




# Opinion Mining in Business

Neagu Ana Maria Cristina



# Table of Contents

A thick yellow horizontal bar spans the width of the slide, with a vertical yellow bar extending downwards from its right end.

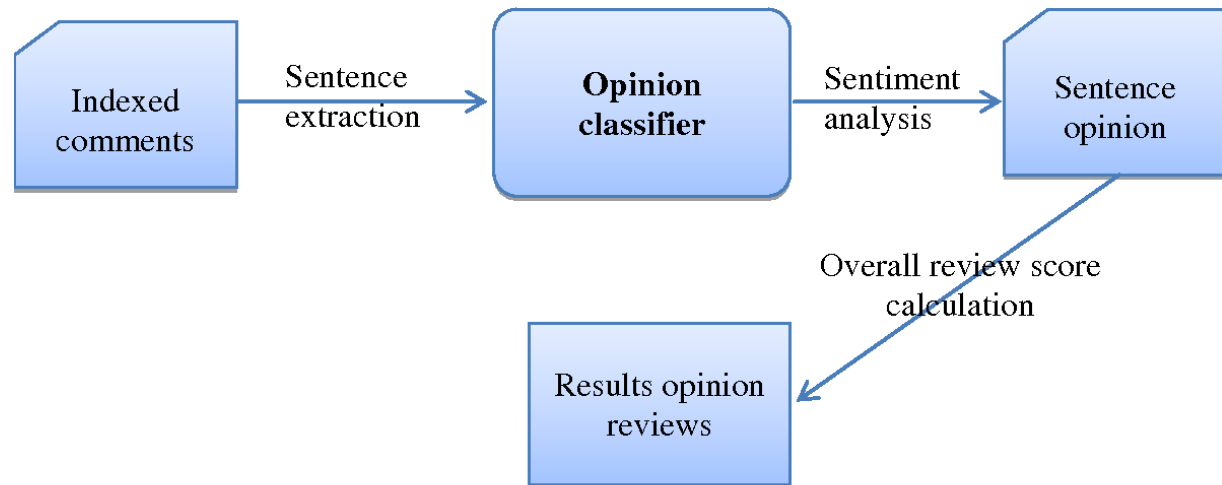
- Introduction
  - Problem description
  - Challenges
  - Opinion mining and ranking algorithms
  - Conclusions and Future work
- 
- A thin grey horizontal bar spans the width of the slide at the bottom.



# Problem description

The opinion mining process:

1. Data collection and pre-processing
2. Classification
3. Aggregation and presentation of results



# Challenges

- Feature extraction
  - “camera”, “sound”, “voice”
- Object identification
  - “Nokia phone”, “moto phone”, “BestBuy”

*“(1) Yesterday, I bought a Nokia phone and my girlfriend bought a moto phone. (2) We called each other when we got home. (3) The voice on my phone was not clear. (4) The camera was good. (5) My girlfriend said the sound of her phone was clear. (6) I wanted a phone with good voice quality. (7) So I was satisfied and returned the phone to BestBuy yesterday.”*

# Challenges

- Opinion orientation classification
- Different people, different writing style
- Opinions change in time
- Misleading opinions due to sarcasm and irony



# Algorithms

## 1. The High Adjective Count algorithm (HAC)

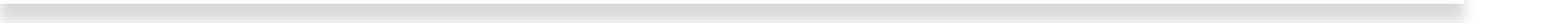
- Identifies and extracts the potential features from the reviews of the product
- Starts by identifying adjectives and nouns (POS tagging)
- The closest adjective is more likely to describe the noun

Tag	Meaning	English Examples
ADJ	adjective	<i>new, good, high, special, big, local</i>
ADP	adposition	<i>on, of, at, with, by, into, under</i>
ADV	adverb	<i>really, already, still, early, now</i>
CONJ	conjunction	<i>and, or, but, if, while, although</i>
DET	determiner, article	<i>the, a, some, most, every, no, which</i>
NOUN	noun	<i>year, home, costs, time, Africa</i>
NUM	numeral	<i>twenty-four, fourth, 1991, 14:24</i>
PRT	particle	<i>at, on, out, over per, that, up, with</i>
PRON	pronoun	<i>he, their, her, its, my, I, us</i>
VERB	verb	<i>is, say, told, given, playing, would</i>
.	punctuation marks	<i>. , ; !</i>
X	other	<i>ersatz, esprit, dunno, gr8, univeristy</i>

# Algorithms

A thick yellow horizontal bar spans the width of the slide, with a vertical yellow bar extending downwards from its right end.

## 2. The Max Opinion Score algorithm (MOS)

- Takes these potential features as input
  - Assigns scores to them and finally helps in classifying every review as positive, negative or neutral
  - A higher score indicates a stronger opinion than lower score
- 
- A thick grey horizontal bar spans the width of the slide at the bottom.



# Algorithms

## Datasets

- customer reviews of 4 products:
- digital camera: Canon G3
- digital camera: Nikon coolpix 4300
- cellular phone: Nokia 6610
- dvd player: Apex AD2600  
Progressive-scan DVD player

CANON G3		DVD		NIKON		NOKIA 6610	
Feature	Score	Feature	Score	Feature	Score	Feature	Score
life	3.600	audio	4	couple	2.800	network	4
viewfinder	3.333	support	2.800	time	2.733	Samsung	3.100
feel	3.200	review	2.800	quality	2.358	service	2.889
megapixel	3.200	color	2.457	battery	2.220	clarity	2.800
zoom	2.581	sound	2.083	feature	2.200	quality	2.179
month	2.400	machine	2.027	service	2	reception	2.101
print	2.400	Disney	2	assist	2	option	2.067
setting	2.150	mention	2	camera	1.974	speakerphone	1.911
resolution	1.693	feature	1.980	get	1.947	radio	1.818
plenty	1.600	purchase	1.800	software	1.840	thing	1.720
quality	1.535	picture	1.720	Nikon	1.733	deal	1.700
use	1.488	use	1.711	scene	1.667	see	1.642
camera	1.450	amazon	1.655	use	1.587	camera	1.636
...	...	...	...	...	...	...	...
battery	-0.150	button	0.286	card	1.032	work	0.050
card	-0.267	something	0.238	speed	0.880	plan	-0.372
focus	-0.471	day	0.133	print	0.541	way	-0.500
bit	-0.975	unit	0.120	way	0.400	operate	-0.833
lot	-1.200	thing	-0.250	photography	-0.227	power	-1
strap	-2.100	display	-0.250	set	-0.286	customer	-1.087
hand	-2.800	jpeg	-0.500	backup	-0.500	input	-1.200

# Conclusions and Future work

- Content on social media that contains the thoughts and desires of clients is known as important material from which organizations can find relevant information.
- **Future work:**
  - classify the reviews into positive, negative or neutral, based on the features identified and their scores;
  - Calculate the precision and the accuracy of the algorithm proposed;

