Trial Evaluation of Automatic Lung Cancer Staging from Pathology Reports

Darren Moore[a], Iain McCowan[a], Anthony Nguyen[a], Mary-Jane Courage[b]

[a] *CSIRO e-Health Research Centre, Brisbane, Australia*
[b] *Queensland Cancer Control Analysis Team (QCCAT), Queensland Health, Brisbane, Australia*

Abstract

This paper presents a system that classifies lung cancer stage using automatic text categorisation of free-text pathology reports. The system has been evaluated in a trial where its output was compared to that of two clinical experts for 179 lung cancer cases. The system achieved 77% accuracy for T staging and 87% for N staging. Inter-expert agreement was also studied.

Keywords: Cancer Staging; Lung Cancer; Machine Learning; Clinical Decision Support Systems.

Introduction

Evidence-based treatment guidelines for lung cancer treatment are informed by analysis of patient outcomes, where data is first stratified into comparable cases according to the AJCC TNM (tumour, nodes, metastases) staging standard [1]. The preferred method for staging lung cancer is through multi-disciplinary team (MDT) conferences. In Queensland, the Integrated Lung Cancer Outcomes Project (QILCOP) collects formal stage data from MDTs. However, due to the resource- and time-intensive nature of MDTs, the state-wide coverage of QILCOP stage data is approximately 50-60% of all lung cancer cases.

The purpose of this study was to develop a prototype system to automatically determine a T stage (TX, T1-T4) and N stage (NX, N0-N2) for lung cancer patients from free-text pathology reports stored in clinical information systems. As metastatic lung cancer is defined as involvement of other organs, it is not usually assessable from pathological studies of the lung, and therefore the current system does not attempt to determine the M stage. The system uses automatic text categorisation techniques to detect individual observations within reports that are relevant to staging, and to automatically assign a stage. Such a system could be used to obtain stage data for patients not formally staged by an MDT, allowing more comprehensive population-level analysis of lung cancer outcomes.

This paper reports findings from a trial comparing automatic staging to that of two clinical experts on a set of 179 lung cancer cases.

Method

The system architecture is illustrated in Figure 1. The input to the system is the set of lung cancer related pathology reports for a patient. All report text is first standardised (spelling, acronyms, numbers, removal of punctuation, etc.) and sentence boundaries are identified using a set of search and replace regular expressions. Sequences of words are then transformed into codes from the UMLS Specialist Lexicon using a dynamic programming search for optimal allocation. Phrases implying negation are identified and associated with surrounding terms using the NegEx algorithm [2]. Each transformed report is then classified for relevance to T and N staging by support vector machines. Reports deemed irrelevant are omitted from further processing. If all reports for a patient are classified as irrelevant to T or N staging then the patient is assigned a stage of TX or NX respectively (i.e. stage cannot be assessed). Support vector machines were implemented using the SVM^{light} package [3].

Each sentence of each relevant report is then input to a series of sentence level classifiers corresponding to specific factors from the staging guidelines (e.g. extension of primary tumour into the visceral pleura, involvement of mediastinal lymph nodes, etc). A broad keyphrase filtering step first disregards completely unrelated sentences. Remaining sentences are then passed to the classification step. For factors that were sufficiently well represented in the training set, the classification step is implemented as a two level support vector machine (SVM). The first level binary SVM classifies a bag-of-words representation of each sentence as relevant (or not) to the factor in question. Relevant sentences are then classified as supporting either a positive or negative finding by the second level SVM. For staging factors that were not well represented in the development data set, manually coded rule-based classifiers make a decision based on the proximity of specific sets of words or phrases. The final stage assignment is the highest stage associated with any of the factors classified as positive across all sentences for that patient.

The system was developed on a set of pathology reports for 710 lung cancer patients. Gold standard T and N stages were obtained from a database of pathological TNM stages previously assigned by expert pathologists or multidisciplinary team meetings. Unbiased accuracy results of 77.6% and 81.8% for T and N staging respectively were obtained across the complete development set.

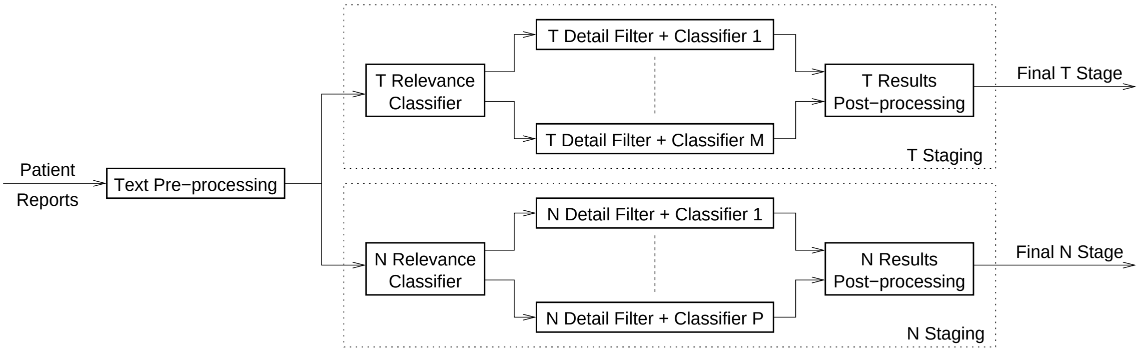
Patient
Reports
Text Pre−processing
T Relevance Classifier
T Detail Filter + Classifier 1
T Detail Filter + Classifier M
T Results Post−processing
Final T Stage
T Staging
N Relevance Classifier
N Detail Filter + Classifier 1
N Detail Filter + Classifier P
N Results Post−processing
Final N Stage
N Staging

Figure 1 - Automatic staging system architecture

The system was then validated in an independent trial, where automatically assigned T and N stages were compared with stages assigned by two pathologists with expert knowledge of the TNM staging guidelines for lung cancer. The trial set consisted of pathology reports for 179 lung cancer patients previously unseen by the development team. The two main objectives of the trial were:

1. To study the level of agreement in expert staging decisions, and use this to establish a consistent gold standard for evaluating the automatic system.

2. To evaluate the reliability of the automatic staging system.

A post-trial meeting was convened with the clinical experts in order to discuss cases where different stages had been assigned. A consensus stage decision was assigned for as many cases as possible. Where a consensus could not be reached, both stages were retained as the gold standard for evaluating the automatic system performance.

Results

The main trial findings were:

1. On the 179 case trial set, the inter-expert agreement was 89.9% (18 disagreements) and 97.8% (4 disagreements) for T and N staging respectively. Most disagreement was due to ambiguity in the reporting, which resulted in experts applying different assumptions and interpretations to reach a final decision. In the post-trial meeting with the clinical experts, a consensus decision was reached on all but 8 T staging decisions.

2. The automatic system was evaluated against the expert-assigned T and N stages. T staging performance was 75.4% and N staging performance was 87.4%, corresponding to the confusion matrices in Table 1. The observed results are similar to those obtained during system development, and the difference in T and N staging performance mirrors the difference in expert agreement levels.

		System			
		T1	T2	T3	T4
Experts	T1	**39**	10	0	1
	T2	5	**81**	2	13
	T3	1	7	**2**	0
	T4	1	4	0	**13**

		System			
		NX	N0	N1	N2
Experts	NX	**10**	6	1	0
	N0	2	**105**	0	1
	N1	0	8	**27**	1
	N2	0	3	2	**13**

Table 1 - Trial set confusion matrices for automatic staging

Conclusion

The prototype system for automatically staging lung cancer was validated against expert decisions in a trial setting with promising results. Future work will focus on improving the current system and adapting the automatic techniques for staging other cancer types and staging protocols.

References

[1] AJCC cancer staging manual. F.L. Greene, D.L. Page, I.D. Fleming, A. Fritz, C.M. Balch, D.G. Haller and M. Morrow (eds.), Springer, 6th edition, 2002.

[2] W.W. Chapman, W. Bridewell, P. Hanbury, G.F. Cooper, and B.G. Buchanan. A simple algorithm for identifying negated findings and diseases in discharge summaries. Journal of Biomedical Informatics, 34:301–310, 2001.

[3] T. Joachims, Making large-scale SVM learning practical. Advances in Kernel Methods - Support Vector Learning, B. Schölkopf, C. Burges and A. Smola (eds.), MIT-Press, 1999.

Darren.Moore@csiro.au,
PO Box 10842, Adelaide St, Brisbane Q 4000, Australia