Recommending Relevant Open Source Projects on GitHub using a Collaborative-Filtering Technique

Mohamed Guendouz, GeCoDe Laboratory, Tahar Moulay University of Saida, Saida, Algeria
Abdelmalek Amine, GeCoDe Laboratory, Tahar Moulay University of Saida, Saida, Algeria
Reda Mohamed Hamou, GeCoDe Laboratory, Tahar Moulay University of Saida, Saida, Algeria

ABSTRACT

The GitHub website represents nowadays an essential tool for developers from around the world; it is considered as a social network for them in which they can share their open source projects with others in a form of repositories. This paper presents and discusses the design and the implementation of a new recommender system for GitHub repositories based on a collaborative-filtering approach, which can be useful in many ways in the process of searching for the right solutions to build projects. The GitHub website is becoming very popular these days, a lot of projects are shared by millions of developers, building this recommender system can reduce searching time and make search results more and more relevant. The authors evaluate their system by conducting a set of experiments on a real data set using different well-known metrics and the k-fold cross validation method. Results obtained from these experiments are very promising, the authors found that their recommender system can reaches better precision and recall accuracy.

Keywords: Collaborative-Filtering, GitHub, Open Source, Recommender Systems, Social Networks

INTRODUCTION

GitHub is a very popular crowdfunding software development platform, a social coding platform and a web based Git repository hosting service, allowing anyone to participate in open source project documentation, design, coding and testing in a social way. In order to participate in these activities, a developer must create an account, allowing him to share his own projects, forking other’s projects and following other developers, Figure 1 shows a sample GitHub profile.

DOI: 10.4018/IJOSSP.2015010101
One of the most helpful implemented features on GitHub is the fork feature, which means making a full copy of the repository of the original project. Forking a repository allows the developer to freely experiment with the project without affecting the original copy, forking is considered as the first task to do in order to make contributions to an existing project. Another implemented feature is the Star feature, when a developer gives a star to a repository it means that he is interested in this project. For example, a developer who is interested in mobile game development may give stars to some 2d mobile game libraries like: AndEngine, LibGDX, cocos1d-x and others.

Developers are always searching for good open source projects to make project prototypes or to enhance their own software projects with new features, GitHub provides them a search functionality to do this manually without any automatic recommendations provided, Figure 2 shows a sample search page. However, searching for suitable repositories can be a difficult task and may take a long time, it can also interpret the development process of a project, for that reason the existence of an automatic recommender system for GitHub repositories may be very helpful for developers to reduce search time and make search results more relevant and organized, these are the main benefits of such a system for all developers. However, developers may benefit differently from it according to their profile type and their professional skills, for instance: a professional developer is probably searching for new programming challenges or even for business opportunities, while a beginner is probably looking for good stuff to learn something new or to improve its skills, or he is simply searching for repositories to work on. The issue that arises in these cases is how we can find a relevant content on GitHub and recommend it to a user.

In this paper, the authors present a new system for recommending relevant GitHub repositories for developers; they use a collaborative-filtering approach and they model the user behaviors as a
Open Source Developer Layer Assessment: Open Onion
www.igi-global.com/article/open-source-developer-layer-assessment/170475?camid=4v1a

A Cost Model of Open Source Software Adoption
www.igi-global.com/article/cost-model-open-source-software/38906?camid=4v1a