Automatic Processing of Pathological Reports for Classification of Brain Tumors

Yi-Chen Ku [a], Tai-Tong Wong [b], Der-Ming Liou [c], Jen-Hsiang Chuang [d]

[acd] Institute of Health Informatics and Decision Making, National Yang-Ming University, Taipei 112, Taiwan

[b] Neurosurgery Neurological Institute, Taipei Veterans General Hospital Taipei, Taipei 112, Taiwan

Abstract

There are over 120 different types of brain tumors, making effective treatment very complicated. Classification of brain tumors accurately can not only help the doctors to treat the patients correctly but also help doctors to do research and teaching in this field efficiently. The objective of our study was to classify pathological reports into different classes of brain tumors automatically according to World Health Organization 2000 classification of brain tumors. We developed pattern-matching rules called Brain-Tumor Classifier processing pathological reports and classifying brain tumors automatically. We compared Brain-Tumor Classifier against a gold standard that was established by three experts judging 276 records. In this testing set, Brain-Tumor Classifier had a specificity of 99.74% (versus 99.79 ~ 99.9 % for the physicians), a positive predictive value of 91.67% (versus 82.35 ~ 94.92 % for the physicians) while maintaining a reasonable sensitivity of 90.83% (versus 85.91 ~ 97.93 % for the physicians). In addition, it had accuracy of 91.1%. We conclude that automatic processing of pathological reports for classification of brain tumors is feasible and useful.

Keywords: pattern-matching rule, text classification, Natural Language Processing, brain tumor

Introduction

A brain tumor is an abnormal growth of tissue in the brain that can be either malignant or benign. In 1999, the American Cancer Society estimated that 16,800 new cases of intracranial tumors were diagnosed and that 13,100 patients would die of brain tumors in a year[1;2]. There are over 120 different types of brain tumors, making effective treatment very complicated[3]. Brain tumors in children are different from those in adults and are often treated differently[4]. Enhancing quality of life of people with brain tumors requires access to quality specialty care, clinical trials, follow-up care and rehabilitative services. Improving the outlook for adults and children with brain tumors requires research into the causes of and better treatments of brain tumors[4].

In medicine, a neurosurgeon reads utilizes patient medical records daily, focuses on diagnoses and treatments of patients, and classifies different types of patients, which can then be used for research and teaching purposes. Currently, some physicians classify brain tumors according to individual physician's "classification standards" manually. Although this type of classification is very detail and contains individual physician's personal diagnostic experiences, it may not be able to be compared with other international standards.

Tumors of the Central Nervous System (CNS) are a very heterogenous group, including benign as well as highly malignant neoplasm with obvious differences among rates of incidence and prevalence, and mortality in different types[5]. There are many different classification systems for brain tumors, making different classification results [3]. The third World Health Organization (WHO) classification of brain tumors was published in 2000. This classification is based on the consensus recommendation of an international WHO working group of experts that convened in Lyon in July 1999. The editors are P. Kleihues and W.K. Cavanee. The publisher is IARC Press at the International Agency for Research on Cancer (IARC) in Loyn, France, a cancer research institute of the WHO. The 2000 WHO classification is published in the form of the new Blue Book series, which contains, in addition to definitions and codes of the International Classification of Diseases-Oncology, the comprehensive chapters describing the epidemiological, clinical, radiological, histopathological, biological and predictive features of each entity. During the past decade our knowledge of the genetic basis of human neoplasms has increased greatly and histological classification of neoplasms is now increasingly supplemented by genetic profiling [6].

Narrative text reports are a significant source of clinical data because much of the clinical information contained in patient medical records is in narrative form[7]. However, it is difficult for using medical text reports for data analysis and decision support applications directly. Natural Language Processing (NLP) can be used for extracting information from narrative data[8]. The pattern-matching rules[9], which is a method of the text categorization in NLP, is a handcrafted rule based method to recognize key concepts in documents and to assign appropriate categories to them.

The purposes of this study were to develop an automatic method to classify pathological reports into different classes of brain tumours by using the pattern-matching rules and comply with WHO 2000 classification of brain tumors[6].

Method

Pattern-matching Rules

This study used the pattern-matching rules for text categorization. The patterns that we would like to match were the keywords in the pathological reports and these keywords were learned from the training set. In addition, we found many of classification issues after analyzing these pathological reports. For example, in different pathological reports, there were many different keywords with the same meaning. They should be classified to the same class. Therefore, after analyzing these pathological reports and discussing with experts, we have created nine rules. Obviously, it is not enough to only use keyword search because a pathological report includes many text strings and there are some important, related words in the pathological reports, such as “revise” and “neck”, which will influence the performance. We used the pattern-matching rules and implemented the rules by using Structured Query Language (SQL) to classify pathological reports.

Training and Testing Set

Data sources were pathological reports obtained from Pediatric Neurosurgery at a medical center in Northern Taiwan. There were 2,793 pathological reports between December 10, 1971 and September 10, 2003. Of these, 1,122 pathological reports were eligible for our study after excluding cytology records. We randomly divided the 1,122 pathological reports into two parts. A total of 846 pathological reports were placed in the training set, and we used the rest of 276 pathological reports as the testing set. The training set was used for learning the keywords for brain tumor classification and for refining the pattern-matching rules. The testing set was used for evaluating the accuracy of our pattern-matching rules.

Gold Standard

Three physicians judged the pathological reports in the testing set to establish a “gold standard” against the pattern-matching rules. Each physician read two thirds of the 276 pathological reports and classified pathological reports as (a) one of the 147 classes from WHO 2000 classification of brain tumors (b) Brain tumor but not include in WHO brain tumor or (c) Unclassified tumors, Images or Clinical features. To determine inter-rater reliability kappa value, a pair of physician judged 92 same pathological reports. If two physicians judged the same pathological report and did not reach agreement on the results, the classification of the pathological report was determined by the third physician. One of three physicians is an attending physician. The second is a chief resident, and the last is a R2 resident.

Performance Evaluation

In our study, sensitivity, specificity and positive predictive value (PPV) [10] were calculated for each class. In addition, we plotted the Receiver Operating Characteristic (ROC) curve [10], to evaluate the performance of the raters and Brain-Tumor classifier. There are a variety of measures for assessing how well humans agree on these judgments. Probably the best known and most widely used among these is the kappa statistic [11]. In general, the following kappa values indicate the stated amount of agreement:

Poor	<0.4
Fair	0.4~0.6
Good	0.6~0.8
Excellent	>0.8

Figure1. Kappa values indication

Our study used statistics software SAS 8.1 to calculate Kappa.

Results

Reliability of three physicians

In the testing set, we calculated kappa values between any two of the three physicians to assess the inter-rater reliability. The result is as follows:

Table 1 Kappa values of three physicians

	Rater1& 2	**Rater2 & 3**	**Rater1 & 3**
Kappa value	0.9523	0.8727	0.8820

According to kappa values indicated in Figure 1, the reliability of the physicians was excellent.

Comparisons between Brain-Tumor Classifier and raters

Table 2 and Table 3 present performance statistics of Brain-Tumor Classifier and rater 1, rater 2, rater 3, respectively. The accuracy was the probability of correct classification in all reports. Because there are four reports that were classified into more than one classes by brain-tumor classifier, total number was 281. In these 4 cases, there were four classes classified correctly and five incorrectly. As can be seen from the Table 2, the accuracy of Brain-Tumor Classifier reached to 91.1% and the raters were 96.74%, 96.2% and 93.48% respectively. Moreover, we calculated average sensitivity, average specificity and average PPV, as shown in Table 3. The average sensitive of Brain-Tumor Classifier was 0.9083, average specificity was 0.9974 and average PPV was 0.9167. According to the 95 percent confidence intervals (95 % C.I.), there was no difference between Brain-Tumor Classifier and raters in statistics on average sensitive, average specificity and average PPV. The performance of Brain-Tumor Classifier was comparable with the experts. Figure 2 shows the ROC curve for Brain-Tumor Classifier and each rater. From the figure,

Rater	**Accuracy (%)**
Brain-Tumor Classifier (n=281)	91.10
Rater1 (n=186)	96.74
Rater2 (n=186)	96.20
Rater3 (n=186)	93.48

rater 2 had the best performance, and Brain-Tumor Classifier had similar performance to rater 3.

Table 2 Comparisons of accuracy among Brain-Tumor Classifier and raters

Table 3 Comparisons of average sensitivity, average specificity and average PPV among Brain-Tumor Classifier and raters

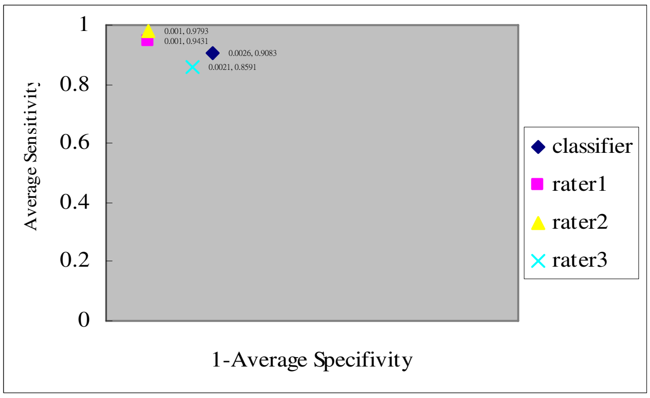

Figure 2 The ROC curve for the Brain-Tumor Classifier and each rater

Discussion

To arrive at the target of classifying pathological reports for brain tumors, we have developed an automatic classification method according to WHO 2000 classification of brain tumors.

The following is the discussions for our study:

Sensitivity of Brain-Tumor Classifier to detect four classes which were lower than or equal to 0.5, and PPV of Brain-Tumor Classifier to detect four classes, which were lower than or equal to 0.5, too. We analyzed the results some of the errors due to misspellings, synonyms of the classes, new tumor names and related words that did not appear in the training set, some other errors may be due to the rules did not appear in the training set. Because of the above-mentioned reasons, 25 records were classified incorrectly.

Although the performance of Brain-Tumor classifier to classify brain tumors based on pathological reports is comparable to some experts, the system performance could be improved further after we analyzed the reasons for incorrect classification.

It took 1.5 hours for classifying pathological reports in the testing set manually. However, it only took 45 seconds for Brain-Tumor Classifier to finish the task.

There are many different classification systems for brain

Rater	Average Sensitivity (95%CI)	Average Specificity (95%CI)	Average PPV (95%CI)
Brain-Tumor Classifier	0.9083 (0.82~0.99)	0.9974 (0.98~1)	0.9167 (0.83~1)
Rater1	0.9431 (0.86~1)	0.999 (0.99~1)	0.9492 (0.87~1)
Rater2	0.9793 (0.93~1)	0.999 (0.99~1)	0.9424 (0.87~1)
Rater3	0.8352 (0.71~0.97)	0.9979 (0.98~1)	0.8235 (0.70~0.95)

tumors; the result for the classification of pathological report is different too. Therefore, using the most suitable standard for classification with public trust is important. We adopted WHO 2000 classification of brain tumors.

Because a pathological report included a lot of keywords and related words, such as "revise" and "neck", only using keyword search was unable to classify pathological report accurately. Our study proposed nine pattern-matching rules, to solve the mentioned problem.

The accuracy of the classification results was influenced by the completeness of the pattern-matching rules. For example, the more complete the synonyms and misspellings is, the more accurate the classification results is.

Because of the human error, the physicians who wrote the pathological reports misspelt the keywords. If we could detect more rules for misspellings, we should be able to improve the performance of Brain-Tumor Classifier.

The 25 records were classified incorrectly in the testing set, because keywords and rules had not appeared in the training set. Therefore the more representative the training set is, the more complete the keywords and rules is.

Through analyzing the reasons for the 25 records which were been classified incorrectly in the testing set, we could increase the new keywords and new rules to improve the sensitive, specificity, PPV and accuracy of

Brain-Tumor Classifier.

Conclusion

Our study proposed nine pattern-matching rules and keywords for classifying brain tumors classes. We have developed an automatic method which complies with WHO 2000 classification of brain tumors to classify pathological reports with good performance

Acknowledgments

We thank Muh-Lii Liang, M.D., Chun-Fu Lin, M.D., and Huai-Che Yang, M.D., at Neurosurgery Neurological Institute, Taipei Veterans General Hospital Taipei, who offered a lot of helps and suggestions in developing rules and keywords of Brain-Tumor Classifier and judged the pathological reports.

Reference

[1]DeAngels LM. Brain tumors. The New England Journal of Medicine 2001; 344(2):1.

[2]Landis SH, Murray T, Bolden S, Wingo PA. Cancer statistics. CA Cancer J Clin 1998;(48):6-29.

[3]Central Brain Tumor Registry of the United States, 1997 Annual Report. 1997.

[4]Brain tumors: brain tumor treatment, symptoms & malignant brain tumor support. http://www.braintumor.org . 2004.

[5]Kepes JJ, Chen WY, Pang LC, Kepes M. Tumors of the central nervous system in Taiwan, Republic of China. Surg Neurol 1984; 22(2):149-156.

[6]yes Kleihues P, Louis DN, Scheithauer BW, Rorke LB, Reifenberger G, Burger PC et al. The WHO classification of tumors of the nervous system. J Neuropathol Exp Neurol 2002; 61(3):215-25.

[7]Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform 2001; 34(5):301-310.

[8]Wilcox A, Hripcsak G. Classification algorithms applied to narrative reports. Proc AMIA Symp 1999;455-9.

[9]Peter Jackson, Isabelle Moulinier. Natural language Processing. Natural Language Processing for Online Applications: Text Retrieval, Extraction & Categorization. 2002: 1-17.

[10]Friedman GD. Epidemiology and Patient Care. Primer of epidemiology. 1994: 268-284.

[11]William R.Hersh. System Evaluation. Information Retrieval: A Health and Biomedical Perspective. 2003: 83-113.