User Centred Networked Health Care
A. Moen et al. (Eds.)
IOS Press, 2011

doi:10.3233/978-1-60750-806-9-477

Using Multimodal Mining to Drive Clinical Guidelines Development

Emilie PASCHE[a,b1], Julien GOBEILL[a,c], Douglas TEODORO[a,b], Dina VISHNYAKOVA[a,b], Arnaud GAUDINAT[a,c], Patrick RUCH[a,c], Christian LOVIS[b]
[a] *BiTeM Group, Geneva, Switzerland*
[b] *Division of Medical Information Sciences, University Hospitals of Geneva and University of Geneva, Geneva, Switzerland*
[c] *Information Science Department, University of Applied Sciences, Geneva, Switzerland*

Abstract. We present exploratory investigations of multimodal mining to help designing clinical guidelines for antibiotherapy. Our approach is based on the assumption that combining various sources of data, such as the literature, a clinical datawarehouse, as well as information regarding costs will result in better recommendations. Compared to our baseline recommendation system based on a question-answering engine built on top of PubMed, an improvement of +16% is observed when clinical data (i.e. resistance profiles) are injected into the model. In complement to PubMed, an alternative search strategy is reported, which is significantly improved by the use of the combined multimodal approach. These results suggest that combining literature-based discovery with structured data mining can significantly improve effectiveness of decision-support systems for authors of clinical practice guidelines.

Keywords. Multimodal mining, information retrieval, clinical guidelines, resistance profile, antibiotic cost.

1. Introduction

Since the early use of antibiotics, it was observed that the selection pressure imposed by their massive employ led to a gradual acquisition of bacterial resistance to antibiotics, rendering them ineffective to treat infectious diseases. Thus it became a priority to regulate antibiotic use and clinical guidelines were developed in that intention. Evidence-based approach is being adopted by most of the organizations developing clinical guidelines, since it provides a very rigorous basis by directly linking the recommendation to evidence [1]. However, the systematic review of the literature required by this approach is a time-consuming and labor-intensive process [2].

As part of the DebugIT (Detecting and Eliminating Bacteria UsinG Information Technology) FP7 European project [3], we aim at facilitating clinical guidelines development and maintenance with the creation of an innovative tool called KART (Knowledge Authoring and Refinement Tool), which gathers literature search and information extraction capabilities based on an advanced question-answering framework. In a previous report [4], we presented an approach to help generating

[1] Corresponding author: Emilie Pasche, University Hospitals of Geneva, Division of Medical Information Sciences, Rue Gabrielle-Perret-Gentil 4, 1211 Geneva 14, Switzerland; E-mail: emilie.pasche@unige.ch.

guidelines based exclusively on text-mining. A question-answering engine performed an automatic literature scanning, followed by the identification of hypothetical treatments, thus accelerating systematic reviews. Infectious disease experts can then validate the correct propositions out of the automatically-generated treatments. In this report, we describe how non-textual modalities and in particular clinical data as stored in operational clinical databases can be injected into the baseline system to improve recommendations, using an association model directly inspired by Aronson et al [5]. The structured data used in our experiments gathers clinically-observed resistance profiles, since it is well-known that performing antibiograms before prescription is the optimal way to prescribe an appropriate antibiotic, and prescription cost-related information, assuming healthcare should minimize health costs.

The number of data analysis methods that can be used to combine multimodal contents is virtually infinite since learning algorithms and distance calculi are in general highly data independent. In our experiment we borrowed the methodological framework from the Cranfield paradigm [6] and the linear combination approach pioneered by Fox et al [7]. Numerous subsequent works have been reported to improve the basic method; however the original approach applied strictly to textual observations as for instance when several engines are combined to generate a meta-engine. In contrast, our fusion experiments merge text-generated associations with prior probabilities directly extracted from a clinical datawarehouse.

2. Data and Methods

In this study, clinical guidelines are represented using a simplified design, assuming the following hypothesis: *disease* + *pathogen* + *conditions* = *antibiotics*. A question-answering engine, EAGLi (Engine for Question-Answering in Genomics Literature, http://eagl.unige.ch/EAGLi) [8], is queried with the parameters disease, pathogen and conditions to retrieve a set of the most-cited antibiotics ranked by relevance. The computation of this set is based on the screening of 50 documents from which possible answers are extracted. In our experiments, the target terminology, corresponding to the space of possible answers, consists of 70 antibiotics normalized by their respective WHO-ATC code. A set of synonyms derived from the Medical Subject Headings (MeSH) is used to augment the recall of the answers. Two search engines are used; PubMed, a Boolean and chronological ranking and easyIR, a vector-based similarity ranking.

The set of the most-cited antibiotics is then re-ordered based on the injection of costs and resistance profiles. The re-ranking is based on the attribution of penalty or bonus on the original relevance scores, resulting in a new ranking. Thus, expensive antibiotics and antibiotics with high resistance get lower ranks, while cheapest antibiotics and antibiotics with low resistance obtain a better rank.

The injection of cost is based on a cost list containing 129 products, corresponding to 17 distinct substances, provided by the HUG (University Hospitals of Geneva) pharmacy supply chain. The very same substance can be mentioned several times (Table 1), representing different routes and/or dosages. We attempt to obtain a daily cost for each antibiotic present in the list. Prescription data of the HUG are used to obtain the number of daily doses usually prescribed given a route/dosage for each product. Moreover, as our system is based on the substance and not the marketed product, different products corresponding to the same substance must be aggregated.

This is based on the prescription frequency of each product in the clinical data of the HUG. Finally, for antibiotics without cost information, we attribute an arbitrary cost. This value is set during the tuning phase by varying the bonus/penalty value from 0 (which expresses a minimal price) to 100 (which expresses a maximal price).

The injection of resistance profiles is based on antibiograms present in the HUG Clinical Data Repository [9] of the DebugIT project. As antibiograms for the pair pathogen-antibiotic were retrieved for only 5% of the data, we decided to search antibiograms for the antibiotic only, disregarding the targeted pathogen. From these antibiograms, we extracted the number of resistant (R) and susceptible (S) outcomes. A susceptibility score is calculated for each antibiotic: S/(S+R). When no antibiogram data is available, an arbitrary susceptibility score is assigned. This score is obtained during the tuning phase by varying the bonus/penalty value from 0 (always resistant) to 1 (always susceptible).

Table 1. Extract of the HUG's cost table. Column *Identifier ATC* indicates the ATC identifier of the antibiotic. Column *Term ATC* displays the substance name. Column *Int_Art_Ach* mentions the name of the drug, as well as its form, dosage and number of doses in the box. Column *Public cost* indicates the cost of the article in Swiss Franc (for sake of confidentiality, real prices are not displayed).

Identifier ATC	Term ATC	Int_Art_Ach	Public cost
J01MA02	Ciprofloxacin	Ciproxinep.osusp 5g=100ml (pce)	62.90
J01MA02	Ciprofloxacin	Ciproxinep.osusp 10g=100ml (pce)	104.95
J01MA02	Ciprofloxacin	Ciproxinefiol 400mg=200ml (pce)	50.95
J01MA02	Ciprofloxacin	Ciprofloxcpr 250mg (1x20)	37.50

A collection of 72 rules extracted from the geriatrics guidelines of the HUG is manually translated and normalized to obtain a machine-readable benchmark [10], following the schema of our simplified clinical guidelines: *disease* + *pathogen* + *conditions* = *antibiotics*. The collection is divided into two sets: a tuning set of 23 rules used for the design of the optimal recommendation system and an evaluation set of 49 rules used for the validation of the final system on previously unseen contents.

3. Results

In Table 2, we provide results of our baseline system, i.e. text-mining results as obtained without any additional knowledge show a top-precision of 40.37% when we use the PubMed engine and 34.28% when the easyIR [8] engine is used. Thus, we can compare two different search models. Although PubMed shows a higher precision, it is worth observing that the relative recall is much lower for PubMed. Thus, the PubMed-based search is able to answer 32 questions out of 49, while easyIR is able to provide answers to all questions.

Results obtained when tuning the model with cost-related information are shown in Figure 1A. The best results for easyIR have been found when a null cost is attributed to antibiotics for which no cost information is available. Performances of PubMed-based search decrease with the injection of costs. Final results based on the evaluation set show a top-precision of 43.31% (+9.03%, $p<0.05$) with easyIR and of 40.28% (-0.09%, not statistically significant) with PubMed using the less penalizing settings (Table 2).

Results of the tuning based on the resistance profile are shown in Figure 1B. The best top-precisions are obtained when a value of 0.1 is assigned to antibiotics without resistance profile information for easyIR, meaning that bacteria are most of the time resistant to this set of molecules and a value of 0.0 for PubMed, meaning that bacteria

are always resistant to these antibiotics. Applying these parameters on the evaluation set (Table 2) results in a top-precision of 39.86% (+5.58%, $p<0.05$) for the easyIR search engine, while the PubMed search engine increases its top-precision up to 56.41%, corresponding to a gain of +16.04% ($p<0.01$).

Table 2. Text-only and multimodal results on the evaluation set. Column *Engine* indicates the search model used. Column *Multimodal* is the type of additional knowledge injected in the model. Column *NbA* is the number of answers to which the search model succeeds to answer. Column *P0* is the top-precision, i.e. the precision at recall = 0. Column *MAP* is the mean average precision. Column *R5* is the recall at position 5.

Engine	Multimodal	NbA	P0	MAP	R5
easyIR	No injection	49/49	0.3428	0.1899	0.2619
easyIR	Cost	49/49	0.4331 (+9.03%)	0.2211	0.3197
easyIR	Resistance profile	49/49	0.3986 (+5.58%)	0.2246	0.3299
PubMed	No injection	32/49	0.4037	0.2354	0.3281
PubMed	Cost	32/49	0.4028 (-0.09%)	0.2311	0.2656
PubMed	Resistance profile	32/49	0.5641 (+16.04%)	0.3337	0.4653

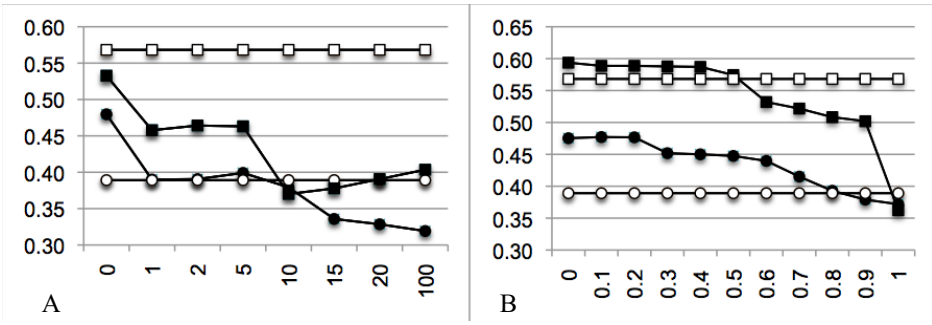

Figure 1. A: Tuning of the cost injection. B: Tuning of the resistance profile injection. Axis x represents the value assigned to the data without injection information and axis y represents the top precision. Baseline top-precisions are represented with ○ for easyIR and □ for PubMed. Multimodal re-ranked top-precisions are represented with ● for easyIR and ■ for PubMed.

4. Conclusion

In our experiments, we have shown that combining textual contents with antibiotics costs result in fairly contrasted results. Indeed, while the relevance-driven model (so-called easyIR) clearly showed improvement when using cost-related features (+9%), such an improvement is not observed when a Boolean search strategy (PubMed) is used. We hypothesized that this is due to the limited set of costs we have had access. This set is even narrower for experiments made with the Boolean search strategy due to the higher number of queries returning no result. Based on the improvements obtained by the relevance-driven engine, we believe that cost information is worth to be used to improve compliance with clinical practice guidelines. In order to overpass such limitations in the future, we plan to use broader coverage resources, for instance, the Swiss Kompendium (http://www.kompendium.ch), a database of drugs, supplied with pricing information. Although those prices may not reflect real HUG's prices, it could potentially provide the information that is currently missing in our cost-based model. Further, during the normalization process, we used the cost of one day of treatment.

But this is a reducing view over antibiotic prescription, since it could be more consistent to determine the cost of the treatment until recovery.

Injection of resistance profiles extracted from the clinical datawarehouses into results obtained by text-mining clearly showed an improvement of the top-precision, especially using the PubMed engine (+16%). Thus, resistance profiles are appropriate features to significantly improve effectiveness of our model when compared to evidence-based knowledge and so for decision-support applied to computerized order-entry systems. Although today's results are already encouraging, we expect that further improvements can be obtained by aggregating species and antibiogram-related information at higher taxonomic levels. Indeed, using exact organisms names such as *Staphylococcus aureus* or *Staphylococcus epidermidis* for retrieving antibiograms can sometimes results in very sparse data, while aggregating all *Staphylococcus* into a phylogenetically-related set seems an effective way to augment recall without affecting significantly precision.

Acknowledgements. The DebugIT project (http://www.debugit.eu) is receiving funding from the European Community's Seventh Framework Programme under grant agreement n°FP7-217139, which is gratefully acknowledged. The information in this document reflects solely the views of the authors and no guarantee or warranty is given that it is fit for any particular purpose. The European Commission, Directorate General Information Society and Media, Brussels, is not liable for any use that may be made of the information contained therein. The EAGLi question-answering framework has been developed thanks to the SNF Grant # 325230-120758.

References

[1] Burgers JS, Grol R, Klazinga NS, Mäkelä M, Zaat J. AGREE Collaboration. Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs. *Int J Qual Health Care*, 2003 Feb;15(1):31-45.

[2] Browman GP. Development and aftercare of clinical guidelines: the balance between rigor and pragmatism. *JAMA. 2001* Sep 26;286(12):1509-11.

[3] Lovis C, Colaert D, Stroetmann VN. DebugIT for patient safety – improving the treatment with antibiotics through multimedia data mining of heterogeneous clinical data. *Stud Health Technol Inform*. 2008;136:641-6.

[4] Pasche E, Teodoro D, Gobeill J, Ruch P, Lovis C. QA-driven guidelines generation for bacteriotherapy. *AMIA Annu Symp Proc. 2009* Nov 14;2009:509-13.

[5] Aronson A, Demner-Fushman D, Humphrey S, et al. Fusion of knowledge-intensive and statistical approaches for retrieving and annotating textual genomics documents. *In Proceedings of the 15th Text REtrieval conference TREC-15*. 2005.

[6] Voorhees E. The Philosophy of Information Retrieval Evaluation. *Lecture Notes in Computer Science*. 2002;2406/2002:143-170.

[7] Fox EA, Koushik MP, Shaw J, Modlin R, Rao D. Combining evidence from multiple searches. *In Proceedings of the First Text REtrieval Conference (TREC-1)*. 2001.

[8] Gobeill J, Pasche E, Teodoro D, Veuthey AL, Lovis C, Ruch P. Question answering for biology and medicine. *Information Technology and Application in Biomedicine (ITAB 2009)*. 2009.

[9] Teodoro D, Choquet R, Pasche E, et al. Biomedical data management: a proposal framework. *Stud Health Technol Inform.* 2009;150:175-9.

[10] Voorhees EM, Harman DK. *TREC experiment and evaluation in information retrieval*. 2005.