en extraction d'information biologique se sont orientées à la résolution des tâches plus complexes telles que la coréférence et l'extraction des événements. Dans cet article, nous nous intéressons au particulier à l'extraction des représentations complexes des événements biologiques. Une représentation d'un événement biologique dans un

Biological event extraction using SVM and composite kernel function

langage naturel comprend une ou plusieurs unités linguistiques désignant un déclencheur d'événement et des entités biologiques participantes à cet événement.

Dans cet article, nous avons présenté notre approche d'extraction des événements biologiques qui utilise les machines à vecteurs de support (SVM) et une fonction noyau composite afin d'identifier les déclencheurs des événements (event trigger) et les participants à cet événement (event argument). Aussi, nous avons utilisé un nombre d'attributs basés sur les informations syntaxiques et contextuelles, générés automatiquement à partir des documents d'apprentissage.

Nous avons implémenté notre système d'extraction des événements en utilisant des outils de traitement automatique de langage naturel (TALN). Nous avons obtenu des résultats compétitifs par rapport au benchmark de BioNLP'09 Shared Task.