論文の内容の要旨

論文題目

Automatic Task-based Content Selection and Representation for Content Recommendation

（コンテンツ推薦のためのタスクに基づくコンテンツ選択及び表現方法）

氏　名　深　澤　佑　介

In this thesis, we establish the method of content recommendation from a variety of contents.

Chapter 1 describes the background, related works, and purpose of this study. In the background, we set three requirements as follows: personalized content selection, comprehensible content representation and comprehensible content navigation. Then, we define the term used throughout the thesis. We investigated related works on content recommendation, content representation using keywords, task-based descriptions, domain-models, and task-models for content navigation. Then, we describe the structure of this thesis.

Chapter 2 introduces task-based content selection to achieve personalized content selection in content recommendation from various content. In order to achieve personalized content selection, we propose task-based profile representation in content-based recommendation, and evaluate the recommendation of a variety of contents (i.e. mobile web content and TV programs, restaurants, sightseeing spots, and hotels). Concretely, we tackle the following three research questions: 1) how to acquire a wide variety of tasks, 2) how to represent item profiles, and 3) how to represent user profiles. To answer question 1), we acquire tasks from the web automatically. For 2), we extract feature words for each task and match content descriptions with those feature words. For 3), we use SVM(Support Vector Machine) to automatically acquire user profiles from history data. An experiment on the metric of recommendation accuracy shows that the combination of task-based profile representation and term-based representation yields a 17.7% improvement in MAE compared to term-based profile representation or domain (content-category) based profile representation.

Chapter 3 introduces task-based content representation to achieve comprehensible content representation. In order to satisfy comprehensible content representation, we develop an application called TaskGuideRoad(TGR); it recommends YouTube content from task-based representation of sightseeing spots shown on a map. We extract task-based descriptions given the input of noun features of contents on the web. A user evaluation shows that TGR improved the frequency of watching video by about 15%. From the results of an interview, 90% of the subjects gave TGR high marks in terms of finding new enjoyable videos. Furthermore, we show that TGR allows users to find sightseeing spots from interesting and useful tasks at unknown places.

Chapter 4 introduces the automatic creation of task-model in order to achieve comprehensible content navigation. We develop two methods; 1) PMI(Pointwise Mutual Information) based Clustering: PMI uses the number of search results to measure the closeness of two concepts, and so does not need features of the tasks. 2)PMI-based features for BSK, BUC and FCA: the method creates features based on PMI between tasks to enhance the existing clustering methods of BSK, BUC and FCA. As PMI calculation needs only the number of search results, this method does not need any features to be specified beforehand. In both methods, we need to estimate the parent-children relationships between parent task candidates and children task candidates with the lowest possible error rate. To do that we propose a method that extends the PMI calculation; it divides the representation of tasks into a noun part and a verb part, and calculates the mutual information between them. A preliminary experiment shows that the proposed method can capture almost 80% of the pairs of correct parent task and child task.

Chapter 5 evaluates automatic creation of task-model, PMI-based task clustering and PMI-based task feature creation for existing clustering algorithms. We adopt taxonomic overlap between the created model and a ground truth model as the evaluation metric of the former. A comparison of feature types shows that the model created by BUC using PMI-based feature achieves, on average, 14.0 % taxonomic overlap, which is almost 94.0% of the taxonomic overlap (average 14.9%) of the model created by BUC using token-based features. This shows that if the tasks are not supported by a sufficient number of descriptions, PMI-based creation works well. By comparing the proposed PMI-based Clustering to three existing algorithms, we find that PMI-based clustering achieves 25.87% overlap, which is a 32.0 % improvement over existing methods; i.e. BSK(=15.23%), BUC(=14.94%) and FCA(=17.60%). The worst case calculation

complexity also shows that PMI-based clustering has lower complexity (=O(n)) than BSK(=O(m times n^2)), BUC(=O(m times n^2)) and FCA(=O(c^k)|k = min (m,n)), where n represents the number of tasks, and m represents number of feature elements. This means that, PMI-based Clustering is expected to create precise task-models in shorter time than the other methods.

Chapter 6 describes a user test conducted to evaluate automatic creation of task-model; it investigates the navigability of the task-model from the view point of ``degree of task clustering" and ``ease of reaching the desired cluster". As for the metric ``degree of task clustering", twelve domains in the model created by PMI-based Clustering are judged as ``well clustered" or ``pretty well clustered", which is the best performance among all methods i.e. FCA(=9 domains) and BUC(=10 domains). As for the metric ``ease of reaching the desired cluster", nine domains in the model created by PMI-based Clustering are judged as ``Easy to reach most of the clusters", which superior to the performance of the other methods e.g. BSK(=2 domains), FCA(=4 domains) and BUC(=1 domain). This result shows that the task-models created by PMI-based Clustering also have the best quality from the viewpoint of ``navigability".

Chapter 7 describes the conclusion and some possible extensions of this thesis.