# A study on the integration of machine learning and evidential reasoning in user preferences

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.