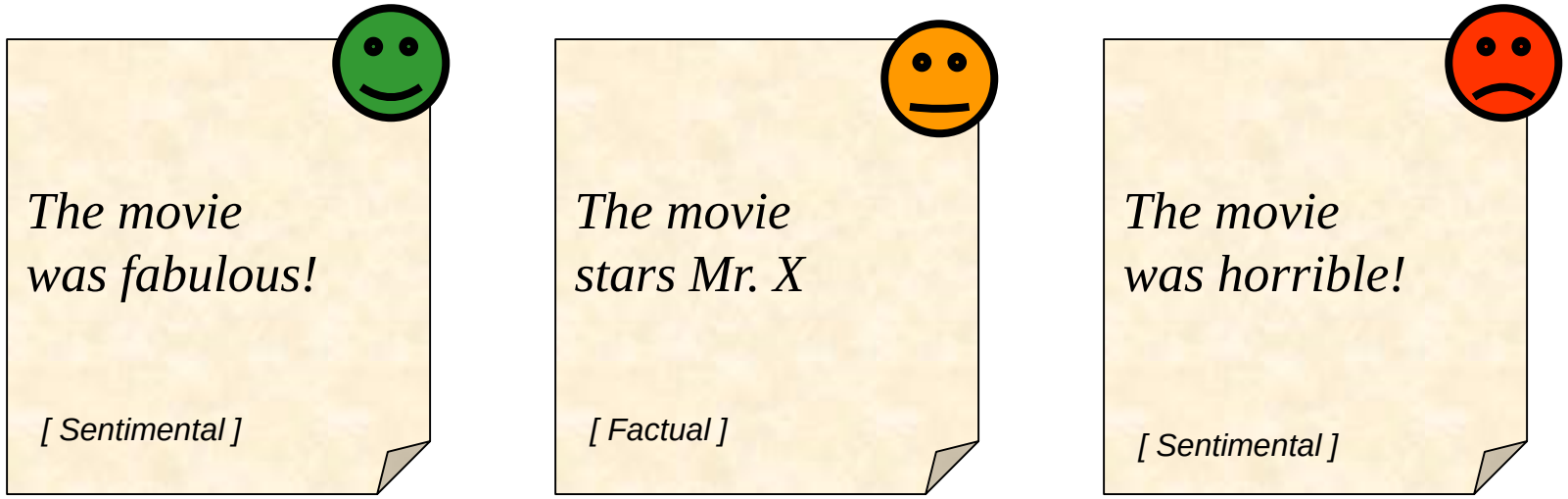
UNIVERSITI KEBANGSAAN MALAYSIA
The National University of Malaysia

Aspect-Based Sentiment Analysis of online reviews

Suhaila Zainudin, Zohreh Madhoushi and Abdul Razak Hamdan
Data Mining and Optimization Research Grup
Center for Artificial Intelligence Technology
Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia

What is sentiment analysis?

Sentiment analysis and opinion mining is the field of study that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions from written language. (Liu 2012)



Why sentiment analysis?

- With huge volume of opinionated text, Normal users and organizations have difficulty summarizing opinions.
- This information is **unstructured**, with **lower quality**, full of **noise** and **spams** and it is not something that is easily **machine processable**.
- Sentiment analysis is hard and a thriving research area in NLP, ML, data and text mining.

Why online reviews?

Pre Web

- Friends and relatives
- Acquaintances
- Consumer Reports

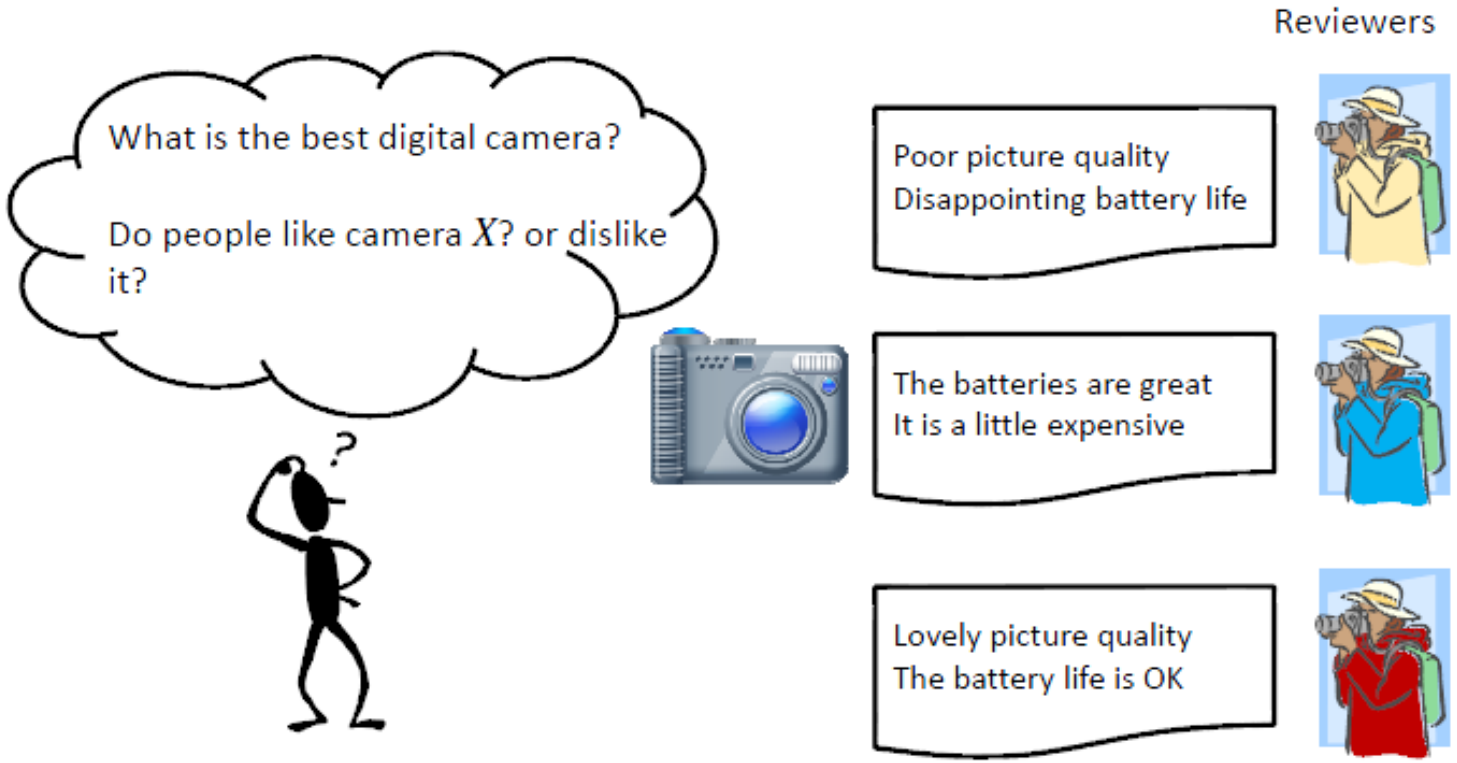
Post Web

"...I don't know who...but apparently it's a good phone. It has good battery life and..."

- Blogs (google blogs, livejournal)
- E-commerce sites (amazon, ebay)
- Review sites (CNET, PC Magazine)
- Discussion forums (*forums.craigslist.org, forums.macrumors.com*)
- Social Networks(Twitter, Facebook)

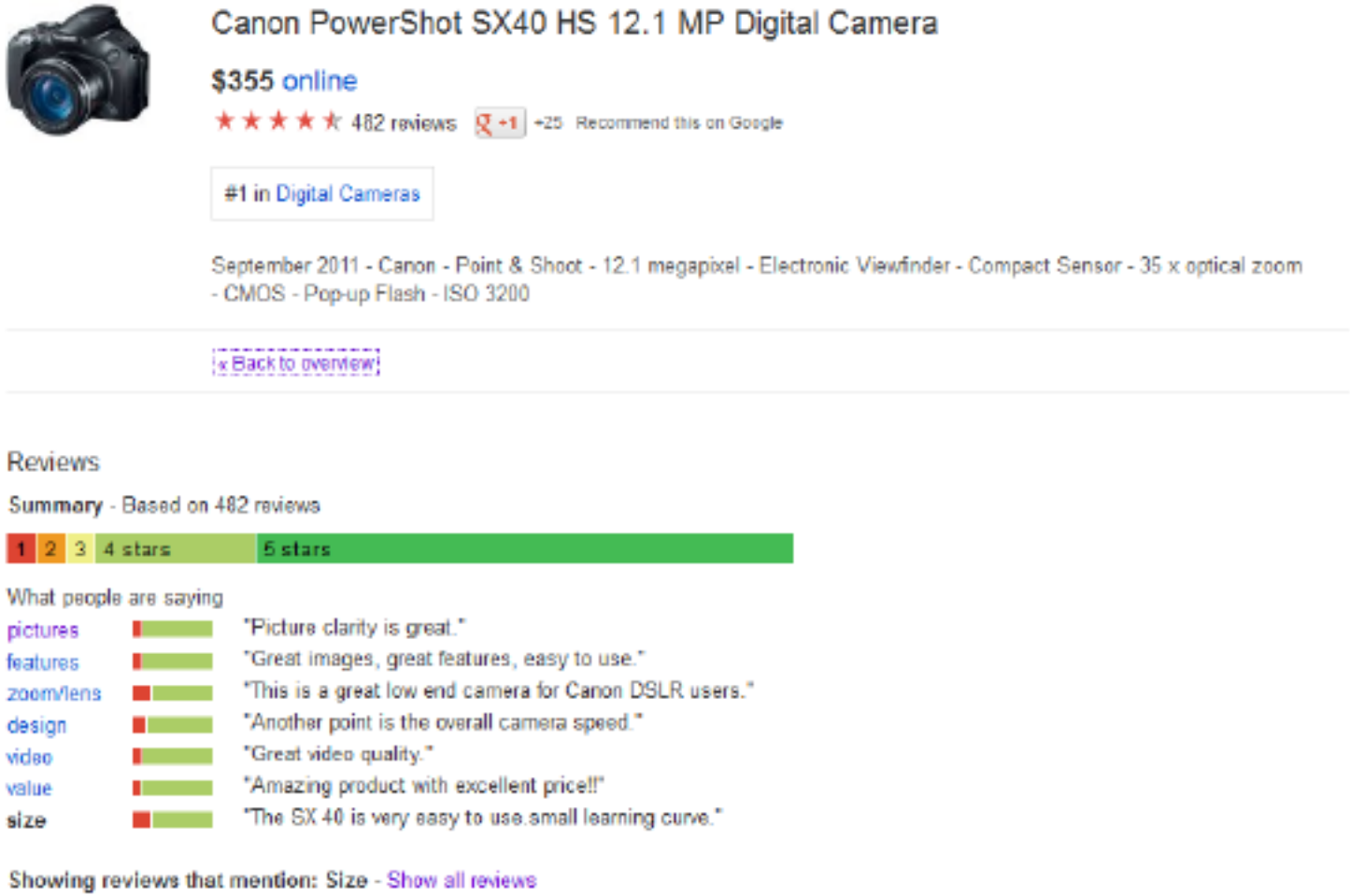
Why online reviews? Cont.…

- There are too many reviews to read



Aspect-Based Sentiment Analysis (ABSA)

Google product



Canon PowerShot SX40 HS 12.1 MP Digital Camera

$355 online

★★★★½ 482 reviews +1 +25 Recommend this on Google

#1 in Digital Cameras

September 2011 - Canon - Point & Shoot - 12.1 megapixel - Electronic Viewfinder - Compact Sensor - 35 x optical zoom - CMOS - Pop-up Flash - ISO 3200

« Back to overview

Reviews

Summary - Based on 482 reviews

1 2 3 4 stars | 5 stars

What people are saying

pictures	"Picture clarity is great."
features	"Great images, great features, easy to use."
zoom/lens	"This is a great low end camera for Canon DSLR users."
design	"Another point is the overall camera speed."
video	"Great video quality."
value	"Amazing product with excellent price!!"
size	"The SX 40 is very easy to use.small learning curve."

Showing reviews that mention: Size - Show all reviews

ABSA

Two main tasks for ABSA:

- aspect extraction
- aspect polarity estimation/Sentiment rating

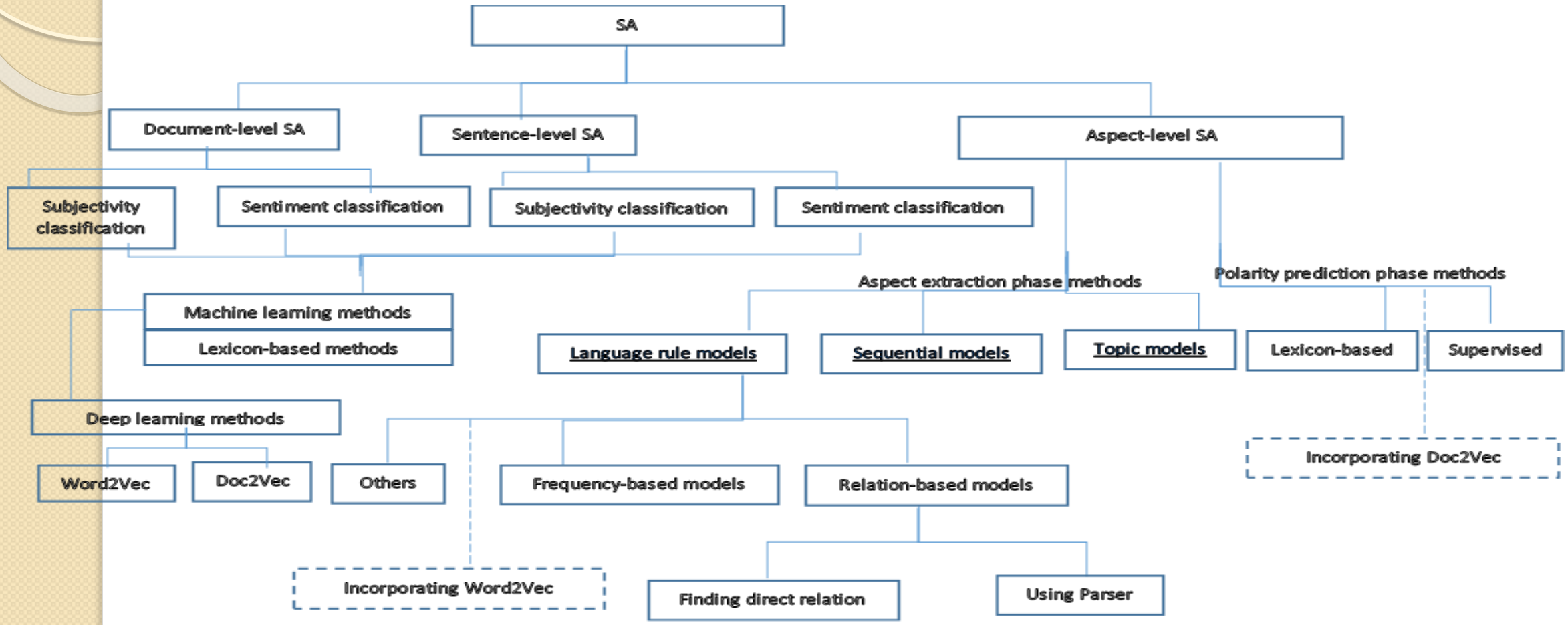
Input

Canon GL2 Mini DVD Camcorder
…excellent zoom …. Blurry lcd … great picture quality….accurate zooming … poor battery … inaccurate screen … good quality … affordable price …

Output

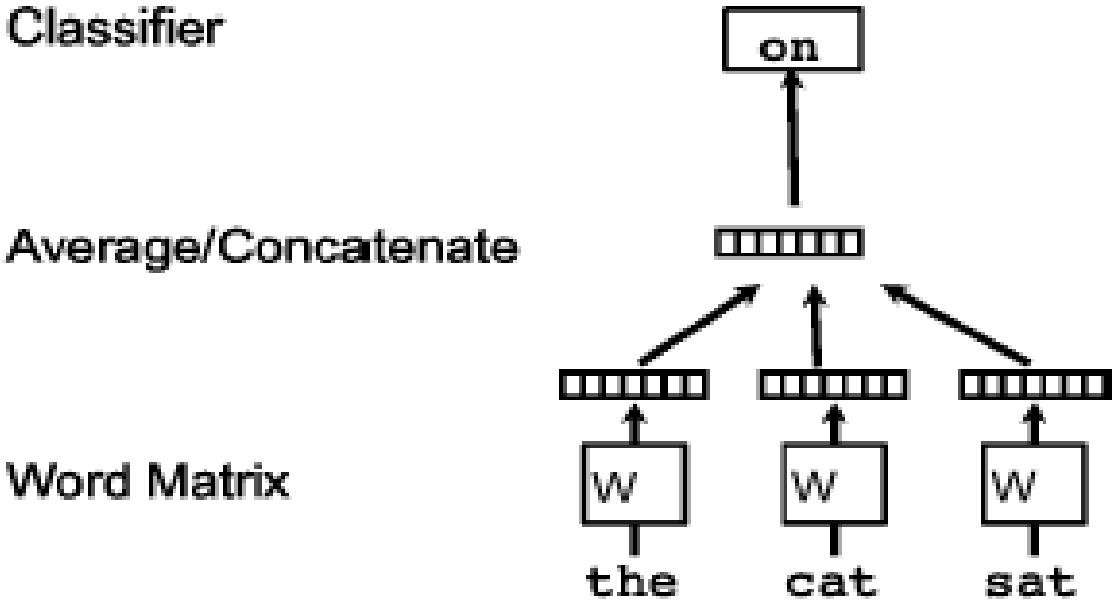
Aspect	Rating	Polarity
zoom	5	Positive
Price	4	Positive
Battery	2	Negative
screen	1	Negative
…	…	…

Background of study



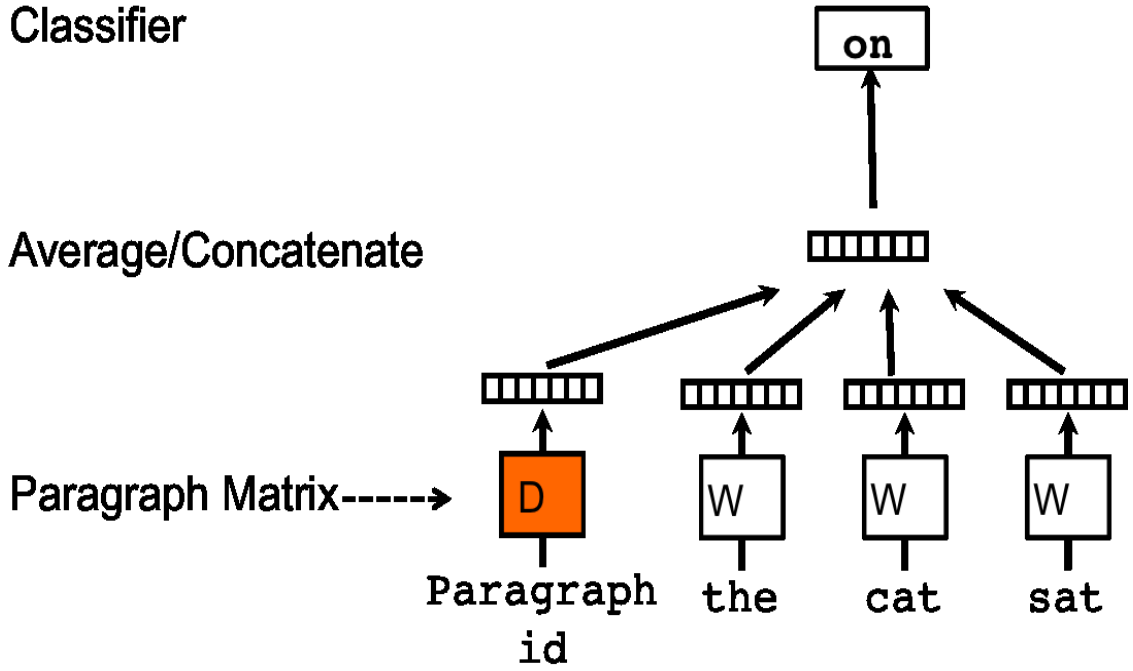
Word2Vec (uses Deep Learning)

- A framework for learning word vectors (Mikolov 2013)



Doc2Vec (uses Deep Learning)

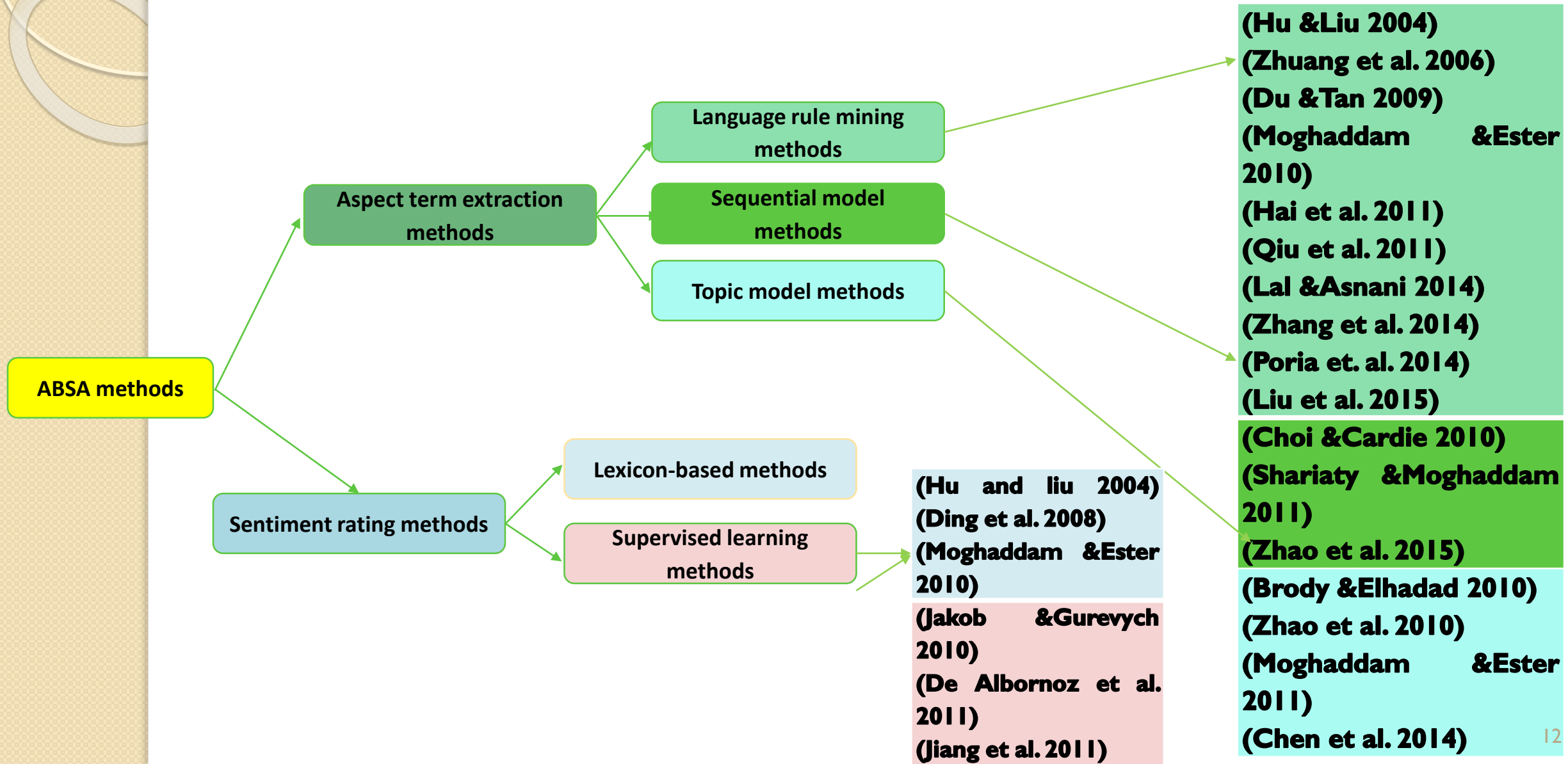
- A framework for learning paragraph vector (Mikolov 2014)



Deep learning

- Learn feature hierarchies, in which features from higher levels of the hierarchy are formed by the composition of lower level features.
- Automatically learning features at multiple levels of abstraction can be regarded as learning complex functions.
- Example-multilayer neural networks, which consists of several layers of non-linear operators for the composition of inputs from lower layers.

ABSA methods



ABSA methods cont...

- **Sequential models** are not suitable in this study due to their supervised nature which makes them domain dependent.
- **Topic models** are too statistic centric.
- Based on our literature review most of the works in ABSA use **language rule models** for this task.
- In this study we focus on language rule methods.

Language rule models

-Finds frequent nouns/noun phrases -Finds adjectives in a window of 5-6 words to create aspect –sentiment pair. • **Liu 2004** • **Ding 2008** • **Liu 2014** • **Lek 2013** • **Marrie 2014** Example: ***Image quality*** of this camera which I bought it in a reasonable price is ***excellent***	Uses dependency parser to find aspect – pair. • **Zhuang 2006** • **Wu 2009** • **Qiu 2011** • **Lizhen 2014**

ABSA challenges

Some of various challenges from the book of liu 2012 that make the problem of ABSA hard:

1. Using **different words** or phrases to express the same aspect, e.g.,
   - ***Photo quality*** is a little better than most of the cameras in this class.
   - That gives the SX40 better ***image quality***, especially in low light, experts say.

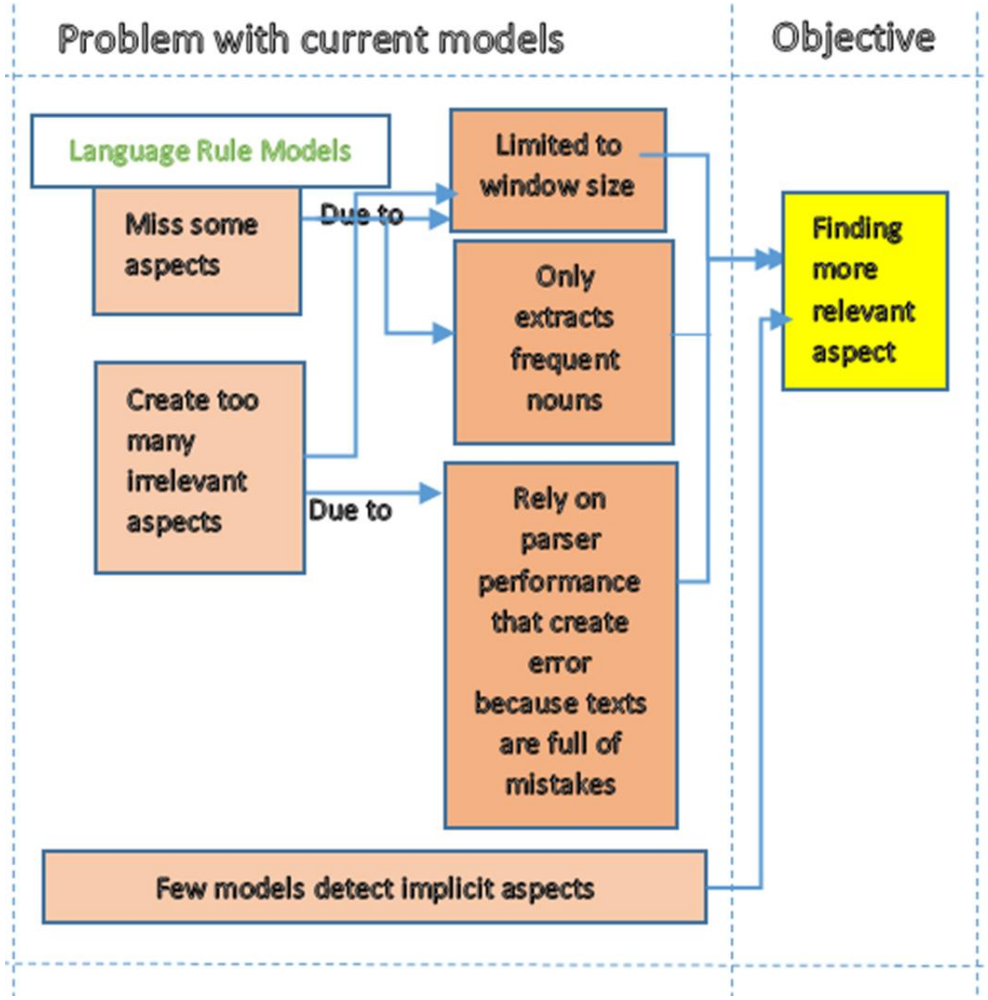
2. using **different sentiments** for expressing the same polarity, e.g.,
   - For a camera of this price, the picture quality is ***amazing***.
   - I am going on a trip to France and wanted something that could take ***stunning*** pictures with, but didn't cost a small fortune.

ABSA challenges…Cont.

3. Reviews include a large amount of **irrelevant information**.
   - I have owned Canon power shot pocket cameras exclusively over the years.
   - I have fat hands but short fingers.

4. While **explicit** aspect/sentiment extraction is easy, extracting **implicit** ones is difficult.
   - This mp3 player is very **affordable**.
   - I **bought** this mp3 for almost nothing!
   - After a twenty-one mile bike ride a four mile backpacking river hike, the size, weight, and performance of this camera ***has been the answer to my needs***.
   - The grip and *weight* ***make it easy to handle*** and the mid zoom pictures have exceeded expectation

Aspect extraction problem



Problem example

- Results are **limited to the window size**.

  Example: ***Image quality*** of this camera which I bought it is in a reasonable price is *excellent*.

- Reviews are **full of irrelevant information**.

  Example:"I have fat ***hand*** and short ***fingers***".

Problem example

- **Explicit aspect**

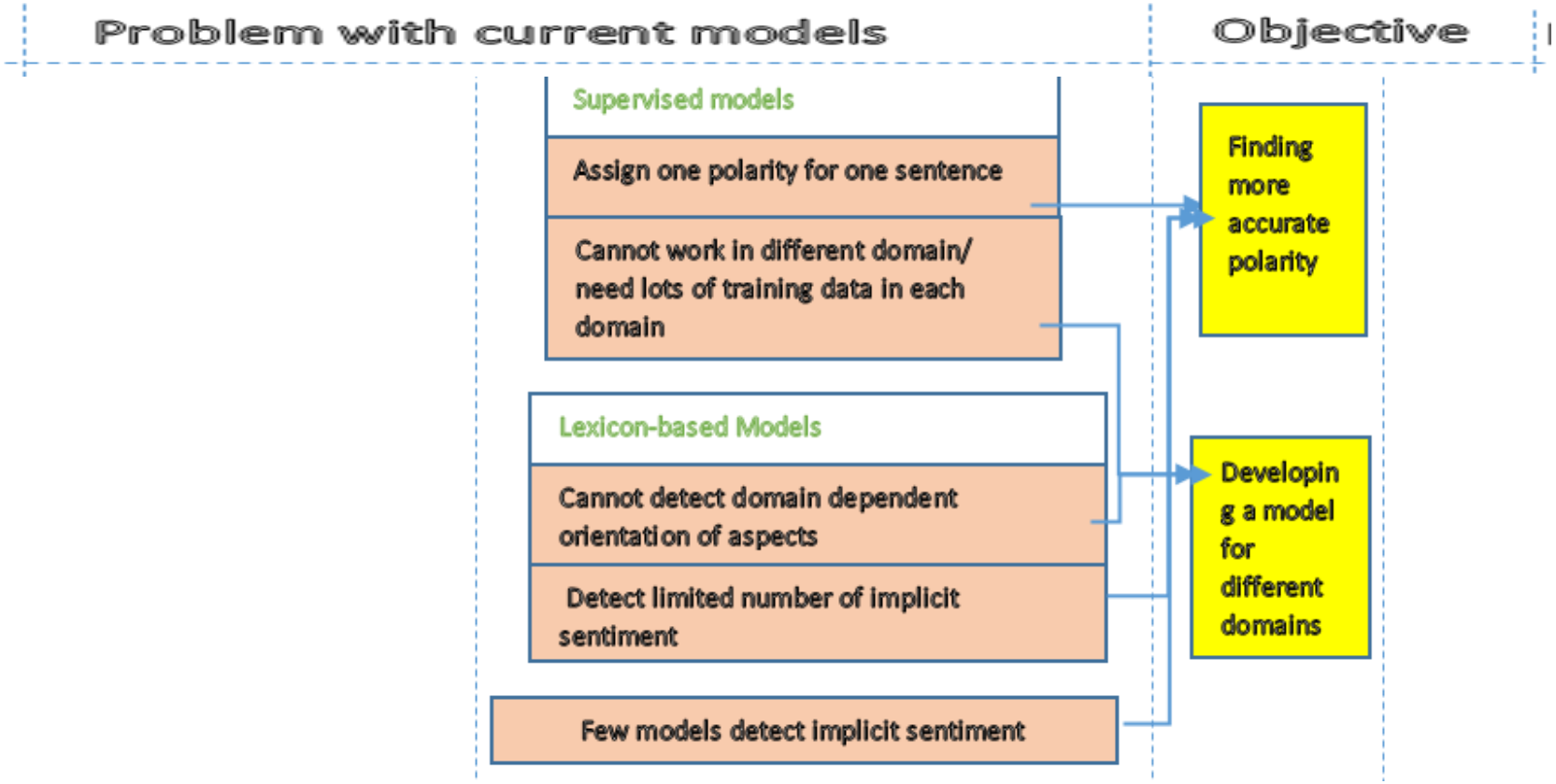
Example: "The ***picture quality*** of this phone is great".

- **Implicit aspect**

Example: "This car is so expensive." → '***price***'

Example: "This phone will not easily fit in a pocket". → '***size***'

Polarity prediction problem



Problem Example

- **More than one aspect in a sentence**

Example: The **image quality** is good but the **price** is high.

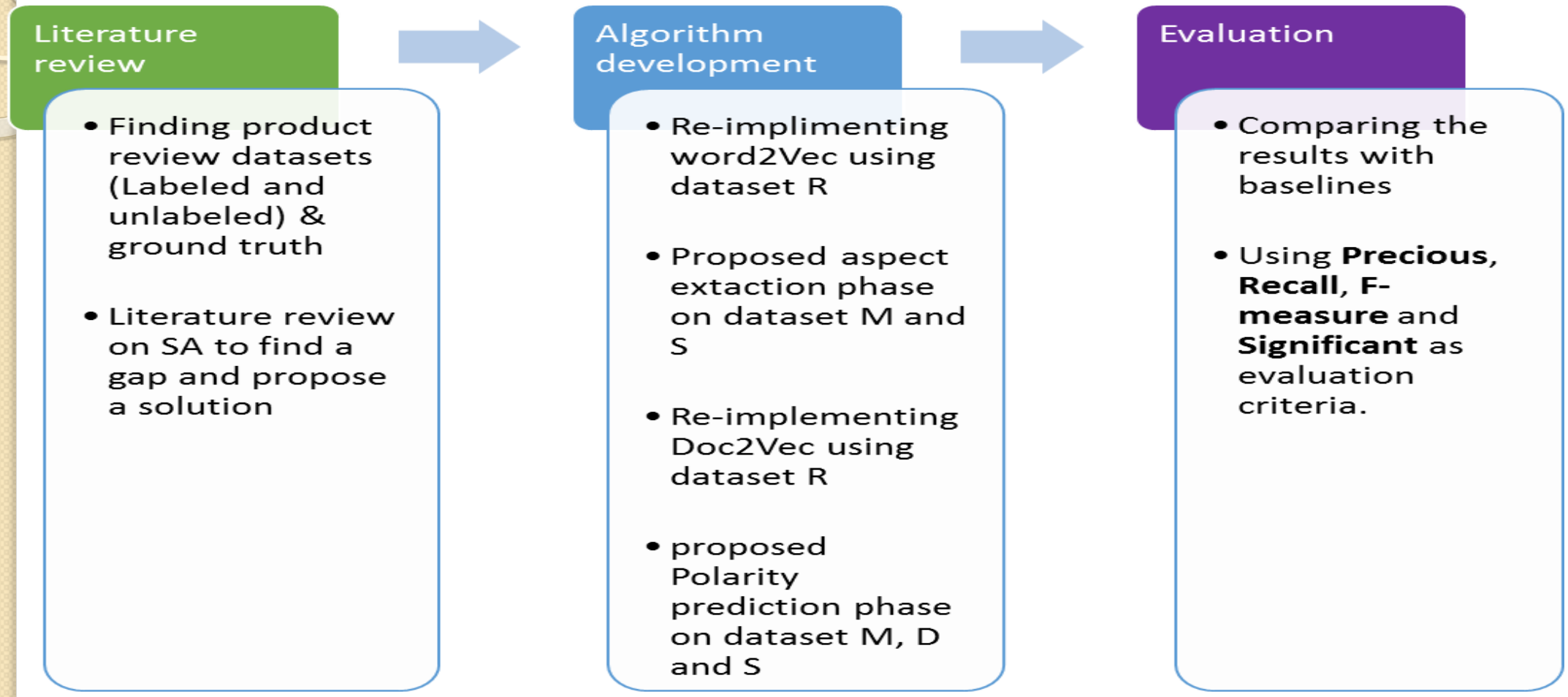
- **Domain dependent orientation** of opinions

Example: ***quite*** is positive in restaurant domain but negative in MP3 player domain.

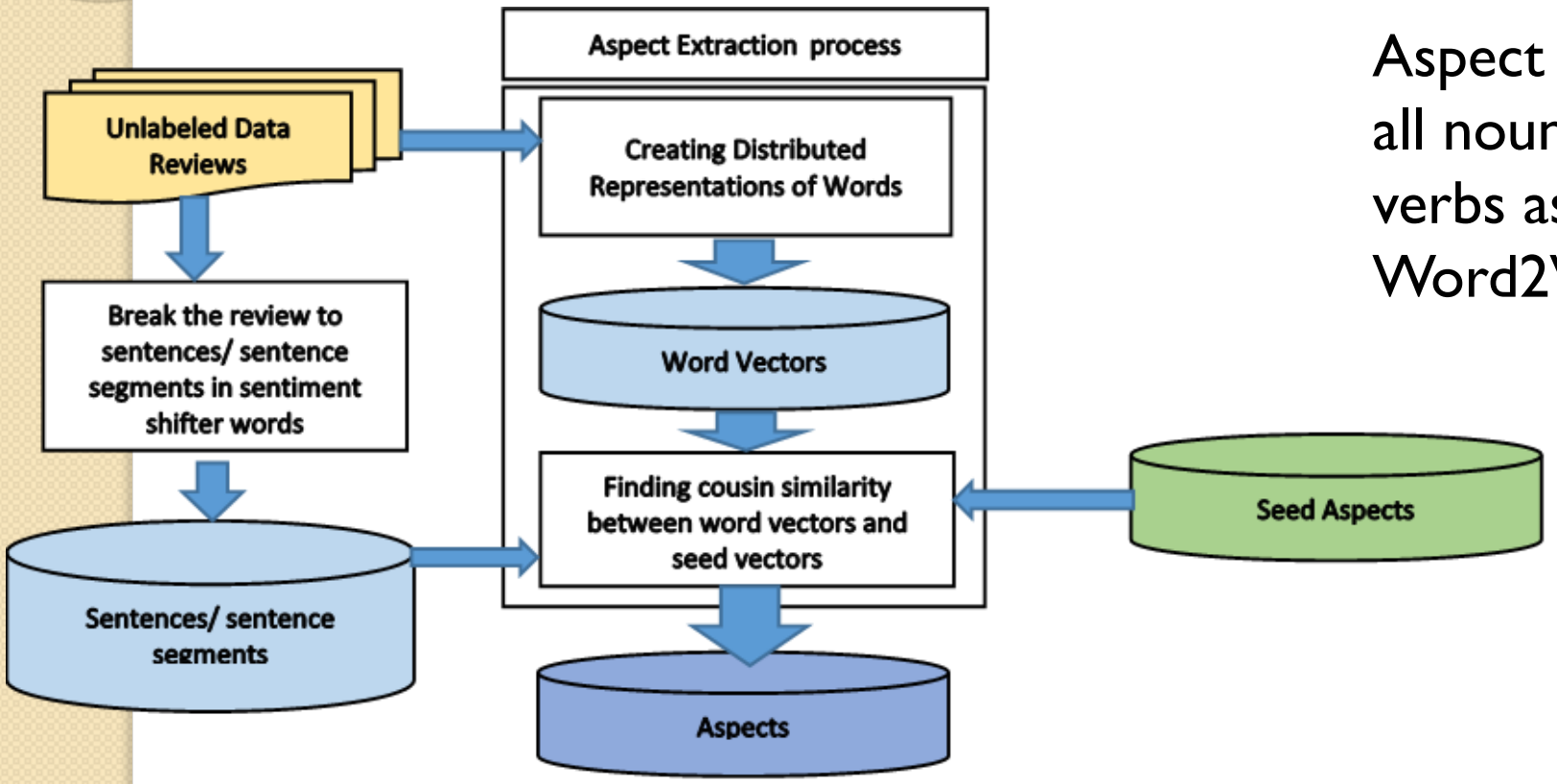
- **Explicit sentiment**

Example: After a twenty-one mile bike ride a four mile backpacking river hike, the size, weight, and performance of this camera ***has been the answer to my needs***.

Research framework

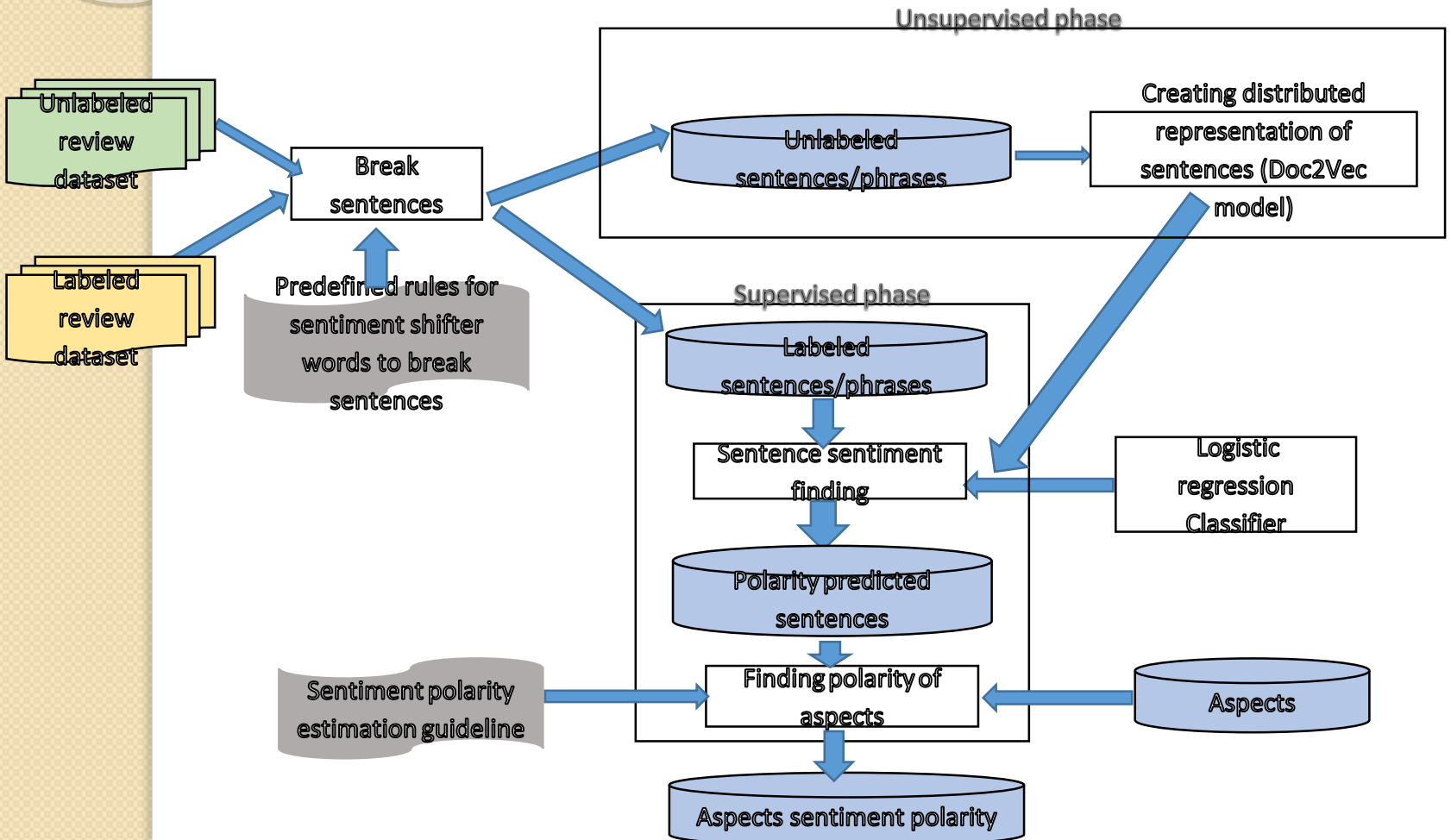


(Our Proposed method) aspect extraction



Aspect extraction-----→ extract all nouns, adjectives, adverbs and verbs as candidate aspect+ uses Word2Vec (mikolov 2013)

(Our Proposed method) sentiment rating



Sentiment rating-----→uses Doc2Vec + (Hu and Liu 2004)

Doc2Vec as unsupervised phase and Logistic regression as supervised phase(mikolov 2014)

Conclusions

The proposed improvments are valuable:-

For Consumers:

- Easing the process of decision making when purchase products or services by providing a decomposed view of rated aspects

For producers:

- Source of consumer feedback.
- Benchmark products and services
- Save lots of money they spend to obtain consumer opinions, using survey, focus group and consultants.

For other systems:

- Opinion summarization systems
- Opinion question answering systems
- Recommendation systems (to provide explanations for recommendation)
- Advertising system (to place an ad of a product with similar rated aspects)
- Many business tasks related to sale management, reputation management, and public relations


Thank You!