A property-based model for lung cancer diagnosis

Alma Barranco-Mendoza[1], Deryck R. Persaud[2], Verónica Dahl[3]

Keywords: cancer diagnosis, biomarkers, molecular targets, logic programming, constraint handling rules, concept formation.

1 Introduction.

To this day, lung cancer remains the leading cause of cancer death for both sexes: almost one-third of cancer deaths among men and almost one-quarter among women. [1] Survival rates for lung cancer are low. It is suggested that only about 15% of patients are diagnosed at the early stages. The average 5-year survival rate for patients that are diagnosed early is 48% compared to 15% for those who were diagnosed at the later stages. [2] It is well known that the survival rates can be improved by the early detection of pre-invasive lesions, which are believed to be the possible precursors of malignant tumours. Although new technology is allowing numerous early lesions to be detected, it is becoming clear that only a small percentage of these will actually progress to cancer. Currently a lot of work is being done on the image analysis field, however, at the early development stages of the lesion, the information that can be obtained from lung imaging analysis (X-ray, CAT scans, MRI, PET scans, etc.) is quite limited. To completely understand the evolution of normal epithelium into invasive neoplasia would require the understanding of the genetic relationship of the cells in a pre-invasive neoplastic lesion during the development into invasive cancer. In recent years, biological research has been done in the area of cancer genetics that has shown that cancer results from an accumulation of key mutations in expanding clones originating from tissue-specific stem cells. [3] The recent availability of the human genome sequence, and the development of high throughput genomic technologies and methods for isolating selected cell populations have started to give us the opportunities for understanding how human cancers develop. This information will drastically improve cancer diagnosis and treatment through the discovery of disease-specific molecular targets. As well, recent research has also been focusing on the detection of biomarkers obtained from serum and sputum proteomic analysis [4]. Unfortunately, there is not much work done in terms of computational tools to assist in the analysis of all the genetic and molecular information in addition to the radiological, serum and sputum data that could determine with better accuracy whether an early lesion would progress into cancer or not. As well, for a system to be more valuable it should be able to provide some kind of diagnosis even if given incomplete patient information, as not all tests can or will be done on said patient. Our research intends to address this with the development of a multidisciplinary property-based model for early lung cancer diagnosis.

2 Concept Formation Rules

Concept Formation Rules (CFR) [5] is a directly executable new cognitive model of knowledge construction inspired in constructivist theory as well as in recent natural language processing

[1]School of Computing Science, Simon Fraser University, 8888 University Drive, Burnaby, BC, V5A 1S6, Canada. E-mail: abm@cs.sfu.ca

[2] Infogenetica Bioinformatics, 3197 Tahsis Avenue, Coquitlam, BC, V3B 6E2, Canada E-mail: deryck@infogenetica.com

[3] School of Computing Science, Simon Fraser University, 8888 University Drive, Burnaby, BC, V5A 1S6, Canada. E-mail: veronica@cs.sfu.ca

methodologies. This system accommodates user definition of properties between concepts as well as user commands to relax their enforcement; accepts concept formation from concepts that violate principles that have been declared as relaxable, and produces a list of satisfied properties and a list of violated properties as a side effect of the normal operation of the rules.

We have used CFR to implement our model for lung cancer diagnosis. As part of the input concepts it accepts the patient's age, smoking history, malignancy history, radiological, serum and sputum data. The knowledge store includes the properties that should be evaluated for each input data element as well as the relations amongst them. The diagnosis is given as a probability of cancer that is calculated as a function of the concepts used in the analysis. As well, the diagnosis will list those diagnostic properties that were satisfied and those that were not. For example:

const(Prob),age(x,**A**),history(x,smoker,**T**),serum_data(x,marker_type,in_range)<=>
marker(x,marker_type,in_range,**P**,B),acceptable(marker(x,marker_type,in_range, **P**),B),
probability(**P**,Prob,x, B) acceptable(probability(**P**,Prob,x),B)| possible_lung_cancer(yes,Prob,x).

relax(marker(x,marker_type,in_range,**P**,B)).

This rule evaluates for a patient x if a specific biomarker, *marker_type*, found in serum data is within a certain value range for a patient with an age of **A** who is a type **T** smoker (**T** depends on the number of cigarettes or cigars smoked daily). If true, then the diagnosis of possible lung cancer is going to be true with a probability increase of **P** (where **P** is a function of the patient's age, health history, and this particular biomarker presence). But if we relax the requirement of the presence of the biomarker, then the system can evaluate patient records that do not have this particular information and report in the diagnosis listing that this information was not included in the record, which could be valuable information as recommended follow-up tests for that particular patient.

3 References

[1] Canadian Cancer Society. *Lung cancer stats*, August 2003
http://www.bc.cancer.ca/ccs/internet/standard/0,3182,3278_14459_371459_langId-en,00.html

[5] Dahl, V. 2004. Concept Formation Rules: an executable cognitive model of knowledge construction. In *Proc. First International Workshop on Natural Language Understanding and Cognitive Science*, Porto, Portugal, April 2004.

[4] Gealy, R., Zhang, L., Siegfried, J.M., Lutetich, J.D., and Keohavong, P. 1999. Comparison of mutations in the p53 and K-ras genes in lung carcinomas from smoking and nonsmoking women. *Cancer Epidemiology, Biomarkers, & Prevention* 8:297-302.

[3] Marx, J. 2003. Mutant Stem Cells May Seed Cancer. News Focus. In: *Science* 301:1308-1310.

[2] Varner, L., Norris, C. 2002. Lung Cancer: Battling the Number One Cancer Killer. Medicine on the Horizon. In: *ExploreHealth with Sentara* http://12.42.224.152/healthnews/MedicineontheHorizon/moth042002.htm