

Grouping pharmacovigilance terms with semantic distance

Marie DUPUCH^{ab}, Magnus LERCH^c, Anne JAMET^{bd}, Marie-Christine JAULENT^{ab}, Reinhard FESCHAREK^e, Natalia GRABAR^f

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

Abstract. *Pharmacovigilance is the activity related to the collection, analysis and prevention of adverse drug reactions (ADRs) induced by drugs or biologics. Besides other methods, statistical algorithms are used to detect previously unknown ADRs, and it was noted that groupings of ADR terms can further improve safety signal detection. Standardised MedDRA Queries are developed to assist retrieval and evaluation of MedDRA-coded ADR reports. Dependent on the context of their application, different SMQs show varying degrees of specificity and sensitivity; some appear to be over-inclusive, some might miss relevant terms. Moreover, several important safety topics are not yet fully covered by SMQs. The objective of this work is to propose an automatic method for the creation of groupings of terms. This method is based on the application of the semantic distance between MedDRA terms. Several experiments are performed, showing a promising precision and an acceptable recall.*

Keywords. Natural Language Processing, Medical informatics, Drug safety, Pharmacovigilance, Signal detection, Drug toxicity, 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 coded with terms from dedicated terminologies, e.g., WHO-ART (World Health Organization - Adverse Reaction Terminology) and MedDRA (Medical Dictionary for Regulatory Activities). Safety signal detection – i.e., the detection of previously unexpected potentially causal associations between drugs and ADRs – depends on the quality and specific features of ADR coding. Besides traditional pharmacovigilance methods, statistical algorithms are increasingly utilized to detect signals in large safety databases [1, 2]. To improve signal detection, these methods benefit from groupings of related ADRs [3]. Grouping of terms is especially relevant because the structure of MedDRA is granular and closely related terms can be spread over the whole terminology: the use of very specific terms for coding ADRs may cause a dilution of signals [4]. Thus, various hierarchical levels of MedDRA (PT, HLT, SOC) and manually built SMQs (Standardised MedDRA Queries) have been used for

signal detection. The PT (Preferred Term) level - which is most often used in quantitative signal detection - corresponds mainly to specific ADRs, while HLTs (High Level Terms) and SOCs (System Organ Classes) are hierarchical levels above the PT level. As for the SMQs, their objective is to link terms relevant to a medical condition. SMQs are designed by a group of experts, who start with the scientific definition of the medical condition of interest, followed by *manual* identification of relevant MedDRA terms [5]. This is a labour-intensive task. Evaluation studies of the SMQs have demonstrated that SMQs often present the highest sensitivity [6, 7], but can be over-inclusive [7] and, because the reports found might lack specificity, their evaluation can be time-consuming. Finally, relevant PTs may be missing in SMQs [7], and several serious safety topics are not yet addressed.

In order to ease and systematize the process of creation of term groupings, automatic methods can be applied. One existing work proposes hierarchical groupings of ADRs [8], but this approach does not necessarily respect medical reasoning. For instance in 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 related to laboratory results or medical procedures. Semantic distance may lead to the creation of groupings which respect medical reasoning significantly better: it has previously been applied to subsets of terms from MedDRA [9] and WHO-ART [10]. In the WHO-ART related work [10], 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. However, no evaluation has yet been performed comparing system-generated with existing MedDRA groupings (SMQs, HLTs, SOCs). Semantic distance has also been applied to other biomedical terminologies (Gene Ontology [11], MeSH and SNOMED CT[12], and UMLS [13]), with a manual rating and evaluation of pairs of terms. We propose a better adaptation of the semantic distance approaches for the creation of ADR term groupings. The whole set of MedDRA terms is used, and special attention is paid to compare the obtained groupings with reference SMQs, which have become a widely accepted standard in pharmacovigilance organizations.

2. Material and Method

Material. Material used is issued from MedDRA 13.0 [14]: ontoEIM and SMQs. The ADR ontology ontoEIM [8] has been created thanks to the projection of MedDRA on SNOMED CT (SNCT) [15] through the UMLS [16]. Only 46% of MedDRA terms are aligned. The terminological representation of MedDRA terms is enriched: their structure is improved and becomes parallel to the structuring in SNCT, and terms receive formal definitions (terms can be defined on four SNCT axes: morphology, topography, causality and expression). Our second material, SMQs, are groupings of MedDRA terms related to a given medical condition of interest (*e.g.*, *Acute renal failure*, *Hepatic disorders*). SMQs, consisting of MedDRA PTs and LLTs, are created to assist users in searching ADR reports related to a medical condition. Currently, 84 SMQs have been released. In our experiments, ontoEIM is the source we use to create groupings of ADR terms, while SMQs in their broad version are used as gold standard for the evaluation.

Method. Semantic distance between terms is often computed within terminologies. It depends on the number of edges (or the shortest path) between two terms (e.g., four edges between the terms *Abdominal abscess* and *Pharyngeal abscess* in Fig. 1), although other factors may be taken into account. We present here the main step of the method, relative to the computing of semantic distance [17] between MedDRA PT and LLT terms through the ontoEIM. We use either a) the ADR terms only (which belong mainly to the Clinical disorder axis *D*), or b) ADR terms and their formal definitions. Within the formal definitions, we use elements provided by two axes: morphology *M* (kind of the abnormality) and topography *T* (anatomical localization). These axes are often involved in the definition of ADRs [18] and they are also frequently represented in ontoEIM, as in this example for 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*). The weight of the 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 *sp* values: $sp_D = 4$, $sp_T = 10$ and $sp_M = 0$.

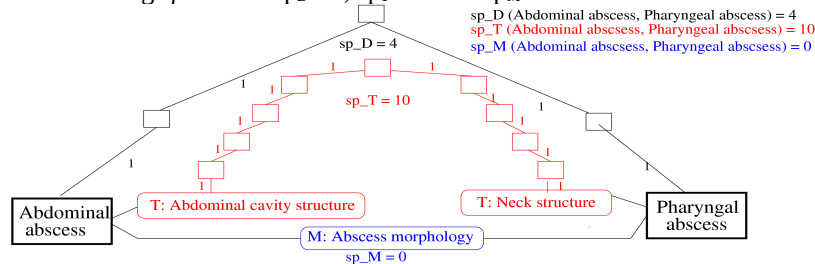


Figure 1: The shortest paths *sp* between *Abdominal abscess* and *Pharyngeal abscess* computed on the three axes: clinical disorder (*D*), topography (*T*) and morphology (*M*).

Semantic distance is computed which allows to generate a semi-matrix and to apply an ascendant hierarchical classification for the creation of groupings of terms. The minimal threshold is set to 2. We perform several experiments in which we evaluate: one axis (*D*) vs three axes (*D*, *M*, *T*); all terms in SMQs vs only terms aligned with SNCT; the best grouping for a given SMQ vs merged groupings. Groupings are compared with 9 SMQs (*Acute renal failure*, *Agranulocytosis*, *Anaphylactic reaction*, *Cytopenia*, *Gastrointestinal haemorrhages*, *Peripheral neuropathy*, *Rhabdomyolysis*, *Severe cutaneous adverse reaction*, *Thrombocytopenia*). Evaluation is performed with three classical measures: precision *P* (number of relevant grouped terms as a percentage of the total number of the grouped terms), recall *R* (number of relevant grouped terms 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 shows our results – mean, min and max values for precision, recall and F-measure - obtained from eight experiments. Four experiments are performed with the best

grouping for each SMQ, i) using one axis with the complete set of SMQ terms ($1a-bc$) or with the aligned terms only ($1a-ba$), ii) using three axes with the complete set of SMQ terms ($3a-bc$) or with the aligned terms only ($3a-ba$). Four more experiments are similar to the above, but are performed with merged groupings. The first four experiments show a promising precision, whereas recall and F-measure are low. Although unsatisfactory for recall and F-measure, such precision nevertheless seems to meet the expectations of pharmacovigilance experts looking for highly specific groupings. The four additional experiments, where we merged the n-best groupings for each SMQ together, were expected to increase the overall performance. As the graphs in Figure 2 show, we can indeed improve the recall with only small deterioration of precision. The F-measure is also improved. When we take into account only one axis (D) the performance is always better than using three axes (D, M, T). This seems to be due to the incompleteness of the available formal definitions. Another factor influencing the performance is related to the use of the complete set of SMQ terms vs the reduced set of only aligned SMQ terms. With the reduced set of terms, we observe a positive effect on the evolution of recall and F-measure (number of terms to be found is reduced), although we observe a negative effect on precision. Additionally, we indicate the min and max values for the three measures. The min-max intervals are visibly very large, which means that there is considerable performance variability for the different SMQs, and that probably various strategies should be used to achieve optimal results for all medical conditions of interest.

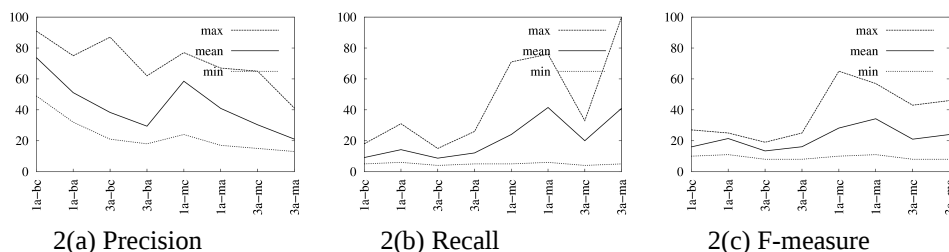


Figure 2: Mean, min and max values for precision, recall and F-measure.

4. Conclusion and Perspectives

The proposed method applies the semantic distance to the creation of groupings of ADR terms. Such groupings, especially when they show a high specificity, may be a useful tool for the detection of signals in pharmacovigilance database. The method, once it is optimized and improved, may also be helpful during the creation of new or improvement of existing SMQs. Furthermore, it could be used to create groupings representing the same medical concept in different terminologies (e.g., MedDRA, WHO-ART, SNOMED CT, UMLS). This in turn, would enable researchers to apply the same term groupings to safety databases independent of the terminology used for ADR coding. In our work, several experiments have been performed to compare system-generated term groupings with SMQs in their broad version. Future experiments will include comparison with the narrow versions of SMQs. A novel aspect of our work, which consists of the merging of n-best

groupings, allows to improve recall without a significant deterioration of precision, which is a positive result. Future studies may lead to an adjustment of thresholds and variables (edge weights, coefficients of axes) and to the identification and 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, taking into account the manifold and diverse medical conditions covered by SMQs. Besides, 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 (Pharmacoepidemiological 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, N. Noren, V. Pinkston, E. Sadou, J. Souvignet, T. Vardar), but views expressed are those of the authors only.

References

- [1] Bate A., Lindquist M., Edwards I., Olsson S., Orre R., Lansner A. & De Freitas R. (1998). A bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol*, **54**(4), 315–21.
- [2] Meyboom R., Lindquist M., Egberts A. & Edwards I.(2002). Signal selection and follow-up in pharmacovigilance. *Drug Saf*, **25**(6), 459–65.
- [3] Hauben M. & Bate A. (2009). Decision support methods for the detection of adverse events in post-marketing data. *Drug Discov Today*, **14**(7-8), 343–57.
- [4] Fescharek R., Kübler J., Elsasser U., Frank M. & Güthlein P. (2004). Medical dictionary for regulatory activities (MedDRA): Data retrieval and presentation. *Int J Pharm Med*, **18**(5), 259–269.
- [5] CIOMS (August 2004). Development and Rational Use of Standardised MedDRA Queries (SMQs): Retrieving Adverse Drug Reactions with MedDRA. Report of the CIOMS Working Group, CIOMS.
- [6] Mozzicato P.(2007). Standardised MedDRA queries: their role in signal detection. *Drug Saf*, **30**(7), 617–9.
- [7] Pearson R, Hauben M, Goldsmith D, Gould A, Madigan D, O'Hara D, Reisinger S, Hochberg A.(2009). Influence of the MedDRA hierarchy on pharmacovigilance data mining results. *Int J Med Inform*, **78**(12), 97–103.
- [8] Alecu I., Bousquet C., Jaulent MC. (2008). A case report: using SNOMED CT for grouping adverse drug reactions terms. *BMC Med Inform Decis Mak*, **8**(S1), 4.
- [9] Bousquet C., Henegar C., Louët A., Degoulet P. & Jaulent M. (2005). Implementation of automated signal generation in pharmacovigilance using a knowledge-based approach. *Int J Med Inform*, **74**(7-8), 563–71.
- [10] Iavindrasana J., Bousquet C., Degoulet P. & Jaulent M.(2006). Clustering WHO-ART terms using semantic distance and machine algorithms. In *AMIA Annu Symp Proc*, p. 369–73.
- [11] Lord PW, Stevens RD, Brass A & Goble CA. (2003). Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation. *Bioinformatics* **19**(10): 1275-1283
- [12] Caviedes JE, Cimino JJ. (2004). Towards the development of a conceptual distance metric for the UMLS. *Journal of Biomedical Informatics* **37**:77-85
- [13] Al-Mubaid H, Nguyen HA. (2009). Measuring semantic similarity between biomedical concepts within multiple ontologies. *Trans. Sys. Man Cyber Part C*, **39**(4):389--398
- [14] Brown E., Wood L. & Wood S. (1999). The medical dictionary for regulatory activities (MedDRA). *Drug Saf.*, **20**(2), 109–17.
- [15] Stearns M., Price C., Spackman K. & Wang A. (2001). SNOMED clinical terms: overview of the development process and project status. In *AMIA*, p. 662–666.
- [16] NLM (2008). UMLS Knowledge Sources Manual. National Library of Medicine, Bethesda, Maryland. www.nlm.nih.gov/research/umls/.
- [17] Rada R., Mili H., Bicknell E. & Blettner M. (1989). Development and application of a metric on semantic nets. *IEEE Transactions on systems, man and cybernetics*, **19**(1), 17–30.
- [18] Spackman K. & Campbell K. (1998). Compositional concept representation using SNOMED: Towards further convergence of clinical terminologies. In *AMIA 1998*, p. 740–744.