Challenge LLL
Scoring Software User's Manual

Julien Gobeill

24rd march 2005

1.INTRODUCTION

In order to evaluate systems competing in LL Challenge, the Organization committee will compute scores and measures with a scoring software. This is downloadable and usable by participants. The purpose of this document is to explain in detail its functionality and usage.

The official scoring software for Challenge LLL is called giec-eval.pl ; it is available to download on the web page of the Challenge (http://genome.jouy.inra.fr/texte/LLLchallenge/). The software is implemented in Perl.

2.USAGE

The software requires a Perl interpreter. The syntax of the command line for running it is the following one :

perl giec-eval.pl -k <keys-file> -p <predictions-file> -d <dictionary-file> (-v | -s: | -w | -b:)

The software requires at least three parameters:

- k <keys-file> : the path to the keys-file. The keys-file is the file containing the correct interactions as found by a human expert. It can be for example the training datas, or a subpart of it. More details about this file and its format are given in section 3.1.

- p <predictions-file> : the path to the predictions-file. The predictions-file is the file containing the interactions to be evaluated, usually found by the competing system. More details about this file and its format are given in section 3.2 .

- d <dictionary-file> : the path to the dictionary-file. The dictionary-file is the file containing all supported canonical forms of protein and gene names, and their synonyms. It is downloadable from the web page of the Challenge. More details about this file are given in section 3.3 .

Other options, indicated between brackets, are not required to run the software. They are detailed in section 5.

An example of a typical call of the program is:

```
perl giec-eval.pl -k exempleOfKeysFile.txt -p exempleOfPredictionsFile.txt
            -d exempleOfDictionary_data.txt
```

3. INPUT FILE FORMAT

3.1 keys-file

The keys-file contains the training data as downloadable from the web page Challenge ; notice that both the "basic" and the "enriched" version are supported by the software. One can also modify this file, or create his one.

More formally, the keys-file contains sets of 6 lines in the "basic" version (respectively 8 lines in the "enriched" version), each set corresponding to one sentence. Here is an example of such a set for "basic" format:

```
ID          10515909-4
sentence    Expression of the sigma(K)-dependent cwlH gene depended on gerE.
words       word(0,'Expression',0,9)   word(1,'of',11,12)          word(2,'the',14,16)       word
(3,'sigma(K)',18,25)        word(4,'dependent',27,35)       word(5,'cwlH',37,40)      word
(6,'gene',42,45)            word(7,'depended',47,54)        word(8,'on',56,57)        word
(9,'gerE',59,62)
agents      agent(3)    agent(9)
targets     target(5)
genic_interactions         genic_interaction(3,5)     genic_interaction(9,5)
```

Each line starts with a field name ("ID", "sentence", "genic_interactions"...) followed by the literals, separated by a tabulation. People are free to insert blank lines or comments between sets, but not inside. The processing of a set is triggered by the keyword ID at the beginning of a line.

Please notice that *the control of the syntax is strict.* For more details, refer to the description of the training data format as defined by the Challenge web page. Agents, targets and genic_interaction arguments are IDs. Each agent or target declared has to be used at least once in a genic interaction and each interaction must contain one declared agent and one declared target.

3.2 predictions-file

The predictions-file is computed by a competing software. It contains sets of 4 lines, each set corresponding to one sentence. Here is an example of such a set :

```
ID              10515909-4
agents          agent('sigK') agent('gerE')
targets         target('cwlH')
genic_interactions        genic_interaction('sigK','cwlH')        genic_interaction('gerE','cwlH')
```

The format is the same as the keys-file one, except that *agents and targets are identified by their biological names* (between simple quotes) and not by their IDs. As announced, *only canonical forms of agents and targets are allowed*. As for the key-file, people are free to insert blank lines or comments between sets, but not inside. For each sentence, even if there are no agents, no targets and no interactions to found, the field names "agents", "targets" and "genic_interactions" must be present.

3.3 Dictionary file

The dictionary-file is the named-entity dictionary downloadable from the web page. One can modify this file, or create his own one. Synonyms are separated by tabulations.

4.OUTPUT

The scoring software computes distinct scores for each sentence, and scores and measures for the whole file.

Correct answer for each sentence is computed in a strict way: for a given ID, to be counted as correct, an interaction in the result-file must be exactly the same as in the key-file. The computed scores for a sentence S are:

COR (correct): number of interactions of S in the result file that are in the key-file for this sentence.
MIS (missing): number of interactions of S in the key-file that are missing, ie are not in the predictions-file for this sentence.
SPU (spurious): number of interactions of S in the result-file that are wrong, ie are not in the keys-file for this sentence.

The scores sentence by sentence (*scoring details*) are not given in the default mode. It can be obtained by the -v option. In all cases follows the *scoring summary*, that is the scores and measures for the whole file. It is simply the sum of the scores for each sentence.

The software outputs then 3 measures, computed from the scoring summary:

$$\text{PRE (precision)} = \frac{COR}{(COR + SPU)}$$

$$\text{REC (recall)} = \frac{COR}{(COR + MIS)}$$

$$\text{FM (f-measure)} = \frac{((beta^2 + 1) * PRE * REC)}{((beta^2 * PRE) + REC)}$$

These 3 measures are given with 2 significant digits; this number can be modified with the -s option. F-measure is computed with beta=1 ; beta can be modified with the -b option (see below).

Finally the software outputs warnings if some sentences of the key-file are missing in the result-file or if some sentences of the result-file refers to no sentence of the key-file; this warning is output by default, but can be omitted with the -w option.

The output is the standard output.

5.OPTIONS AND PARAMETERS

The supported options are the following ones:

-h (help) : displays the usage and help, does not run the program
-k (keys-file) :   path of the keys-file. *Mandatory*.
-p (predictions-file) : path of the predictions-file. *Mandatory*.
-d (dictionary-file) : path of the dictionary-file. *Mandatory*.
-v (verbose) : outputs the detail scores, sentence by sentence. *Optional*.
-s $[0..8] \in \mathbb{N}$ (significant digits) : sets the number of significant digits of the measures. Expects an integer as argument, which must be between 1 and 8. *Optional*
-w (warnings) : set off the default output of warnings, which report missing IDs in the key-file or in the result-file. *Optional*
-b $[0..10] \in \mathbb{R}$ (beta) : sets the value of beta for computing the F-measure. Expects a float as argument, which must be between 0 and 10. *Optional*