Helping Physicians to Organize Guidelines within Conceptual Hierarchies

Diego Sona[1], Paolo Avesani[1], and Robert Moskovitch[2]

[1] ITC/irst - Trento, Italy
[2] SafeCrop Centre, IASMA, via Mach 1, S. Michele all'Adige, TN, 38010, Italy

Abstract. Clinical Practice Guidelines (CPGs) are increasingly common in clinical medicine for prescribing a set of rules that a physician should follow. Recent interest is in accurate retrieval of CPGs at the point of care. Examples are the CPGs digital libraries National Guideline Clearinghouse (NGC) or Vaidurya, which are organized along predefined concept hierarchies. In this case, both browsing and concept-based search can be applied. However, mandatory step in enabling both ways to CPGs retrieval is manual classification of CPGs along the concepts hierarchy, which is extremely time consuming. Supervised learning approaches are usually not satisfying, since commonly too few or no CPGs are provided as training set for each class.
In this paper we apply *TaxSOM* for multiple classification. *TaxSOM* is an unsupervised model that supports the physician in the classification of CPGs along the concepts hierarchy, even when no labeled examples are available. This model exploits lexical and topological information on the hierarchy to elaborate a classification hypothesis for any given CPG. We argue that such a kind of unsupervised classification can support a physician to classify CPGs by recommending the most probable classes. An experimental evaluation on various concept hierarchies with hundreds of CPGs and categories provides the empirical evidence of the proposed technique.

1 Introduction

Clinical practice guidelines (CPGs) are an increasingly common and important format in clinical medicine for prescribing a set of rules and policies that a physician should follow. It would be best if automated support could be offered to guideline-based care at the point of care. It is often important to be able to quickly retrieve a set of guidelines most appropriate for a particular patient or task. Correctly classifying the guidelines, along as many semantic categories as relevant (e.g., therapy modes, disorder types, sighs and symptoms), supports easier and more accurate retrieval of the relevant guidelines using concept based search.

Traditional text retrieval systems use the *vector-space-model*, in which retrieval is based on the terms appearance in the text. Concept based search extends this approach, in which documents and queries are mapped into concepts taxonomy based on their contents. Such an approach is implemented in

Vaidurya - a concept based and context sensitive search engine for clinical guidelines [1]. Electronic CPGs repository, such as the National Guideline Clearinghouse (NGC) provide a hierarchical access to electronic CPGs in a free-text or semi-structured format (see <http://www.ngc.org>). However, to implement an accurate concept based search, manual classification of CPGs should be applied by an expert. This task is extremely time consuming, therefore, an automatic process that classifies CPGs along the concepts hierarchy is crucial, while very challenging.

The main aim of this paper is to provide a tool that assists the domain expert (physician), who classifies the CPGs, suggesting the most probable classes for each CPG, even when concept hierarchies is built from scratch, and no examples of labeled CPGs are provided for each class. In this case there is no premise for a successful training of any existing supervised classifier, therefore, recommendations can be given only using an unsupervised model, a task known as the *bootstrapping* problem [2]. Then, once the physician is provided with the set of recommended classes for each CPG she can select the most appropriate ones.

The interesting part of this approach is that while the physician manufactures the concept hierarchy she also inserts some prior knowledge on the desired organization of data. Thus, each concept is described by some keywords and has several descendants sub-concepts, which are more specific. This prior knowledge is exploited by the proposed model in order to perform a preliminary classification of CPGs according to their contents and the desired organization within the hierarchy.

An evaluation of the proposed approach was made on the NGC dataset with encouraging results. Section 2 gives a description of the addressed task with some references to related works. Section 3 introduces the model used to test the proposed solution. Section 4 describes the experimental setup and gives some results. Finally, Section 5 draw some conclusions.

2 Task Definition

A concept hierarchy (taxonomy) is a hierarchy of categories (classes), represented in a *tree-like* structure. Each node is described in terms of both linguistic keywords, ideally denoting their "semantic meaning", and relationships with other categories. The leaves of the tree represent the most specific concepts while nodes near the root represent more general concepts. In our particular task, each node of the hierarchy, also intermediate, can contain CPGs and, in general, each CPG can belong to more than one category (multiple classification).

A typical task in text categorization is to identify a set of categories that best describe the content of a new unclassified CPG. A wide range of statistical and machine learning techniques have been applied to text categorization [3–9]. However, these techniques are all based on having some initial labeled examples, which are used to train a (semi)-supervised model.

These problems are partially solved within the setup we use in the *TaxSOM* model [10]. The model uses the prior knowledge to drive a clustering process and, as a result, it organizes the CPGs on a given concept hierarchy without any supervised training stage. Basically, the model bootstraps the given taxonomy with a preliminary classification of CPGs that afterward need to be reviewed by the taxonomy editor. The basic idea of the *bootstrapping* process is to support and alleviate the manual labeling of a set of unlabeled examples, providing the user with an automatically determined preliminary hypothesis of classification. Specifically, the goal here is to recommend the user with a list of the most probable k classes for each CPG.

3 Classification Models

A simple strategy to classify documents only using prior knowledge is to build a reference vector for each category, through the encoding of all keywords in the current node and in all its ancestors, i.e., all keywords of the nodes in the path from the root to the current node. In this way the codebooks contain both the lexical and the topological information. The documents are then associated to the category having the nearest reference vector. This is a standard prototype based minimum error classifier that in the following will be referred to as *baseline* approach.

The above idea has been further extended in the *TaxSOM* model [10]. Specifically, a *TaxSOM* is a collection of computational units connected so as to form a graph having the shape isomorphic to the given taxonomy. Such computational units, namely codebooks, are initialized only using local lexical information. Then an unsupervised training algorithm (similar to Self Organizing Maps training) adapts these codebooks considering both the documents similarity and the constraints determined by the node keywords and the relationships between units.

Once a *TaxSOM* has been properly trained the final configuration of the codebooks describes a clustered organization of documents that tailors the desired relationships between concepts. The learning procedure of a *TaxSOM* is designed as an iterative process, which can be divided into two main steps: a competitive step and a cooperative step. During the *competitive* step the codebook most similar to the current input vector (a document) is chosen as the *winner* unit. In the *cooperative* stage all codebooks are moved closer to the input vector, with a learning rate proportional to the inverse of their topological distance from the winner unit. The iterations of the two steps are interleaved with an additional phase where the codebooks are constrained by the prior lexical knowledge localized on the nodes (refers to [10] for a detailed description).

4 Experimental Evaluation

We used the NGC CPGs collection to evaluate the suggested approach. The CPGs in the NGC repository are classified along two hierarchical concept trees (*Disorders* and *Therapies*) having each roughly 1,000 unique concepts, where in some regions the concepts trees are 10 levels deep, but the mean is 4 to 6 levels.

There are 1136 CPGs, each CPG may have multiple classifications at different nodes. CPGs have a mean of 10 classifications, not necessarily on the leaves. To evaluate the model with a plurality of taxonomies, we have split the two original trees into eight smaller and different datasets. The variability of the hierarchies structure (number of concepts and CPGs) enabled us to evaluate the model without any biases caused by any prior knowledge of the taxonomy.

The depth of the eight trees ranges from 5 to 9, where the number of concepts ranges from 124 to 326, and the mean ratio of documents per concept is 2.6. The content of documents and the category keywords were cleaned removing stop–words, stemming, and reducing the resulting vocabulary to 500 important words plus the keywords. This selection was done separately for each taxonomy using the notion of Shannon Entropy[3]. Finally, CPG contents were encoded with a *set–of–words* representation (i.e., binary vectors).

As previously outlined, since to our knowledge there are not models devised to solve the proposed bootstrapping problem, we compared *TaxSOM* with the *baseline* approach. The model was tested on each taxonomy performing an hypothesis of classification for all CPGs, and the results were then compared with the original labeling. Actually, the addressed task requires the multi-classification of CPGs, therefore, given a CPG, both models generate a membership value for each class. These membership values are then used to rank the classes, and this ranking is then used to select the best classes to recommend to the user.

For the evaluation we devised a specific measure – the *multi-classification k-coverage precision*. This measure allows a comparison of models rather than an objective evaluation. The measure counts the percentage of CPGs "correctly" classified with respect to the total number of CPGs. The definition of *k-coverage* is strongly related to the definition of "precise classification". A document is correctly classified when all the first k recommended classes include the real classes. Thus, a model having *k-coverage* equal to 60% for $k = 10$ means that for 60% of the documents all the corresponding correct classes appear in the first ten ranked classes.

The results in Table 1 are shown for two settings of k. Based on the results it seems that for all taxonomies *TaxSOM* outperforms the *baseline* approach. The table also provides the results of the same type of analysis done for the original two NGC hierarchies, indicating the same magnitude but with lower performance caused by the lower rate of documents per concepts.

5 Conclusions

In the paper we presented an approach for helping physicians to organize CPGs within hierarchies of concepts. The challenge was twofold: to avoid the need of labeled documents in advance and to exploit relational knowledge encoded by a taxonomy. Preliminary evaluation on a collection of CPGs is encouraging. We are in the process of evaluating the system based on other settings of k.

[3] Entropy is a standard information theoretic approach used to evaluate the amount of information provided by the presence of a word in the dataset.

Table 1. Results of *baseline* and *TaxSOM* *k-coverage* with two different values for k.

	k-coverage			
	baseline		***TaxSOM***	
taxonomies	$k = 10$	$k = 20$	$k = 10$	$k = 20$
diagnosis	11.3	23.3	32.9	46.8
neoplasms	38.2	46.2	47.2	63.7
organic chemicals	60.7	71.0	73.1	80.0
pathol. sympt. cond. signs	42.8	55.4	64.2	76.5
surgical operative proced.	44.1	70.4	68.8	76.9
system diseases nervous	26.1	38.9	53.5	71.2
therapeutics	23.9	39.2	52.5	69.0
virus diseases	27.5	48.9	58.0	72.5
disease condition	8.0	14.0	21.6	34.7
treatment intervention	3.9	5.9	11.2	20.0

6 Acknowledgements

This research was supported partially by SafeCrop Centre, funded by Fondo per la Ricerca, Autonomous Province of Trento.

References

1. Moskovitch, R., Hessing, A., Shahar, Y.: Vaidurya - a concept-based, context-sensitive search engine for clinical guidelines. In: Proc. of the joint conf. of AMIA04 and Medinfo-2004, San Francisco, CA, US (2004)
2. Adami, G., Avesani, P., Sona, D.: Bootstrapping for hierarchical document classification. In: Proc. of CIKM-03, 12th ACM Int. Conf. on Information and Knowledge Management, ACM Press, New York, US (2003) 295–302
3. Ceci, M., Malerba, D.: Hierarchical classification of html documents with WebClassII. In: Proc. of the 25th European Conf. on Information Retrieval (ECIR'03). Volume 2633 of Lecture Notes in Computer Science. (2003) 57–72
4. Cheng, C., Tang, J., Fu, A., King, I.: Hierarchical classification of documents with error control. In: PAKDD 2001 - Proc. of 5th Pacific-Asia Conf. on Knowledge Discovery and Data Mining. Volume 2035 of Lecture Notes in Computer Science. (2001) 433–443
5. Doan, H., Domingos, P., Halevy, A.: Learning to match the schemas of data sources: A multistrategy approach. Machine Learning **50** (2003) 279–301
6. Dumais, S., Chen, H.: Hierarchical classification of web document. In: Proc. of the 23rd ACM Int. Conf. on Research and Development in Information Retrieval (SIGIR'00). (2000)
7. Moskovitch, R., Cohen-Kashi, S., Dror, U., amd A. Maimon, I.L., Shahar, Y.: Multiple hierarchical classification of free-text clinical guidelines. In: Intelligent Data Analysis In Medicine And Pharmacology (IDAMAP-04) Stanford University CA. (2004)
8. Ruiz, M., Srinivasan, P.: Hierarchical text categorization using neural networks. Information Retrieval **5** (2002) 87–118

9. Sun, A., Lim, E.: Hierarchical text classification and evaluation. In Cercone, N., Lin, T., Wu, X., eds.: ICDM 2001 - Proc. of the 2001 IEEE Int. Conf. on Data Mining, IEEE Computer Society (2001) 521–528
10. Adami, G., Avesani, P., Sona, D.: Clustering documents in a web directory. In: Proc. of WIDM-03, 5th ACM Int. Workshop on Web Information and Data Management, ACM Press, New York, US (2003) 66–73