**KDD Cup 2016 Proposal**

**A Novel Collaborative Filtering Algorithm by Bit Mining Frequent Itemsets**

|  |  |
| --- | --- |
| *Sending with full of respect*: | KDD Cup 2016 Organization Committee |
| *Submitted by*: | Loc Nguyen |
| *Organization*: | Sunflower Soft Company, Ho Chi Minh city, Vietnam |
| *Contact address*: | 1/4B Ton Duc Thang street, My Binh ward, Long Xuyen city, An Giang province 881092, Vietnam |
| *Email*: | [ng\_phloc@yahoo.com](mailto:ng_phloc@yahoo.com) |
| *Phone*: | +84975250362 |

I am Loc Nguyen, founder of Sunflower Soft Company, who submit a proposal of “a novel collaborative filtering algorithm by bit mining frequent itemsets” to KDD Cup 2016. The algorithm is named **GreenFall**, which serves in recommendation application. GreenFall discovers online customers’ purchase patterns by bit mining frequent itemsets and then, uses these patterns to recommend customers favorite items in the fastest speed. The strongest point of GreenFall algorithm is the high speed based on its simplicity, in which it only rolls bit sets with bitwise operations. As a result, it can respond recommendation request immediately. Fast speed, the strongest point of GreenFall, is motivated from the requirement of instant response in online commercial website. According to a survey of recommendation study, most current collaborative filtering algorithms do not consider mining frequent itemsets. Even in the area of bit mining frequent itemsets, other relevant researches make a lot of complex computations such as Index-BitTable, candidate itemsets generation and projection of bit vectors whereas GreenFall does not require any complex structure and complicated computation. So it is asserted that GreenFall is a novel algorithm.

GreenFall is evaluated on the popular MovieLens dataset created by GroupLens Research project. The evaluation proves preeminence in time consuming of GreenFall. There are 7 metrics used in this evaluation: MAE, MSE, precision, recall, F1, ARHR and time. These metrics are standardized in recommendation study. GreenFall is implemented by the multiplatform language Java which is simple and suitable for scientific purpose. So it is easy to understand the source code of GreenFall and improve it. The aforementioned challenge of this propose, immediate response in real-time recommendation application, is solved by GreenFall algorithm but I need to write its document and guidance report, which support scientists and software developers to use and improve it.

The proposed algorithm GreenFall is built upon Hudup “A framework of e-commercial recommendation algorithms”. Hudup is a middleware framework or “operating system” for e-commercial recommendation software, which support scientists and software developers build up their own recommendation algorithms based on this framework with low cost, high achievement and fast speed. I build up Hudup by myself. The trial version of Hudup is available at <http://www.locnguyen.net/st/products/hudup> and the default algorithm built-in Hudup is GreenFall and so, you can download this trial version to evaluate GreenFall. Hudup framework proposes simplest specifications on which GreenFall is based. Consequently, it is easy to access and improve GreenFall. According to such specifications, scientists can extend or modify GreenFall without excessive domain knowledge. Please read the introduction of Hudup available at <http://article.aascit.org/file/html/9400857.html> for more details about these specifications. There is no additional requirement of software and hardware when setting up Hudup framework except that you must install Java runtime environment (JRE).

The GreenFall algorithm was invented, implemented, and tested from 2011 to 2012. According to the first estimated plan, it requires 10 persons to complete project. However, only two persons get involved the project. One is responsible for composing and implementing the algorithm and one is responsible for testing and commenting on the algorithm. From 2011 to 2012, the salary per month for scientist in Vietnam is 800 USD.

* It takes 1 month to draft GreenFall.
* It takes 4 months to compose and revising the paper relevant to GreenFall. Note that the paper have been revised until now.
* It takes 4 months to implement GreenFall.
* It takes 3 months to test and fix GreenFall.

In general, the cost of GreenFall is 19,200 USD given two persons within 12 months.

GreenFall algorithm is evaluated on the dataset *MovieLens* dataset owned by GroupLens Research project. MovieLens is available at <http://grouplens.org/datasets/movielens>.

The source code of GreenFall is written by an open language Java. I am very enthusiastic to submit the source code. GreenFall is developed at Sunflower Soft Company whose Vietnamese name is Huong Duong. The company is established in October 29, 2011 with business license number 0311286418 in Ho Chi Minh city, Vietnam. The main operation in company is to do scientific projects relevant to data analysis and scientific publications. The company aims to creativity and all products focus on science and intelligence. These are strong point as well as the unique aspect of company. Followings are main achievements of the company:

* Releasing 2 scientific products: *Hudup* – A framework of e-commercial recommendation algorithm and *Phoebe* – A framework of estimating fetus weight and age. Please visit the web page <http://www.locnguyen.net/st/products> for more details about these products.
* Releasing about 9 scientific papers relevant to computer science, statistics and mathematics. Please visit the web page <http://www.locnguyen.net/st/articles> for more details about these papers.

I am founder of Huong Duong Company and so I endorse this proposal.

Best regards,

Loc Nguyen