

**Hashtagger+: Efficient High-Coverage Social Tagging of Streaming News**

**Abstract:**

News and social media now play a synergistic role and neither domain can be grasped in isolation. On one hand, platforms such as Twitter have taken a central role in the dissemination and consumption of news. On the other hand, news editors rely on social media for following their audience’s attention and for crowd-sourcing news stories. Twitter hashtags function as a key connection between Twitter crowds and the news media, by naturally naming and contextualizing stories, grouping the discussion of news and marking topic trends. In this work, we propose Hashtagger+, an efficient learning-to-rank framework for merging news and social streams in real-time, by recommending Twitter hashtags to news articles. We provide an extensive study of different approaches for streaming hashtag recommendation, and show that pointwise learning-to-rank is more effective than multi-class classification as well as more complex learning-to-rank approaches. We improve the efficiency and coverage of a state-of-the-art hashtag recommendation model by proposing new techniques for data collection and feature computation. In our comprehensive evaluation on real-data, we show that we drastically outperform the accuracy and efficiency of prior methods. Our prototype system delivers recommendations in under 1 minute, with a Precision@1 of 94 percent and article coverage of 80 percent. This is an order of magnitude faster than prior approaches, and brings improvements of 5 percent in precision and 20 percent in coverage. By effectively linking the news stream to the social stream via the recommended hashtags, we open the door to solving many challenging problems related to story detection and tracking. To showcase this potential, we present an application of our recommendations to automated news story tracking via social tags. Our recommendation framework is implemented in a real-time Web system available from insight4news.ucd.ie.

**Existing System:**

Our previous work has introduced an accurate learning-to-rank (L2R) approach for streaming hashtag recommendation to news, but its efficiency and coverage is still not appropriate for practical use. The model in has a time consuming data collection stage for each new article, and thus requires 12 hours to deliver 89 percent precision and 66 percent article coverage.

The relationship between the news story and the hashtags is very dynamic, with new hashtags being created and adopted by Twitter users at a rapid pace. It may be seen from this example that for applications aiming to exploit hashtagging, it is critical to capture the dynamic co-evolution of news and hashtags, as the news story evolution influences the Twitter discussions, which in turn may affect the news. We note that the content of some articles may not be obviously related to a story, but a hashtag recommender can use the social discourse to create a bridge between news articles.

**Proposed System:**

We propose a real-time hashtag recommendation approach that is able to efficiently and effectively capture the dynamic evolution of news and hashtags. Most prior approaches for hashtag recommendation work on static datasets and do not account for the emergence and disappearance of hashtags. Many approaches use topic/class modeling by considering hashtags as topics, and mapping news articles to topics using content similarity. As the relevant hashtags change quickly and the news and Twitter environments are highly dynamic, approaches that use multi-class classification need continuous retraining to adapt to new content. Additionally, to train models, these methods rely on tweets that contain both hashtags and URLs. Such tweets are very few and tend to be noisy, which may explain the low accuracy of prior methods.