Multilingual Assistant for Medical Diagnosing and Drug Prescription Based on Category Ranking

Fernando Ruiz-Rico
University of Alicante
frr@alu.ua.es

Jose-Luis Vicedo
University of Alicante
vicedo@dlsi.ua.es

María-Consuelo Rubio-Sánchez
University of Alicante
mcrs7@alu.ua.es

Abstract

This paper presents a real-world application for assisting medical diagnosis and drug prescription, which relies on the exclusive use of machine learning techniques. We have automatically processed an extensive biomedical literature to train a categorization algorithm in order to provide it with the capability of matching symptoms to MeSH descriptors. To interact with the classifier, we have developed a multilingual web interface so that professionals in medicine can easily get some help in their decisions about diagnoses (lookfordiagnosis.com) and prescriptions (lookfortherapy.com). We also demonstrate the effectiveness of this approach with a test set containing several hundreds of real clinical histories.

1 Introduction

Text categorization consists of automatically assigning documents to pre-defined classes. It has been extensively applied to many fields and in particular, some efforts have been focused on MEDLINE abstracts classification (Ibushi and Tsujii, 1999). However, as far as we are concerned, it has never been used to assist multilingual medical diagnosing and drug prescription by using the textual information provided by biomedical literature together with patient histories.

Every year, thousands of documents are added to the *National Library of Medicine* and the *National Institutes of Health* databases[1]. Most of them have been manually indexed by assigning each document to one or several entries in a controlled vocabulary called MeSH[2] (Medical Subject Headings). The MeSH tree is a hierarchical structure of medical terms which are used to define the main subjects that a medical article or report is about. Due to the wide use of this terminology, we can find translations into several languages such as Portuguese and Spanish (i.e. DeCS[3] - Health Science Descriptors). This paper focuses on both the diseases sub-tree (from C01 to C23) and drugs sub-tree (from D01 to D20). The first one defines on its own more than 4,000 pathological states, and also offers the chance to search for documented case reports related to each of them. The drugs sub-tree provides the capability of arranging around 8,000 active principles, which can be directly matched to commercial drugs.

Our proposal tries to estimate a ranked list of diagnoses and possible prescriptions from a patient history. To tackle this problem, we have selected an existing categorization algorithm, and we have trained it using the textual information provided by lots of previously reported cases and laboratory findings. This way, a detailed symptomatic description is sufficient to obtain a list of possible diseases and prescriptions, along with an estimation of probabilities and bibliography.

We have not used binary decisions from binary categorization methods, since they might leave some interesting MeSH entries out, which should probably be taken into consideration. Instead, we have chosen a category ranking algorithm to obtain an ordered list of all possible diagnoses and pre-

---

[1] http://www.pubmed.gov
[2] http://www.nlm.nih.gov/mesh
[3] http://decs.bvs.br/I/homepagei.htm

scriptions so that the user can finally decide which of them better suits the clinical history.

In this paper, first of all, we will explain the way we have developed our experiments, including a full description of the sources and methods used to get both training and test data. Secondly, we will provide an example of a patient history and both the expected and provided diagnoses. We will also show the suggested drugs recommended by the algorithm for a common disease. And we will finish by showing and commenting several evaluation results on.

2 Procedures

2.1 Medical Diagnosis

We have extracted the training data from the *PubMed* database[1] by selecting every case reports on diseases written in English including abstract and related to human beings. These documents were extracted by using the "diseases category[MAJR]" query, where [MAJR] stands for "MeSH Major Topic", asking the system for retrieving only documents whose subject is mainly a disease. The query provided us with 483,726 documents[4] leading us to 4,024 classes with at least one training sample each.

With respect to the test set, we have used 400 medical histories from the School of Medicine of the University of Pittsburgh (Department of Pathology[5]). Although, so far the web page contains more than 500 histories[4], not all of them are suitable for our purposes. There are some which do not provide a concrete diagnosis but only a discussion about the case, and some others do not have a direct matching to the MeSH tree. We have used from each document both the title and all the clinical history, including radiological findings, gross and microscopic descriptions, etc. To get the expected output, we extracted the top level MeSH diseases categories corresponding to the diagnoses given on the titles of the "final diagnosis" files (dx.html).

As the ranking algorithm, we have chosen the Sum of Weights (SOW) approach (Ruiz-Rico et al., 2006), that is more suitable than the rest for its efficiency, accuracy and incremental training capacity. Since medical databases are frequently updated and they also grow continuously, we have preferred using a fast and unattended approach

that lets us perform updates easily with no substantial performance degradation after incrementing the number of categories or training samples. The restrictive complexity of other classifiers such as SVM could derivate to an intractable problem, as stated by (Ruch, 2005).

To evaluate how worth our suggestion is, we have measured accuracy through three common ranking performance measures (Ruiz-Rico et al., 2006): Precision at recall = 0 ($P_{r=0}$), mean average precision (*AvgP*) and Precision/Recall break even point (*BEP*). Sometimes, only one diagnosis is valid for a particular patient. In these cases, $P_{r=0}$ let us quantify the mistaken answers, since it indicates the proportion of correct topics given at the top ranked position. To know about the quality of the full ranking list, we use the AvgP, since it goes down the arranged list averaging precision until all possible answers are covered. BEP is the value where precision equals recall, that is, when we consider the maximum number of relevant topics as a threshold. To follow the same procedure as (Joachims, 1998), the performance evaluation has been computed over the top diseases level.

2.2 Drug Prescription

Multilingual drug prescription can be achieved through the international active principles, which are the constituents of drugs on which the characteristic therapeutic action of the substance largely depends. The appropriate nomenclature for the active principles can be found translated to several languages from MeSH, and can lead to the final commercial medicaments in most of the countries around the world.

To train the algorithm for this new purpose, we have launched the following query to the *PubMed* database:

("Plant Families and Groups"[majr] OR "Inorganic Chemicals"[majr] OR "Organic Chemicals"[majr] OR "Heterocyclic Compounds"[majr] OR "Polycyclic Compounds"[majr] OR "Macromolecular Substances"[majr] OR "Hormones, Hormone Substitutes, and Hormone Antagonists"[majr] OR "Enzymes and Coenzymes"[majr] OR "Carbohydrates" OR "Lipids"[majr] OR "Amino Acids, Peptides, and Proteins"[majr] OR "Nucleic Acids, Nucleotides, and Nucleosides"[majr] OR "Complex Mixtures"[majr]) AND "therapeutic use"[sh] NOT ("adverse effects"[sh] OR "contraindications"[sh] OR "poisoning"[sh] OR "radiation effects"[sh] OR "toxicity"[sh])

After filtering only articles written in English which have abstract, a total amount of 540,235[4] training documents are left.

---

[4]Data obtained on February 14th 2007

[5]http://path.upmc.edu/cases

Order	Diseases	Probability	Expand
1	Nervous System Diseases	100	+
2	Neoplasms	93	+
3	Pathological Conditions, Signs and Symptoms	61	+
4	Congenital, Hereditary, and Neonatal Diseases and Abnormalities	25	+
5	Endocrine System Diseases	23	+
...	...	...	...

Figure 1: Example of the first level of a hierarchical diagnosis

2.3 Multilingual Environment

Since all training data is written in English, every symptom provided to the algorithm must also be written in English. For this purpose, an automatic translation tool is used for input data in other languages than English. We also promote the translation by using the MeSH vocabulary, which has been delivered by human experts, and provides a reliable correspondence of thousands of non phrases in many language pairs. Although the automatic translation method is not accurate enough for natural speaking, it may be capable of giving quite good results for independent noun phrases (Ruiz-Rico et al., 2006), which are the pieces of information the ranking algorithm uses.

2.4 Availability and Requirements

No special hardware nor software is necessary to interact with the assistant. Just an Internet connection and a standard browser are enough to access on-line through the following sites: *www.lookfordiagnosis.com* and *www.lookfortherapy.com*.

By using a web interface and by presenting results in text format, we allow users to access from many types of portable devices (laptops, PDA's, etc.). Moreover, they will always have available the latest version, with no need of installing specific applications nor software updates.

3 A Couple of Examples

3.1 Medical Diagnosis

One of the 400 histories included in the test set looks as follows:

Case 177 – Headaches, Lethargy and a Sellar/Suprasellar Mass

A 16 year old female presented with two months of progressively worsening headaches, lethargy and visual disturbances. Her past medical history included developmental delay, shunted hydrocephalus, and tethered cord release ...

The final diagnosis expected for this clinical history is: "Rathke's Cleft Cyst", which is a syn-

Nervous System Diseases
Nervous System Neoplasms
Central Nervous System Neoplasms

Order	Diseases	Probability	Expand
1	Brain Neoplasms	100	+
2	Central Nervous System Cysts	44	+
3	Meningeal Neoplasms	13	+
4	Spinal Cord Neoplasms	3	+

Order	Diseases	Probability
1	Craniopharyngioma	100
2	Cysts	63
3	Pituitary Neoplasms	59
4	Central Nervous System Cysts	48
5	Headache	46
6	Xanthomatosis	44
...	...	...

Figure 2: Output example after manual expansion of high ranked topics (up) and by selecting the flat diagnosis mode (down)

rheumatoid arthritis

Order	Active principles	Rating	Drugs	Expand
1	Gold	100	+	
2	Gold Sodium Thiomalate	72	+	
3	Immunoglobulin G	62	+	->
4	Antibodies, Monoclonal	47	+	
5	Hydroxychloroquine	46	+	
6	Aurothioglucose	45	+	->
7	Isoxazoles	43	+	->
8	Auranofin	36	+	
9	Indomethacin	28	+	->
10	Biological Products	23	+	->
11	Antibodies	19	+	->
12	Glucocorticoids	18	+	
13	Immunoconjugates	18	+	->

Myocrisin

Prescription only

Main Use	Active Ingredient	Manufacturer
Rheumatoid arthritis	Sodium Aurothiomalate.	JHC

How does it work?

Sodium aurothiomalate is an injectable form of the precious metal gold. It is often simply called "Gold".

Whilst it is not known exactly how this medicine works it is thought to damp down some elements of the body's immune response, which is indirectly responsible for the swelling, pain and inflammation of the disease.

Sodium aurothiomalate belongs to a mixed group of medicines called "disease-modifying anti-rheumatic drugs" (DMARDs). Unlike simple painkillers, sodium aurothiomalate actually slows progression of the underlying disease. It does, however, produce more serious side effects.

It is used in the treatment active, progressive rheumatoid arthritis and progressive juvenile chronic arthritis.

What is it used for?

- Inflammatory disease of the joints (rheumatoid arthritis)

Figure 3: Example of the drug prescription suggestions for *rheumatoid arthritis* (up) and the final medicament (down) found through the *drugs* link provided by the assistant.

onym of the preferred term "Central Nervous System Cysts". Translating this into one or several of the 23 top MeSH diseases categories we are led to the following entries: "Neoplasms", "Nervous System Diseases" and "Congenital, Hereditary, and Neonatal Diseases and Abnormalities".

In hierarchical mode, our approach provides automatically a first categorization level with expanding possibilities as shown in figure 1. We provide navigation capabilities to allow the user to go down the tree by selecting different branches, depending on the given probabilities and his/her own criteria. Moreover, a flat diagnosis mode can be activated to directly obtain a ranked list of all diseases, as shown on the lower part of figure 2.

After an individual evaluation of this case, we have obtained the following values: $P_{r=0} = 1$, $AvgP = 0.92$, and $BEP = 0.67$, since the right topics in figure 1 are given at positions 1, 2 and 4.

3.2 Drug Prescription

As an example for drug prescription, figure 3 shows the suggestions that the ranking algorithm provides for *rheumatoid arthritis*, where the user obtains a ranked list of active principles. Finally, we reach the name of one of the possible medicaments containing the selected active principle, along with particular recommendations from pharmacists (secondary effects, etc).

4 Results

Last row in table 1 shows the performance measures calculated for each medical history and its diagnosis, averaged afterwards across all the 400 decisions. $P_{r=0}$ indicates that we get 69% of the histories correctly diagnosed with the top ranked MeSH entry. AvgP value means that the rest of the list also contains quite valid topics, since it reaches a value of 73%.

First row in table 1 provides a comparison between SVM (Joachims, 1998) and sum of weights (Ruiz-Rico et al., 2006) algorithms using the well known OHSUMED evaluation benchmark. Even using a training and test set containing different document types, BEP indicates that the performance is not far away from that achieved in text classification tasks, meaning that category ranking can also be effectively applied to our scenario.

Regarding drug prescription tests, we are still working under the evaluation process, colaborating with companies such as *CMPMedica*, which

Table 1: Averaged performance for both text categorization and diagnosis

Corpus	Algor.	$P_{r=0}$	AvgP	BEP
OHSUMED	SVM	-	-	0.66
	SOW	-	-	0.71
Case reports and patient histories	SOW	0.69	0.73	0.62

is in charge of many sites containing drug compendiums (vademecum.es, vidal.fr, cddata.co.uk, etc.). We have already performed preliminary tests by using the symptoms and diseases in the MeSH tree as the input data, and an arranged list of active principles as the output data. We have reached an *AvgP* around 0.9.

5 Conclusions and Further Work

We believe that category ranking algorithms may help in multilingual medical diagnosing and drug prescription from clinical histories. Although the output of the categorization process should not be directly taken as a medical advice, the accuracy achieved could be good enough to assist human experts. However, due to the large amount of new articles continuously added to biomedical literature, it becomes quite difficult for a practitioner to keep up to date. Further works are focused on providing bibliographic references for each suggestion of the classifier. We pretend to select from the PubMed database those entries most related to the pathological states entered by the user.

References

Ibushi, Katsutoshi, Collier-Nigel and Jun'ichi Tsujii. 1999. Classification of medline abstracts. *Genome Informatics, volume 10*, pages 290–291.

Joachims, Thorsten. 1998. Text categorization with support vector machines: learning with many relevant features. In *Proceedings of ECML-98, 10th European Conference on Machine Learning*, pages 137–142.

Ruch, Patrick. 2005. Automatic assignment of biomedical categories: toward a generic approach. *Bioinformatics, volume 22 no. 6 2006*, pages 658–664.

Ruiz-Rico, Fernando, Jose Luis Vicedo, and María-Consuelo Rubio-Sánchez. 2006. Newpar: an automatic feature selection and weighting schema for category ranking. In *Proceedings of DocEng-06, 6th ACM symposium on Document engineering*, pages 128–137.