An Integrated Text Analytic Framework for Product Defect Discovery

In this study, we propose an integrated text-based analytic framework called SMART (Social Media Analytic fRamework using Text) to quantify social media content and extract important features from such postings that can be subsequently used to discover and analyze product defects

- contributions
- Firstly，We test the new framework in two diverse domains—vehicles and consumer electronics
- Secondly，It incorporates methodological innovations in data acquisition and transformation as we consider seven categories of input cues (lexical, stylistic, social, sentiment, distinctive terms, product features, and semantic cues）
- Thirdly, this study performs a multivariate analysis vs. univariate consideration of a restricted set of input variables

- Fourth, we have added new baselines to show the relative advantages of our new framework over existing text analytical approaches
- Finally, we provide a comprehensive discussion on the business value and organizational implementation of text analytics(in product quality issue discovery, limitations and biases)

- In section 2, we begin with a discussion of the business value of automated defect discovery from social media.

- Reviewing relevant literature on defect management and text classification in section 3（ Defect Management & Extracting Useful Features from Text ）

    1.text mining ： gathering comments➔text preprocessing➔feature extraction➔text modeling and analytics

   2.how to extract features using techniques

      （NLP, 语言分析, 信息检索）

    how to build models using those features to uncover product defects.

Lexical features（词汇特征， ）

- We present an integrated defect discovery framework that considers a comprehensive assortment of firstand second-order input features for text analytics in section 4

Figure 1   SMART: An Integrated Text Analytic Framework for Product Defect Discovery from Social Media
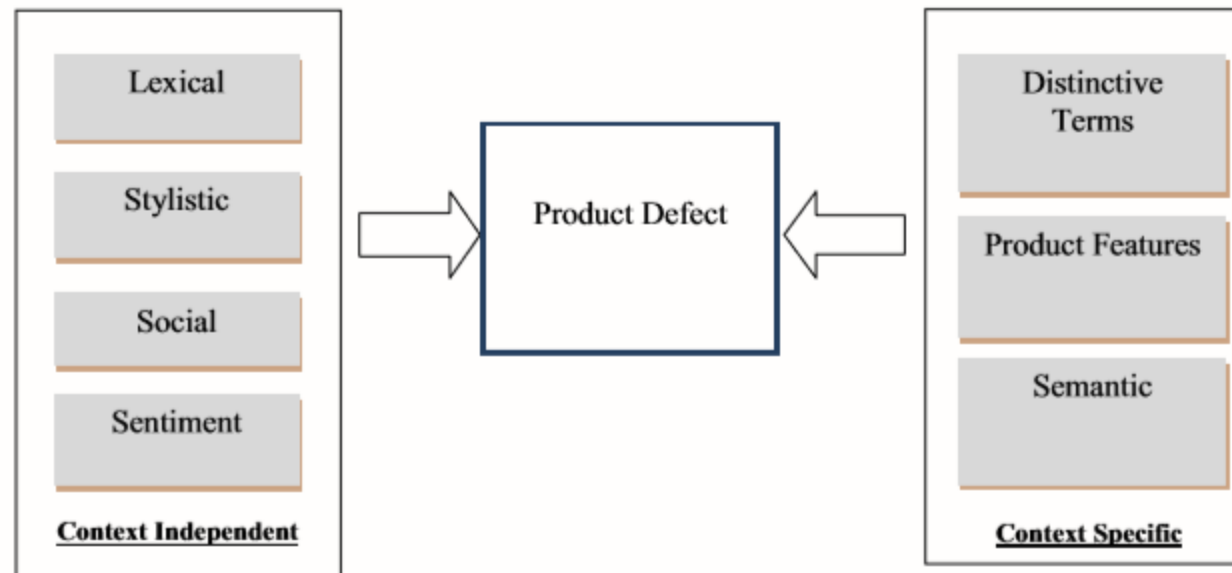
- In section 5, we test our framework using case studies from two domains: automotives, and consumer electronics

   defect classification by each expert rater

   we assessed inter-rater agreement k=0.96

   （Stylistic information &social cues &sentiment cues & distinctive words product features semantic cues）

- Section 6 describes limitations of our study, particularly statistical and human biases in the source data and defect management process

- Section 7 concludes and discusses directions for future research.