Abstract.

E-commerce websites have an important role in business. Many people purchase products by referring to the e-commerce websites every day. Product ratings and reviews written by the crowd on e-commerce shopping websites have become a critical information source for making recommendations and purchasing decisions. An important challenge, however, is that the vast majority of products (e.g., 90% of products on amazon.com) do not receive enough attention and lack sufficient ratings and reviews by the users; hence, they constitute the so-called cold products.

One solution to address cold products, which has already been studied in the literature, is to exploit side information and build representations for system elements (such as products, users, reviews) and predict ratings according to those representations or generate reviews for these products by sampling review sentences from closely related warm products.

A majority of the works in the literature rely on product specification as the side information. This would make limitations to generalizability of the work, since structured specifications are not available in all domains.

Our work in this thesis differs in that it not only employs product specification, but also employs a wide range of information such as free form textual content as well as product-review, product-user, and user-review interactions. We propose to use such side information and exploit neural architectures to build semantic representations for the system entities. The generated representations are exploited to model relations between cold products and other entities. More specifically, we propose to solve the cold product problem in a generalizable form by three steps: 1) We proposed to exploit free form textual content to build semantic representations for products and users. Then we use these representations in the recommender frameworks and improve the recommendation task. 2) We proposed to use a combination of structured specifications and textual content to make representations for products and reviews. The generated representations reveal implicit relations between products and reviews and are used to select reviews for the cold product. 3) We proposed a generalizable framework to select review sentences. As such, we show how neural graph embeddings can be used to encode product, user, and review information into an attributed heterogeneous graph representation. We further propose how review relevance and importance can be considered using graph traversal to select appropriate review sentences for a given cold product.

Finally, we systematically compare the performance of our work with those of several state-of-the-art baselines on various datasets collected from CNET.com and rottentomatoes.com with different characteristics from both quantitative (e.g., the Recall Oriented Understudy for Gisting Evaluation (ROUGE) metrics) and qualitative aspects and show how our proposed approach was able to provide statistically significantly improved performance over various strong baselines.