Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance

Abstract:

The vast majority of existing approaches to opinion feature extraction rely on mining patterns only from a single review corpus, ignoring the nontrivial disparities in word distributional characteristics of opinion features across different corpora. In this paper, we propose a novel method to identify opinion features from online reviews by exploiting the difference in opinion feature statistics across two corpora, one domain-specific corpus (i.e., the given review corpus) and one domain-independent corpus (i.e., the contrasting corpus). We capture this disparity via a measure called domain relevance (DR), which characterizes the relevance of a term to a text collection. We first extract a list of candidate opinion features from the domain review corpus by defining a set of syntactic dependence rules. For each extracted candidate feature, we then estimate its intrinsic-domain relevance (IDR) and extrinsic-domain relevance (EDR) scores on the domain-dependent and domain-independent corpora, respectively. Candidate features that are less generic (EDR score less than a threshold) and more domain-specific (IDR score greater than another threshold) are then confirmed as opinion features. We call this interval thresholding approach the intrinsic and extrinsic domain relevance (IEDR) criterion. Experimental results on two real-world review domains show the proposed IEDR approach to outperform several other well-established methods in identifying opinion features.

450/526, Trichy Main Road, Near Sri Sakthi Kaliamman Temple, Dadhagapatti Gate, Salem-636 006, Tamil Nadu, India.