

Title	ユーザーの趣向分析のための機械学習と証拠推論の統合に関する研究
Author(s)	VO, DUC VINH
Citation	
Issue Date	2022-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/17781
Rights	
Description	Supervisor:Huynh Nam Yan, 先端科学技術研究科, 博士

氏 名	VO, Duc Vinh		
学 位 の 種 類	博士（知識科学）		
学 位 記 番 号	博知第 298 号		
学 位 授 与 年 月 日	令和 4 年 3 月 24 日		
論 文 題 目	A Study on the Integration of Machine Learning and Evidential Reasoning in User Preferences		
論 文 審 査 委 員	主査	HUYNH Van Nam	北陸先端科学技術大学院大学 教授
		橋本 敬	同 教授
		DAM Hieu Chi	同 教授
		白井 清昭	同 准教授
		乾口 雅弘	大阪大学 教授

論文の内容の要旨

With the increase in use of social networks, users on these micro-blogging platforms usually share their thoughts and interests via short texts such as posts, status, or reactions. Capturing user preferences or interests from these kinds of data has attracted much attention. This process is known as the user profiling or user preferences. This problem aims at processing, inferring, and extracting a list of weighted keywords (or a semantic-based structure) that correctly represent the expertise or preferences of a specific user on these networking platforms. The problem has various potential applications in practice such as researcher finding in academic projects, item recommendation in e-commerce systems, or job offering in labor markets. However seeking an efficient solution is not trivial but a challenging task. Researchers working on this problem usually face following common challenges: (1) The *data sparsity* and *cold-start* issues existing in user texts; (2) user preferences dynamically *change* over time; (3) social networking users usually create lots of *short documents*. The consecutive documents are often not very closely related to each other. This causes difficulties for inferring the desired profile; (3) data may come in different formats (e.g., images, texts, or reactions) from multiple sources (e.g., one user may simultaneously have multiple accounts across networking platforms).

This research is motivated by three major factors. The first factor is reasoning ability in evidence theory (also preferred to as Dempster-Shafer theory). This theory is theoretically well-studied to become a full-fledged theory of uncertainty. It has been widely applied to various topics, including machine learning problems (e.g., classification and clustering), problems involving uncertainty (e.g., database management with uncertainty), and multiple-attribute decision making. The second factor is that advancements in machine learning and deep learning have shown significant remarkable achievements in both academic and practice. Machine learning practitioners proposed many robust, data-driven models that can learn from input observations to make accurate predictions, find hidden patterns, or even create new instances that are very similar to the input data. The third factor is the increasing of short texts shared by users on social networking platforms. Additionally, a number of open-sourced libraries make such kind of data obtainable and feasible to be processed.

Taking the aforementioned challenges and motivations into consideration, this research proposed two novel

frameworks for *capturing user profiles using short texts under both static and dynamic scenarios*. The first framework is designed for inferring *static* profiles, which is based on evidence theory and *k*-means clustering. The second framework is designed for inferring *dynamic* profiles, which is based on deep generative networks and evidence theory. In both proposed frameworks, advanced machine learning techniques, such as *k*-means clustering and deep generative neural networks, are used for concept learning from user short texts. The learned concepts form the so-called *frame of discernment* in Dempster-Shafer theory for reasoning process. These concepts are quantitatively transformed into the so-called *mass function* in the evident theory by *maximum a posterior estimation*. The derived mass functions are then considered as pieces of evidence at the reasoning phase. Finally, Dempster's rule and the so-called *pignistic probability distribution* are used for information fusion and profile extraction. The experiments on short text data sets verified that the proposed methods outperform baseline models on many evaluation metrics. Additionally, we also propose an approach for visualizing the fluctuation of user preferences on various topics over time by using the output of the proposed frameworks. This visualization may reveal significant insights that are useful for many practical applications.

Keywords: User profiling, user preferences, Dempster-Shafer theory, user profile visualization, deep generative networks.

論文審査の結果の要旨

Inferring user profiles based on texts created by users on social networks has a variety of applications in Web-based decision support systems and it has received considerable attention from the research community over the last decade. Previous work based on generative topic models such as Latent Dirichlet Allocation (LDA) or its variations usually encounter the sparsity problem of short texts. The main objective of this dissertation was to develop a new integrated approach that combines advanced machine learning (ML) techniques with Dempster-Shafer (DS) theory of evidence for the problem of user profile learning, which is capable of not only working well with short texts but also dealing with uncertainty inherently in user texts. The key idea of the proposed approach is that ML techniques are first utilized for concept learning that determines the frames of discernment in the language of DS theory for the target problem of extracting top-*n* keywords for user profile, and then so-called mass functions are determined via maximum a posterior estimation for text data and combined using Dempster's rule for inferring users' keyword distributions. Two frameworks for user profile learning using short texts have been eventually developed for static and dynamic scenarios respectively.

Particularly, the first framework for inferring static profiles is developed based on DS theory, word embedding, and *k*-means clustering, while the second framework for inferring dynamic profiles is developed by integrating deep generative networks (i.e., variational autoencoder and generative adversarial network) and DS theory to cope with dynamic changes of user preferences over time. Thanks to the ability of combining multiple pieces of information from many documents, the proposed approach is flexible enough to be scaled when input data coming not only in multiple modes but also from different sources on Web environments. The effectiveness of the proposed frameworks is validated by experimental studies conducted on datasets crawled from Twitter and Facebook.

This dissertation has made excellent contributions to methodological and experimental developments within the areas of preference modeling and machine learning. The research work presented in this dissertation has resulted in two high quality journal papers and one conference paper.

In summary, Mr. VO Duc Vinh has completed all the requirements in the doctoral program of the School of Knowledge Science, JAIST and finished the examination on January 31, 2022, all committee members approved awarding him a doctoral degree in Knowledge Science.

Date: *4 February 2022*