Identifying Entity Aspects in Microblog Posts

Introduction

- Scenario: Online reputation management
- What do people say about an entity? (company, organization, individual, product)
- Identification of aspects discussed on microblog posts
- Aspects = products, services, competitors, key people related to the entity

Identifying Entity Aspects

Dataset

- 94 entities, 17,775 tweets, 177 tweets/entity
- Built upon WePS-3 ORM Task Dataset
- Disambiguated company names (e.g., apple fruit vs. Apple Inc.)
- Pooling methodology
- Three annotators with substantial agreement (Cohen’s kappa > 0.6)
- Annotated 2455 terms, 1304 aspects (54.11%)
- Most of the true aspects are nouns (89.72%)

Entity | Aspects
---|---
Apple Inc. | ipad, iphone, prototype, team, gmos, employee advertising, headphones, digital, pro, music, secret, disc, ls, braava, camera, vegas, battery, explosion, iphone, coffee, butter, leg, fragrances, battery, drink, meats
Starbucks | coffee, latte, tea, fragrances, banner, drink, meats
Sony | coffee, latte, tea, fragrances, banner, drink, meats

Experiments and Results

- All Words
  - TF.IDF significantly outperforms PLM and OO in precision.
  - The results for OO are much lower than for the other methods.
- Noun Filter
  - Applying a part-of-speech filter and only consider terms tagged as nouns
  - For all methods, MAP and precision values are slightly higher than in the all-words condition; considering only nouns helps to identify aspects.
  - PLM best method using Noun filter.
- Observations (after manually inspecting the results):
  - Results for TF.IDF, LLR and PLM are very similar.
  - OO tends to return more subjective terms as aspects (syntactic parsing errors).
  - Examples: haha, pl, safety, win
  - OO has more difficulty to filter out generic terms.
  - Examples: new, use, today, come

Conclusions

- Simple statistical methods such as TF.IDF are a strong baseline for the task of identifying entity aspects, significantly outperforming opinion-oriented methods.
- Only considering terms tagged as nouns improves the results for all the methods analyzed.