Computers in Human Behavior 51 (2015) 762–770

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Cloud computing service for knowledge assessment and studies recommendation in crowdsourcing and collaborative learning environments based on social network analysis

Vladimir Stantchev *, Lisardo Prieto-González, Gerrit Tamm

Institute of Information Systems, SRH University Berlin, Berlin, Germany

ARTICLE INFO

Article history:
Available online 17 January 2015

Keywords:
Collaborative learning
Social networks
e-learning
Knowledge representation model
Artificial Immune System
Big data

ABSTRACT

Interactions among people have substantially changed since the emergence of social networks, the expansion of the Internet and the proliferation of connected mobile devices, and so have the possibilities of collaborative learning, with the inclusion of new e-learning platforms. From this point, assessing human knowledge in these virtual environments is not a trivial task. This work presents a novel cloud-computing-based service which relies on advanced artificial intelligence mechanisms to infer knowledge and interest from users considering the aggregated data presented from/to these users in different social networks. This way it is possible to assess with a certain degree of confidence the user knowledge level in different topics as well as recommend additional specific education related to his/her former studies in order to get a better/desired job.

1. Introduction

Technological developments in the last quarter century spurred an unprecedented development of knowledge-oriented occupations, business models and societal paradigms. This is particularly true for the educational sector, where changes in technologies (expansion of computing, portable devices and telecommunication infrastructures) and introduction of Internet based services, such as social networks changed the behavior and interaction among educators and students (Stantchev, Colomo-Palacios, Soto-Acosta, & Misra, 2014).

This work focuses on advanced artificial intelligence-based structures and algorithms to provide a knowledge assessment method as well as a recommendation system. The hypothesis of this research is that such augmentation of the capabilities provided by professional social networks can improve collaborative learning scenarios by matching demand and supply in different knowledge areas. Matching demand and supply in markets for complex goods or services is notoriously hard, due to their inherent information asymmetry (Stantchev & Tamm, 2012) with information substitutes and automated matching presenting agreed upon approaches to alleviate this asymmetry. Both the educational and the employment markets are typical examples – with universities' accreditations, employer rankings, professional references and recommendations, as well as educational degrees and certificates being canonical exhibits of such information substitutes.

The rest of the article is structured as follows: Section 2 covers the State of the Art (SoA) about technologies and systems related to the proposed solution. Section 3 details the proposed model, describing the specific techniques used and the theoretical foundations applied. Section 4 details the adaptation of the proposed model to provide useful services in education and crowdsourcing areas. Section 5 provides an overview of the results obtained from the application of the model, and finally, general conclusions derived from this work and future lines of research are presented in Section 6.

2. State of the Art

In this section, a brief overview about relevant existing technologies will be summarized, since they are important in the implementation of the final system. In first place, the most relevant characteristics of social networks are exposed, as most of the data used to feed the solution is obtained from them. Following, characteristics of Artificial Immune Systems (AIS) as well as the typical modules and models used and taken into consideration in their

* Corresponding author at: SRH Hochschule Berlin (SRH), Institute of Information Systems, Ernst-Reuter-Platz 10, Raum H.804, 10587 Berlin, Germany. Tel.: +49 (0) 30 374 374 370.

E-mail addresses: vladimir.stantchev@srh-hochschule-berlin.de (V. Stantchev), lisardo.gonzalez@srh-hochschule-berlin.de (L. Prieto-González), gerrit.tamm@srh-hochschule-berlin.de (G. Tamm).

http://dx.doi.org/10.1016/j.chb.2014.11.092

design and implementation will be explained. Later, existing knowledge representation models are described, as the skill levels of the different users will be compared against a big set of knowledge areas and relationships among these areas, which will be modelled using a specific novel technique. Finally, basics of recommender systems as well as the profile inference techniques and interoperability between the proposed solution and existing systems will be described.

2.1. Social network sites

As defined by Ellison and Boyd (2007), "social network sites as web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site". A "social network" is a broader term that refers to a group of people, often with emphasis on the cross-disciplinary characteristics of a network (e.g. nodes and ties between them) (Bródka, 2013).

In the last years, the growing of this kind of services has been humongous, as well as their diversity (oriented applications). Nowadays it is possible to find generalistic blog-like social networks (i.e. Facebook, Google +, tumblr.), microblogging networks (Twitter), professional networks (LinkedIn, XING), photography oriented networks (Instagram), video based networks (YouTube, Vine, Instagram video) or personal interests sharing networks (Pinterest). Some users have different profiles in different networks at the same time, taking advantage of the different possibilities.

Social networks (SN) have been part of information systems research since the 1990-ies, back then often known as "online communities" or "virtual communities". A very detailed summary of SN-related research before the explosive growth of Facebook, Twitter and others is presented, together with a life-cycle view on success factors, in Iriberri and Leroy (2009). Some benefits for individual users that this meta-study identified included "Visibility beyond boundaries of local work or geographical community", "Opportunity to build and maintain social ties ...", and "Access to research articles and hyperlinks within the community related to the focus of the community", while benefits for organizations included "Increase access to expert knowledge" and "Increased quality of knowledge and advice". In the proposed life-cycle of *Inception–Creation–Growth–Maturity* the most important success factors in the *Growth*-phase included "Quality Content", "New user Integration" and "Trust". We are convinced that both the expected benefits and the success factors can be positively influenced by the proposed approach – trust can be positively influenced by automated or semi-automated matching in complex electronic marketplaces (Petruch, Stantchev, & Tamm, 2011; Stantchev & Stantcheva, 2012) while the individual perception of quality is strongly influenced by the own educational and cultural background (Maznevski, Gomez, DiStefano, Noorderhaven, & Wu, 2002).

2.2. Artificial Immune Systems

As stated by De Castro and Timmis (2002), "Artificial immune systems can be defined as computational systems inspired by theoretical immunology and observed immune functions, principles and models, which are applied to problem solving". AIS are included in the broad set of bio-inspired algorithms corresponding to bio-informatics area (genetic algorithms, ant colonies, neural networks, etc.) which were introduced in the 80s by Bersin, Farmer, Packard, Perelson and Varela.

AIS base their implementation in two complex systems existing in Natural Immune Systems: innate immunity and acquired immunity. Innate immunity is the capability of elimination a limited set of existent antigens. It is distributed all over the body and it is formed by several types of cellules, such as macrophages and monocytes. On the other hand, Acquired or adaptive immunity allows recognizing and eliminating antigens which had not been faced previously. It also allows remembering this new types of antigens in order to eliminate them easily in the future.

There are some relevant characteristics of the immune systems that make them especially interesting for their use in computation:

- **Decentralization**: It takes place thanks to antibodies' autonomy which can carry out their main function without needing any kind of central control.
- **Diversity**: Mutation allows increasing the amount of different individuals in the system, as well as in other type of genetic algorithms, so it also increases the number of patterns which are recognized.
- **Learning**: It can be defined as the acquisition of a behavior through the experience. The AIS are able to adapt to different antigens and to eliminate then easier and easier.
- **Memory**: "the ability to remember information, experiences, and people". Antibodies which recognized antigens in the past, last in time.
- **Parallelism**: Due to the antibodies' autonomy, pattern recognition, cloning and the rest of actions can be carried out within the system in a parallel way.
- **Pattern recognition**: It is one of the AIS' main characteristics and it allows recognizing the similarity grade between an antigen and an antibody.
- **Self-regulation**: Cellules' time life, as well as their collaboration, makes the system regulate on its own.

AIS present a similar structure, defined in the work of De Castro and Timmis (2002). This structure is divided in three layers that go from the application domain to the final solution. The first layer, representation, is related to the different domain concepts which are going to be faced with the main components present in an information system (antigen representation, antibody representation, definition of knowledge region, storage, etc.). The second layer, affinity grade, referees to the necessary functions to be developed in order to analyze the similarity among two individuals. This is very important, as it is the responsible for the biological system proliferation, as cloning and selecting individuals will be directly related to affinity grade. Some existing techniques to study the affinity are, for example, the Manhattan distance or the Euclidean distance. In order to favor affinity and diversity (required by AIS and other bio-inspired algorithms such as genetic programming), mutation operators are used. Apart from mutators, a selection function proportional to affinity must be defined in order to filter the best candidates. Common used functions are:

- **Bi-class selection**: A percent of best and worst individual are kept in the population whereas the other individuals are chosen randomly.
- **Elitist selection**: It is based on keeping the best individual or a bigger amount of the best tones in the population.
- **Ranking based selection**: It is based in the designation of a reproduction or cloning probability depending on the affinity grade of a given individual (in other words, its quality).
- **Tournaments based selection**: An n individual size set is selected in a random way, and its individuals compete among them to check out which ones have the highest affinity, since they will be kept for the next generation. This process is repeated a certain number of iterations to generate a new population of individuals.