　　　　　　　　　　 　　　　　　　　　　　　　 HCMC, 18th February 2019

Research Proposal for Master’s Thesis Project / Research Project / Survey for Doctoral Research Plan

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Intended Degree** | | **Master of Science** | | | | |
|  | | | | | | |
| Type of Major Research | | ✓ Master’s Thesis Project  □ Research Project  □ Survey for Doctoral Research Plan (Qualifying Examination) | | | | |
|  | | | | | | |
| **Student Name** | NGUYEN VAN QUANG | | | **Program** | ✓ ICT  □ QCF | |
|  | | | | | | |
| Supervisor | Assoc. Prof. Quan Thanh Tho | | **Signature** | **Second Supervisor** |  | **Signature** |
| □ Advisor for Minor Research Project  □ Advisor for Internship | | | | |  | **Signature** |
|  | | | | | | |
| **Title** | | | | | | |
| **Star2vec: from subspace embedding to whole-space embedding for intelligent recommendation system** | | | | | | |
|  | | | | | | |
| **Background** | | | | | | |
| * *Musto, Cataldo, et al. were implement a very simple method to recommendation system using word embedding technique. (1) Embedding words occur in items to vector space, (2) inferred a vector-space representation of the items by summing the representation of the words occurring in the document. (3) Inferred user profile vector by summing each item this user liked (4) finally using classical similarity to find . recommend top K items. The word embedding techniques they used were: Latent Semantic Indexing, Random Indexing and word2vec.* * *Bobadilla, Jesús, et al. compares the similarity between text-processing with recommendation system. They used historical data of users locations to recommend future check-in locations. As a text-processing use word as a unit, they consider each user location check-in as a word to put into the model* | | | | | | |
|  | | | | | | |
| **Objectives** | | | | | | |
| *We build a Content-Base recommendation system using word embedding technique.* | | | | | | |
|  | | | | | | |
| **Originality/Significance** | | | | | | |
| *Base on previous works, we extend:*   * *Recommend items not in one categories but only in many categories. (in different tables if data save in relational database)* * *Computed the distance between each items, which support to recommend in case “like item A and like items B”.* * *Support dislike problem: Recommend item “B” to users, in case they are very strong dislike item “A”,* * *No need to apply indexing techniques or any techniques in relational database.* | | | | | | |
|  | | | | | | |
| **Methodology and Evaluation** | | | | | | |
| * *Methodology:*  1. *Build a undirected graph from raw data to represent relationship between each categories of items.* 2. *Using deep random walk algorithm to generate a large corpus of documents* 3. *Using word2vec techniques to generate vectors represent items* 4. *With a new item, using k-nearest neighbors to find k items*  * *Evaluation:*  1. *Compare result and performance with relational database, especially in complex queries* 2. *Compare result items with other recommend system* | | | | | | |
|  | | | | | | |
| **Research Plan** | | | | | | |
| * *1/2019 : literature review and write thesis proposal* * *2/2019: do experiments* * *3/2019-4/2019: Write a the thesis* | | | | | | |
|  | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Submission Requirements for Research Proposal | | | | | | |
| ① Successfully accumulate all **required** courses. | | | | | | |
| ② The research plan should have sufficient contents | | | | | | |