Motivation and Problem Definition

Keeping up with continuous text streams, like daily news, costs a considerable amount of time. We developed an interactive classification interface for text streams that learns user-specific topics from the user’s labels and partitions incoming data into these topics.

Current approaches that categorize unstructured text documents use pre-trained learning models for text classification. In the case of a continuous text stream, the usefulness is limited, as these models cannot adapt their categories or learn new terminology.

Methodology

To adapt to changing terminology and to learn user-specific topics, we utilize a variant of active learning in an iterative process of model training.

We present visual active learning for text streams by visualizing the topic affiliations in a Star Coordinates visualization. This visualization provides novel direct interaction tools for iterative model training.

Evaluation

We developed a simulation to compare the accuracy of visual active learning and classic active learning.

In a preliminary user study, we compared our visualization to a list-based interface for news retrieval and active learning.

Results

Through our evaluation, we could show that our visualization is a very effective user interface for active learning of streaming data.

Assign document ‘It Looked Like a Beer Belly...’ to topic ‘health’.