Biological event extraction using SVM and composite kernel function

Maha Amami*, Aymen Elkhlifi**, 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.Elkhlifi@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 where 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.

1 Introduction

The past decade has seen an explosive growth in the amount of experimental and computational biological data. This growth is accompanied by an increase in the number of biological texts discussing the results. The MEDLINE [1] database contains in 2010 over 20 million articles, and the database is currently growing at a rate of more than 10% each year (Ananiadou and al., 2006). The availability of huge textual resources provides the scientists with the chance

1. http://www.ncbi.nlm.nih.gov