



Aspect-based Sentiment Analysis: Information Extraction and Applications for Portuguese texts

Student: Pedro Paulo Balage Filho (PhD) - Scholarship: FAPESP

Advisor: Thiago Alexandre Salgueiro Pardo

Interinstitutional Center for Computational Linguistics (NILC)

balage@icmc.usp.br

Registration Date: July, 2012, Qualification Date: January, 2014, Limit Date: December, 2015

Abstract

Sentiment Analysis is a recent area of Natural Language Processing (NLP) which aims to process texts with opinion, sentiment and subjectivity. In this area, there are three levels of analysis: text level, sentence level and **aspect or entity level**. In the aspect-based sentiment analysis, feelings are extracted according to the aspects they modify. The focus of this work is on the **aspect-based sentiment extraction**. The aim of this work is **to produce techniques to extract opinions** and their modifiers. Also, these techniques should **improve other activities in NLP**, such as automatic summarization, question answering, recommendation systems, among others.

Example

O **iPhone** é um bom aparelho. Sua **bateria** é excelente. A **qualidade** é muito boa, mas o **preço** é exorbitante.

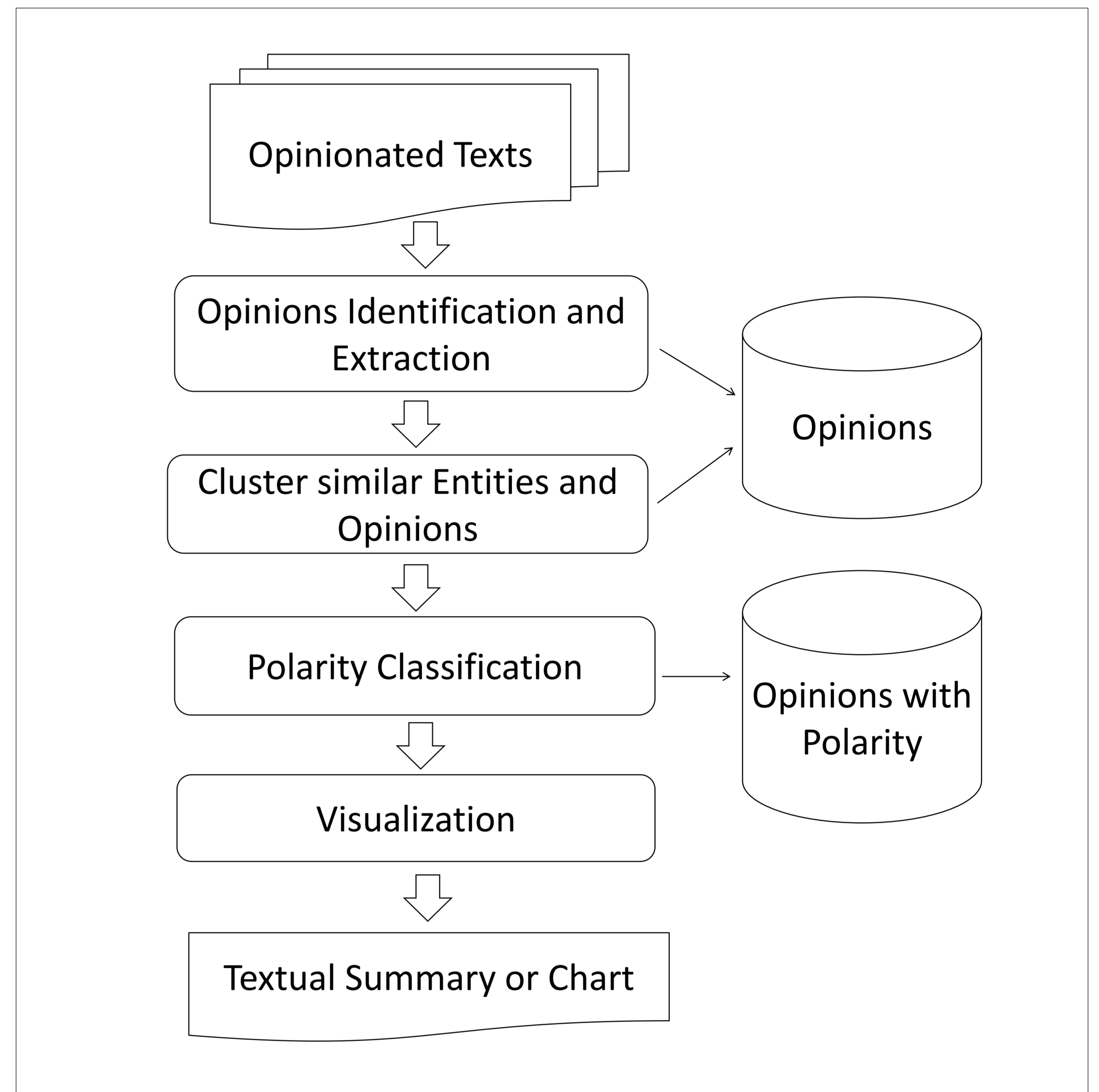
(The iPhone is a good device. Its battery is excellent. The quality is very good, but the price is exorbitant.)

Aspect	Predicate	Sentiment
lphone	Bom (good)	Positive
Bateria (battery)	Excelente (excellent)	Positive
qualidade (quality)	muito boa (very good)	Positive
preço (price)	Exorbitante (exorbitant)	Negative

Another Example - Challenges

*Opinião: Esse **celular** é razoável e tem um bom preço. Mas quem pretende comprar o **LG KP215** precisa saber que: A **câmera** tem muito ruído na imagem (muito mesmo!); o **mp3 player** tem um volume menor que o de outros tocadores de mp3; o **alto-falante externo** é baixo (em lugares muito barulhentos não se pode ouvir nem a campainha); **ele** só aceita vídeos de resoluções baixas; não funciona com a operadora Vivo (já que o **transmissor** dele é dual-band, 900/1800 Mhz, e a **Vivo** opera em São Paulo a 850Mhz. Se você não se importar com esses detalhes, vá em frente. Mas **eu** comprei e estou muito arrependido, pois eu pensei que estava levando para casa um bom celular, mas depois, veio a decepção.*

Basic Architecture



Preliminary Results

- ✓ Statistics obtained from an annotated corpus
- ✓ Opinion identification and extraction techniques
- ✓ Polarity classification system

Publications

Balage Filho, P. P., and Pardo, T. (2013). An Evaluation of the Brazilian Portuguese LIWC Dictionary for Sentiment Analysis. In S. M. Aluísio & V. D. Feltrim (Eds.), Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology (pp. 215–219). Fortaleza, CE, Brazil: Sociedade Brasileira de Computação.

Balage Filho, P. P., & Pardo, T. A. S. (2013). NILC_USP: A Hybrid System for Sentiment Analysis in Twitter Messages. In Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013) (pp. 568–572). Atlanta, Georgia, USA: Association for Computational Linguistics.

Future Steps

- Development of better opinion extraction techniques
- Opinion Clustering
- Intrinsic and Extrinsic Evaluations