Using weak supervision to scale the development of machine-learning models for social media-based marketing research

Jennifer Cutler and Aron Culotta

Abstract

Marketers have expressed substantial enthusiasm about the potential of social media data to enhance marketing research, and the computer science literature provides many examples of using the text and network connections of social media users to infer measurements of interest to marketers. Despite this, the adoption of such machine-learning approaches has been surprisingly limited in marketing practice, in part due to the hurdle of procuring the labelled training data typically used to build such models. This paper discusses how the organic structure of social media can often be leveraged to circumvent the need for such curated data labels. It describes two emerging methodological themes of weak supervision — training on exemplars and training on groups — that are broadly promising towards this goal, providing examples of how they have been applied towards a variety of marketing tasks without requiring any manually labelled training data, and in some cases, requiring nothing more than a single keyword as input. This paper presents these approaches in the hope that examples will inspire and facilitate the development of a broader range of flexible, scalable and cost-effective models for social media-based marketing research, and stimulate additional research in this area.