Chapter 49
A Comparative Study of Associative Classifiers in Mesenchymal Stem Cell Differentiation Analysis

Weiqi Wang
University of Oxford, UK

Yanbo J. Wang
China Minsheng Banking Corporation Ltd., China

Qin Xin
Simula Research Laboratory, Norway

René Bañares-Alcántara
University of Oxford, UK

Frans Coenen
University of Liverpool, UK

Zhanfeng Cui
University of Oxford, UK

ABSTRACT

Discovering how Mesenchymal Stem Cells (MSCs) can be differentiated is an important topic in stem cell therapy and tissue engineering. In a general context, such differentiation analysis can be modeled as a classification problem in data mining. Specifically, this is concerned with the single-label multi-class classification task. Previous studies on this topic suggests the Associative Classification (AC) rather than other alternative (Classification) techniques, and presented classification results based on the CMAR (Classification based on Multiple Association Rules) associative classifier. Other AC algorithms include: CBA (Classification Based on Associations), PRM (Predictive Rule Mining), CPAR (Classification based on Predictive Association Rules) and TFPC (Total From Partial Classification). The main aim of this chapter is to compare the performance of different associative classifiers, in terms of classification accuracy, efficiency, number of rules to be generated, quality of such rules, and the maximum number of attributes in rule-antecedents, with respect to MSC differentiation analysis.

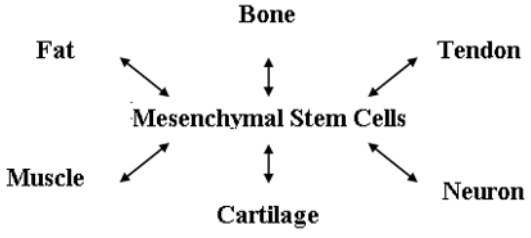
DOI: 10.4018/978-1-4666-2455-9.ch049

INTRODUCTION

Mesenchymal Stem Cells (MSCs) have been claimed to be an integral part of tissue engineering due to their pluripotent differentiation potential both *in vivo* and *in vitro* (Beeres, Atsma, van der Laarse, Pijnappels, van Tuyn, & Fibbe, 2005; Derubeis & Cancedda, 2004; Zhang, Li, Jiang, Wu, & Liu, 2004), and have become one of the most significant research topics in the past few decades. MSCs are able to differentiate along the osteogenic, chondrogenic, adipogenic, myogenic, tendonogenic, and neurogenic lineages under appropriate stimuli (Pittenger, Mackay, Beck, Jaiswal, Douglas, & Mosca, 1999; Roelen & Dijke, 2003; Tuan, Boland, & Tuli, 2003), generating bone, cartilage, fat, muscle, tendon, and neuron cells respectively (Figure 1). Other discoveries on plasticity and immunologic properties of MSCs have further increased the interest in their clinical applications (Krampera, Glennie, Dyson, Scott, Laylor, & Simpson, 2003; Muller, Kordowich, Holzwarth, Spano, Isenee, & Staiber, 2006). The significance of MSCs in clinical therapy has triggered an urgent need for a better understanding and, if possible, computational prediction of MSCs differentiation (Griffith & Swartz, 2006).

In order to obtain a better understanding of MSCs, a significant number of studies have been conducted (Battula, Bareiss, Treml, Conrad, Albert, & Hojak, 2007; Hanada, Dennis, & Caplan, 1997; Lennon, Haynesworth, Young, Dennis, & Caplan, 1995; Magaki, Kurisu, & Okazaki, 2005; Meuleman, Tondreau, Delforge, Dejeneffe, Massy, & Libertalis, 2006; Muller et al., 2006), providing an enormous amount of experimental data for computational prediction. However, those studies and experiments were not interrelated with each other, i.e. different experiments focused on different combinations of factors affecting MSC differentiation, including species of cell donors, *in vitro* vs. *in vivo* environments where the experiments were executed, cell culture media, growth factors and supplements to the culture media, culture dimension (monolayer vs. 3D culture), cell attaching substrate (for monolayer culture) vs. scaffold (for 3D culture), and cell behaviors, especially the differentiation fates of MSCs in terms of the different lineages to which the cells committed (Hanada et al., 1997; Haynesworth, Baber, & Caplan, 1996; Kuznetsov, Friedenstein, & Robey, 1997; Lennon et al., 1995; Muller et al., 2006). The scattered experimental data hence resulted in a large amount of noise in the database and a discrete data structure, which cannot take advantage of traditional mathematical modeling methods. As a consequence, it is extremely difficult to construct intracellular pathway models for MSC metabolism, especially for their differentiation process (Bianco, Riminucci, Gronthos, & Robey, 2001).

Figure 1. Differentiation fates of MSCs



On the other hand, useful information and meaningful prediction for MSC differentiation can be derived based on knowledge discovery via data mining techniques. The nature of data mining is to discover useful, but hidden, information (knowledge) in data. Previous studies under this heading (Wang, Wang, Banares-Alcantara, Coenen, & Cu, in press; Wang, Wang, Banares-Alcantara, Cui, & Coenen, 2009) model the analysis of MSC differentiation as a classification problem (in data mining) — the task of assigning predefined categories (differentiation fates) to "unseen" (MSC) instances. Broadly speaking, classification can be separated into two divisions: *single-label* that assigns exactly one predefined category to each "unseen" instance; and *multi-label* that assigns one or more predefined category to each "unseen"