



# Enabling the use of hereditary information from pedigree tools in medical knowledge-based systems



Pablo Gay\*, Beatriz López, Albert Plà, Jordi Saperas, Carles Pous

University of Girona, Campus Montilivi, P4 Building, Girona E17071, Spain

## ARTICLE INFO

### Article history:

Received 7 January 2013

Accepted 7 June 2013

Available online 15 June 2013

### Keywords:

Decision support

Structured data

Feature construction

Modeling

Medical applications

Knowledge-based system

## ABSTRACT

The use of family information is a key issue to deal with inheritance illnesses. This kind of information use to come in the form of pedigree files, which contain structured information as tree or graphs, which explains the family relationships. Knowledge-based systems should incorporate the information gathered by pedigree tools to assess medical decision making. In this paper, we propose a method to achieve such a goal, which consists on the definition of new indicators, and methods and rules to compute them from family trees. The method is illustrated with several case studies. We provide information about its implementation and integration on a case-based reasoning tool. The method has been experimentally tested with breast cancer diagnosis data. The results show the feasibility of our methodology.

© 2013 Elsevier Inc. All rights reserved.

## 1. Introduction

Family history has been important for preventing several inheritance diseases, as it is one of the key variables in the Gail model [1] for breast cancer diagnosis. Family information is usually gathered thanks to pedigree software (as Cyrillic [2]), which allows to annotate relationships, healthy states, genetic markers, and much more data on patients and relatives, in a tree structured way. Thus, to improve medical care, a knowledge-based system (KBS) should incorporate the information about families collected thanks to this kind of tools.

Data processing on pedigree software has been traditionally faced from a statistical point of view. However, statistics are not easy in this kind of structured scenarios, most popular statistics tools can conduct to inappropriate or absurd conclusions, and other methods for compositional data are required. On the other hand, expert physicians can evaluate at a glance, from the structure, density and another heuristic knowledge, the risk of a member of the family for suffering an inheritance illness. The skill of evaluating the information of the family is something that is acquired by experience, and difficult to transmit to other, novice physicians. Our research concerns the development of tools that capture the heuristic knowledge of expert physicians, finding out measures from the tree structure that conducts as close as possible

to the predictions made by them. Providing a method to extract the relevant information from family trees enables the integration of pedigree tools with medical KBSs so other physicians can also use inheritance data in their decision making.

The contribution of this paper is our methodology towards achieving such integration. It includes the definition of structured data-based indicators which are computed by analyzing the information contained in pedigree files. The methodology is presented first under the assumption of a simple, hierarchical family, and then is extended to cover more complex situations (second marriages, and so on). Our research is constrained to the data we have on breast cancer, an illness in which inheritance has been proved to be a key factor. Nevertheless, we believe that other inheritance illnesses can benefit from our results.

This paper is organized as follows. First we provide information about the structured data on Section 2. Next, in Section 3, we describe our methodology to evaluate a set of indicators from pedigree files. In Section 5 case studies are provided, and in Section 6 the experimentation performed so far is shown and discussed. Then, in Section 7 we expose some related work and, finally, we end the paper in Section 8 with some conclusions and future work.

## 2. Structured family data

Our starting point is the family information gathered in the very well-known standard that nowadays is one of the most used for pedigree information sharing: the GEDCOM format (GEnealogy Data COMmunication) [3]. This format consists of a header section, records, and a trailer section. Within these sections, records repre-

\* Corresponding author.

E-mail addresses: [pablo.gay@udg.edu](mailto:pablo.gay@udg.edu) (P. Gay), [beatriz.lopez@udg.edu](mailto:beatriz.lopez@udg.edu) (B. López), [albert.pla@udg.edu](mailto:albert.pla@udg.edu) (A. Plà), [jordi.saperas@udg.edu](mailto:jordi.saperas@udg.edu) (J. Saperas), [carles.pous@udg.edu](mailto:carles.pous@udg.edu) (C. Pous).



Download English Version:

<https://daneshyari.com/en/article/10355490>

Download Persian Version:

<https://daneshyari.com/article/10355490>

[Daneshyari.com](https://daneshyari.com)