Grouping pharmacovigilance terms with semantic distance

Marie DUPUCHab, Anne JAMETbc, Marie-Christine JAULENTab, Reinhard FESCHAREKd, Natalia GRABAREHe

a Université Pierre et Marie Curie - Paris6, Paris, F-75006 France; b INSERM, U872 eq. 20, Paris, F-75006 France; c HEGP AP-HP, Paris, France; d CSL Behring GmbH, Marburg, Germany; e CNRS UMR 8163 STL, Université Lille 3, France

Abstract. Pharmacovigilance is the activity related to collection, analysis and prevention of adverse drug reactions (ADRs) induced by drugs or biologics. Beside other methods, statistical methods are in use to detect new ADR and it was noted that their combination with groupings of terms gathering similar ADRs allows to improve the detection of new ADRs. SMQs, reference groupings in the pharmacovigilance area, are built thanks to the exploitation of the MedDRA structure. Currently SMQs are over-inclusive, although they can miss several relevant terms. Moreover, several important security topics are not covered by the SMQs. The objective of this work is to propose an automatic method for the creation of groupings of terms. This method is based on exploitation of the semantic distance between MedDRA terms. Through several experiences performed, we obtain a high precision and an acceptable recall.

Keywords. Natural Language Processing, Medical informatics, Drug toxicity, Pharmacoepidemiology, Semantics, Terminology

1. Introduction

Pharmacovigilance is the activity related to the collection, analysis and prevention of adverse drug reactions (ADRs) induced by drugs or biologics. ADRs are encoded with terms from dedicated terminologies, WHO-ART (World Health Organization - Adverse Reaction Terminology) and MedDRA (Medical Dictionary for Drug Regulatory Activities). Signal detection – unexpected relation between a drug and an ADR – exploits this encoding. Statistical methods are typically used for signal detection in data banks with sufficiently large volume of cases [2, 11], but for increasing of the signal intensity, groupings of similar pharmacovigilance cases is suitable [8]. This need becomes more relevant because the structure of MedDRA is very granular: use of very specific terms for encoding of ADRs may cause a dilution of signal [7]. Thus, various hierarchical levels of MedDRA (SOC, HLT, PT) and more recently SMQs (Standartized MedDRA Queries) have been used. The PT (Preferred Terms) level, very often exploited for the signal detection, corresponds mainly to specific ADRs, while SOCs (System Organ Class) and HLTs (High
Level Terms) correspond to MedDRA hierarchical levels, which are higher than PTs and provide with hierarchical groupings of PTs. As for the SMQs, their objective is to group together terms associated with a given diagnosis. SMQs are designed by groups of experts further to the manual analysis of the MedDRA structure and of the scientific literature [5]. This is a long and tedious task. Few evaluation works of these groupings have demonstrated that SMQs often present the highest sensibility [12, 14] but can be over-inclusive [14] and, because the cases found lack specificity, their evaluation is extremely long. Finally, relevant PTs may not be included in SMQs [14], and SMQs are not exhaustive: several serious diagnostics are not yet addressed.

In order to ease and systematize the process of creation of groupings of terms, automatic methods can be exploited. One existing work proposes hierarchical groupings of ADRs [1], which does not necessarily respect the medical reasoning. For instance, in relation to renal diseases, in addition to terms such as Acute nephritis and Insufficiency renal, which have a hierarchical relation among them, it can be also relevant to consider terms relative to laboratory results or to medical procedures. Semantic distance may lead to creation of groupings which respect better the medical reasoning: it has previously been applied to subsets of terms from MedDRA [3] and WHO-ART [9]. In the WHO-ART related work [9], the obtained groupings demonstrated several types of relations: synonyms, antonyms, physiological functions or abnormalities, associated symptoms, abnormal laboratory tests, pathologies and their causes, close anatomical localizations, degrees of severity, and heterogeneous groupings. No evaluation has been realized against existing groupings (SMQs, SOCs, HLTs). We propose to continue the adaptation of the semantic distance approaches to the creation of ADR term groupings. The whole set of MedDRA terms is used, and a special attention is paid to the evaluation of the obtained groupings against the SMQs, which become the reference in the pharmacovigilance area.

2. Material and Method

Material. Material used is issued from MedDRA [4]: ontoEIM and SMQs. The ADR ontology ontoEIM [1] has been created thanks to the projection of WHO-ART and MedDRA on Snomed CT [18], through the exploitation of UMLS [13]. Up to 85% of WHO-ART terms, but only 46% of MedDRA terms are thus aligned. We exploit the resource built with MedDRA, which terms are used for the creation of SMQs. Thanks to this alignment, representation of MedDRA terms is enriched: their structure is improved and becomes parallel to the structuring in Snomed CT, and they receive formal definitions (terms can be defined on four Snomed CT axes: morphology, topography, causality and expression). Our second material, SMQs (Standartized MedDRA Queries), are groupings of MedDRA terms related to a given diagnosis (i.e., Acute renal failure, Hepatic disorders). SMQs are created in order to provide assistance when searching cases relevant to this diagnosis. 82 existing SMQs contain MedDRA PTs and LLTs. In our experience, ontoEIM is the material we use for the creation of groupings of ADR terms, while SMQs in their broad version are used as gold standard for the evaluation of these groupings.

Method. Semantic distance between terms is often computed within terminologies or ontologies. It depends on the number of edges (or the shortest path) between two terms.
(four edges between terms Abdominal abscess and Pharyngeal abscess on figure 1), although other factors may be also exploited. A detailed description of our method is proposed in [6], while here we remind its main steps and new aspects. First, because of the partial alignment of the MedDRA terms (46%), we try to improve it by the transposition of alignments between LLT and PT terms when one of them is aligned and others are missing. Four different cases where distinguished and successfully processed [6]. Then, semantic distance [16] is computed between MedDRA PT and LLT terms through the ontoEIM resource. We exploit for this either the ADR terms only (which belong mainly to the axis Diagnostic D) or ADR terms and their formal definitions. Within the formal definitions, we exploit elements provided by two axes: morphology M (the kind of the abnormality) and topography T (the anatomical localization). These axes are often involved in the definition of diagnostics [17] and they are also the most frequently present in ontoEIM. Let’s consider figure 1 and its two terms Abdominal abscess and Pharyngeal abscess defined as follows:

– Abdominal abscess: M = Abscess morphology, T = Abdominal cavity structure
– Pharyngeal abscess: M = Abscess morphology, T = Neck structure

The shortest paths sp are computed between these two terms (axis D) and between their formal definitions (axes T and M). Weight of edges is set to 1, and the value of each shortest path corresponds to the sum of weights of all its edges. For this pair of terms we obtain the following values: $sp_D = 4$, $sp_T = 10$ and $sp_M = 0$.

Figure 1: The shortest paths $sp$ between Abdominal abscess and Pharyngeal abscess computed on the three axes: Diagnostics (D), formal definitions (T and M).

Semantic distance is computed [6], which allows to generate a semi-matrix and to apply an ascendant hierarchical classification [10] for the creation of groupings of terms. The minimal threshold is set to 2. We perform several experiences, where we exploit: one axis (D) and three axes (D, M, T); all terms in SMQs and only terms aligned with Snomed CT; the best grouping for a given SMQ and merged groupings. Groupings are compared with 9 SMQs [6] issued from [19]. Evaluation is performed with three classical measures: precision $P$ (number of relevant terms grouped as a percentage of the total number of the grouped terms), recall $R$ (number of relevant terms grouped as a percentage of the number of terms in the corresponding SMQ) and F-measure $F$ (the harmonic mean of $P$ and $R$).

3. Results and Discussion

Figure 2 indicates results, mean values for precision, recall and F-measure as well as their
min and max values, obtained with eight experiences performed. Four experiences are done with the best grouping for each SMQ: exploitation of one axis with the complete set of SMQ terms (1a-bc) or with the aligned terms only (1a-ba), exploitation of three axes with the complete set of SMQ terms (3a-bc) or with the aligned terms only (3a-ba). Four more experiences are similar to these but are performed with a merging of groupings (with F-measure higher than 10): exploitation of one axis with the complete set of SMQ terms (1a-mc) or with the aligned terms only (1a-ma), exploitation of three axes with the complete set of SMQ terms (3a-mc) or with the aligned terms only (3a-ma). The first four experiences, performed with the best grouping, show a high precision although recall and F-measure are low [6]. Although unsatisfactory for recall and F-measure, such high precision seems nevertheless to meet the expectations of the pharmacovigilance people looking for highly specific groupings. The four additional experiences, where we apply a merging of n-best groupings for each SMQ, were expected to increase the whole performances. As the graphs show, we can indeed improve the recall with only small deterioration of precision. F-measure is also improved. As a general remark on the obtained results, when we take into account one axis (D) the performances are always higher than when three axes (D, M, T) are exploited, which seems to be due to the incompleteness of the formal definitions. Another factor, which influences the performances of the method, is related to the exploitation of complete SMQs and of SMQs composed of only aligned terms. When the set of terms is reduced to aligned terms it has a positive effect of the evolution of recall and F-measure, because the number of terms to be found is reduced. Although it has a negative effect on precision. Additionally graphs from figure 2 indicate also the min and max values for the three evaluation metrics. It appears very clearly that the interval is very large, which means that there is a great variability of performances according to SMQs, and that probably various strategies should be used for different diagnostics.

4. Conclusion and Perspectives

The proposed method exploits the semantic distance for the creation of groupings of ADR terms. Such groupings, especially when they show a high specificity, may be an useful tool for the signal detection in pharmacovigilance area. The method proposed, once it is optimized and improved, may also be exploited during the creation of new SMQs. In our work, several experiences have been performed and groupings have been evaluated against the SMQs in their broad version. A novel aspect of our work, which consisted in a merging
of n-best groupings, allows to improve recall without a significant deterioration of precision, which is a very positive result. Additional future experiences may lead to an adjustment of thresholds and of variables (edge weights, coefficients of axes) and to the definition of other factors which influence the quality of groupings. We will also provide a detailed analysis of individual SMQs in order to decide if different strategies may lead to better results according to the SMQs. We assume also that methods provided by Natural Language Processing may enrich and improve the groupings.

Acknowledgment. This work was partly supported by funding from the European Community's Seventh Framework Programme (FP7/2007-2013) for the Innovative Medicine Initiative (IMI) under Grant Agreement [1150004]. The research leading to these results was conducted as part of the PROTECT consortium (Pharmacoeconomics, Epidemiology and Research on Outcomes of Therapeutics by a European ConsorTium, www.imi-protect.eu) which is a public-private partnership coordinated by the European Medicines Agency. Authors are thankful to other participants of this task (C. Bousquet, O. Caster, G. Declerck, R. Hill, A. Kluczka, X. Kurz, M. Lerch, N. Noren, V. Pinkston, E. Sadou, J. Souvignet, T. Vardar), but views expressed are those of the authors only.

References