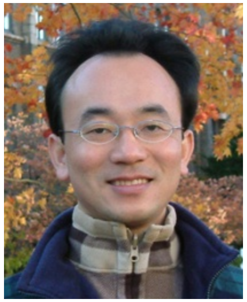
Tamkang University
淡江大學

Social Computing and Big Data Analytics
社群運算與大數據分析

Tamkang University

Social Media Marketing Analytics
(社群媒體行銷分析)

1052SCBDA08
MIS MBA (M2226) (8606)
Wed, 8,9, (15:10-17:00) (L206)



Min-Yuh Day
戴敏育
Assistant Professor
專任助理教授
Dept. of Information Management, Tamkang University
淡江大學 資訊管理學系
http://mail. tku.edu.tw/myday/

2017-04-12

課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
1	2017/02/15	Course Orientation for Social Computing and Big Data Analytics (社群運算與大數據分析課程介紹)
2	2017/02/22	Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data (資料科學與大數據分析：探索、分析、視覺化與呈現資料)
3	2017/03/01	Fundamental Big Data: MapReduce Paradigm, Hadoop and Spark Ecosystem (大數據基礎：MapReduce典範、Hadoop與Spark生態系統)

課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
4	2017/03/08	Big Data Processing Platforms with SMACK: Spark, Mesos, Akka, Cassandra and Kafka (大數據處理平台SMACK： Spark, Mesos, Akka, Cassandra, Kafka)
5	2017/03/15	Big Data Analytics with Numpy in Python (Python Numpy 大數據分析)
6	2017/03/22	Finance Big Data Analytics with Pandas in Python (Python Pandas 財務大數據分析)
7	2017/03/29	Text Mining Techniques and Natural Language Processing (文字探勘分析技術與自然語言處理)
8	2017/04/05	Off-campus study (教學行政觀摩日)

課程大綱 (Syllabus)

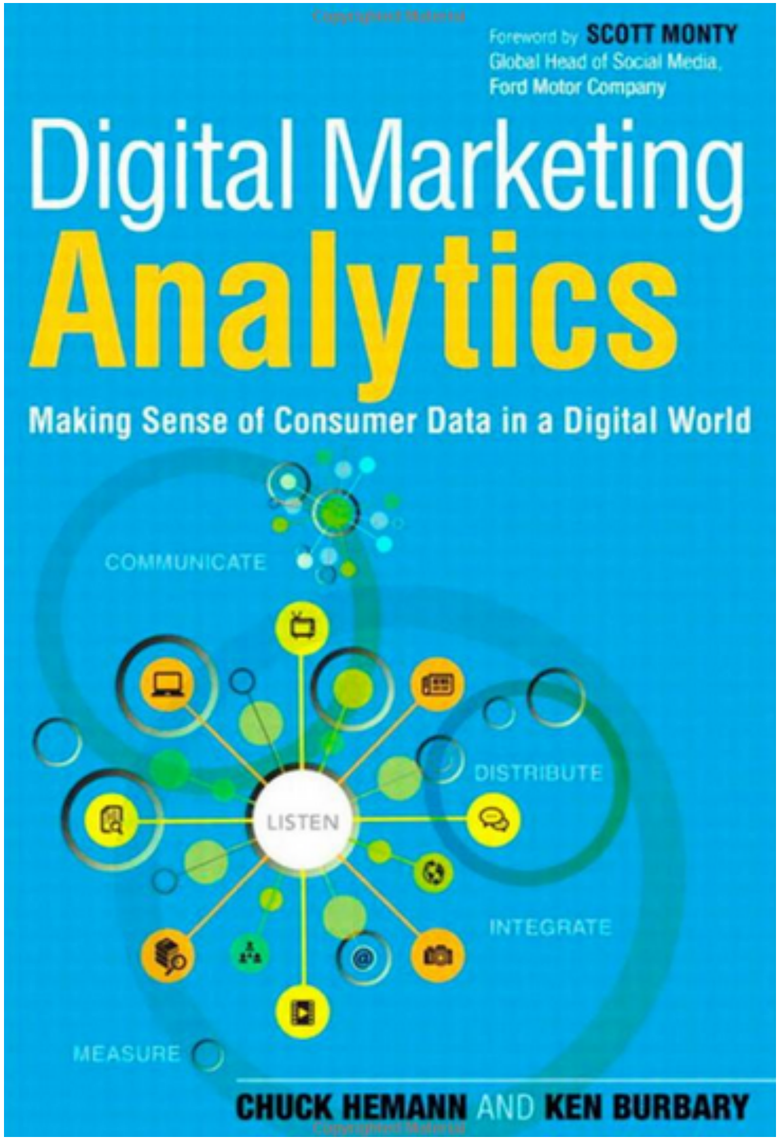
週次 (Week)	日期 (Date)	內容 (Subject/Topics)
9	2017/04/12	Social Media Marketing Analytics (社群媒體行銷分析)
10	2017/04/19	期中報告 (Midterm Project Report)
11	2017/04/26	Deep Learning with Theano and Keras in Python (Python Theano 和 Keras 深度學習)
12	2017/05/03	Deep Learning with Google TensorFlow (Google TensorFlow 深度學習)
13	2017/05/10	Sentiment Analysis on Social Media with Deep Learning (深度學習社群媒體情感分析)

課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
14	2017/05/17	Social Network Analysis (社會網絡分析)
15	2017/05/24	Measurements of Social Network (社會網絡量測)
16	2017/05/31	Tools of Social Network Analysis (社會網絡分析工具)
17	2017/06/07	Final Project Presentation I (期末報告 I)
18	2017/06/14	Final Project Presentation II (期末報告 II)

Social Media Marketing Analytics

Digital Marketing Analytics:
Making Sense of Consumer Data in a Digital World, Chuck Hemann and Ken Burbary, Que. 2013



Source: http://www.amazon.com/Digital-Marketing-Analytics-Consumer-Biz-Tech/dp/0789750309

Consumer Psychology and Behavior on Social Media

Marketing

"Meeting needs profitably"

Source: Philip Kotler & Kevin Lane Keller, Marketing Management, 14th ed., Pearson, 2012

How consumers **think, feel, and act**

Analyzing Consumer Markets

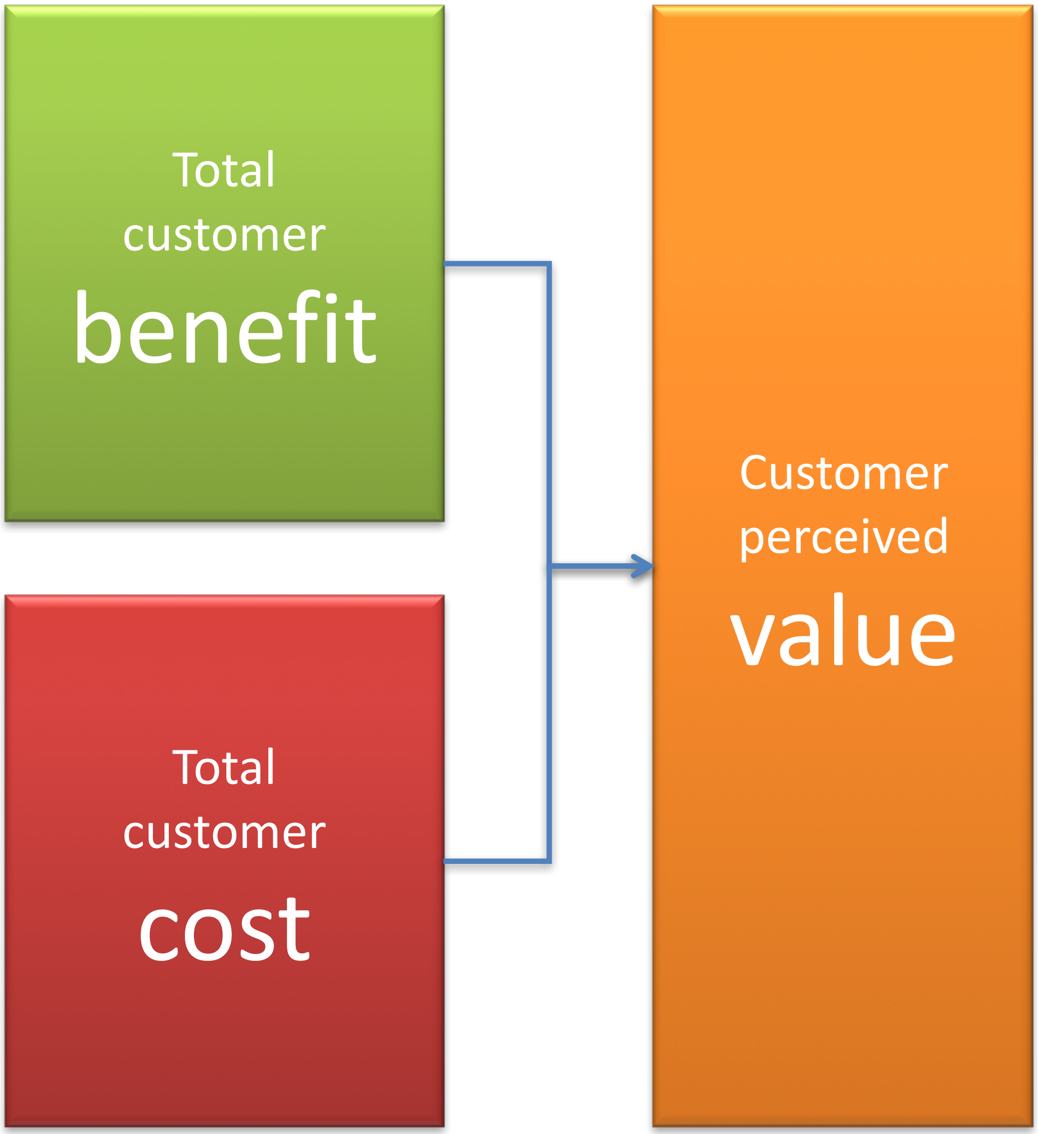
- The aim of marketing is to meet and satisfy target customers' needs and wants better than competitors.
- Marketers must have a thorough understanding of how consumers think, feel, and act and offer clear value to each and every target consumer.

Value

the sum of the tangible and intangible benefits and costs

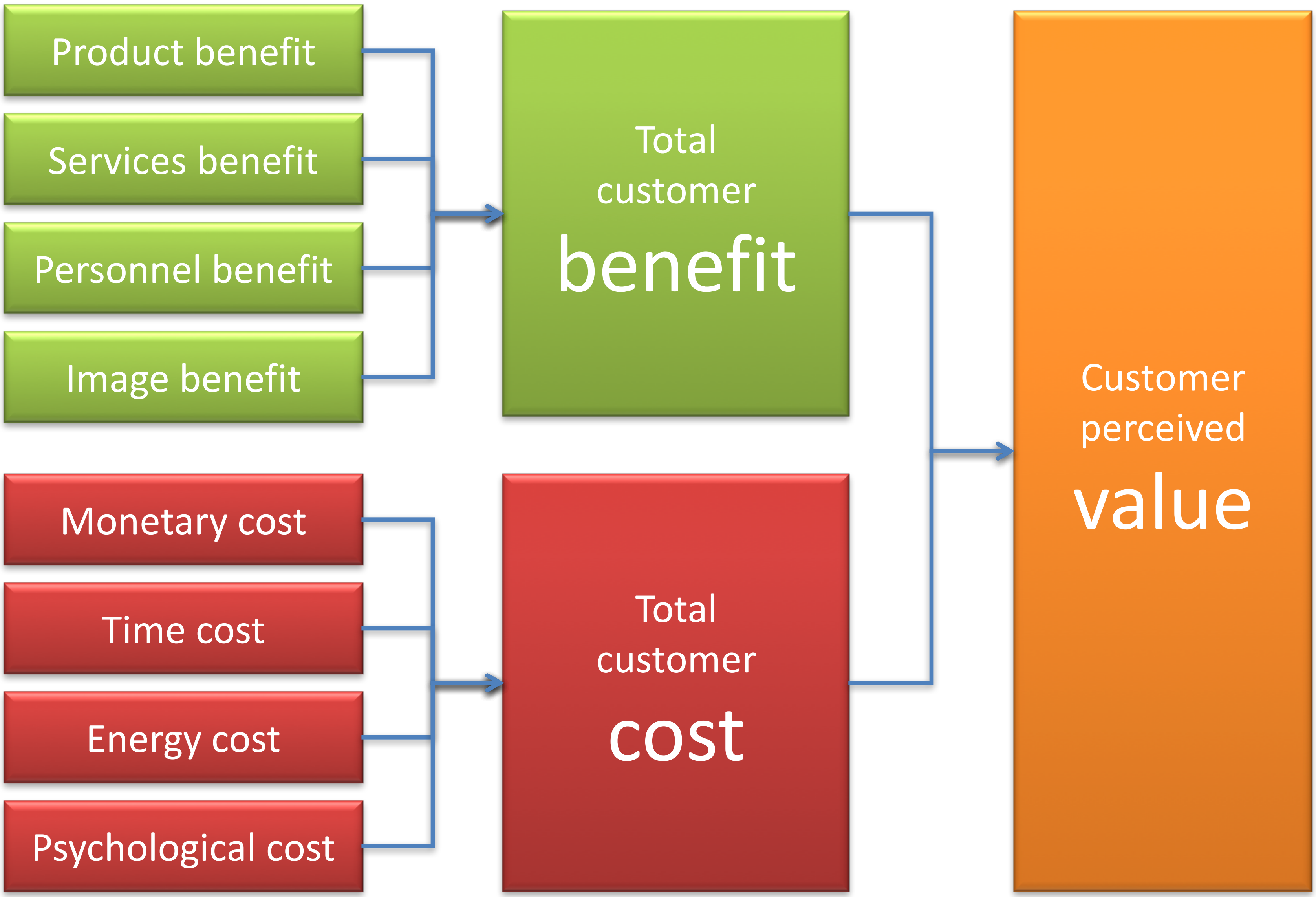
Source: Philip Kotler & Kevin Lane Keller, Marketing Management, 14th ed., Pearson, 2012

Value



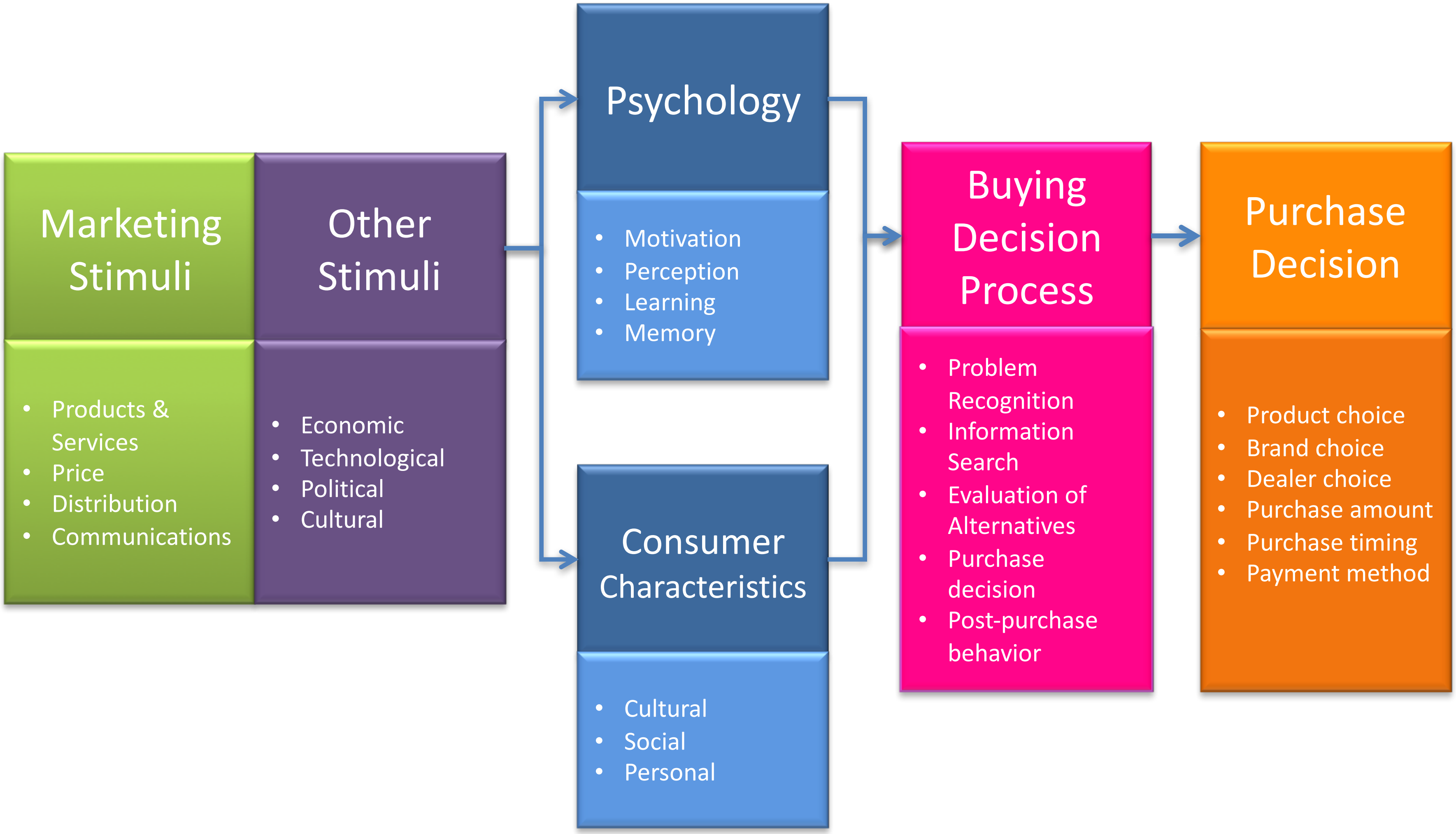
Source: Philip Kotler & Kevin Lane Keller, Marketing Management, 14th ed., Pearson, 2012

Customer Perceived Value



Source: Philip Kotler & Kevin Lane Keller, Marketing Management, 14th ed., Pearson, 2012

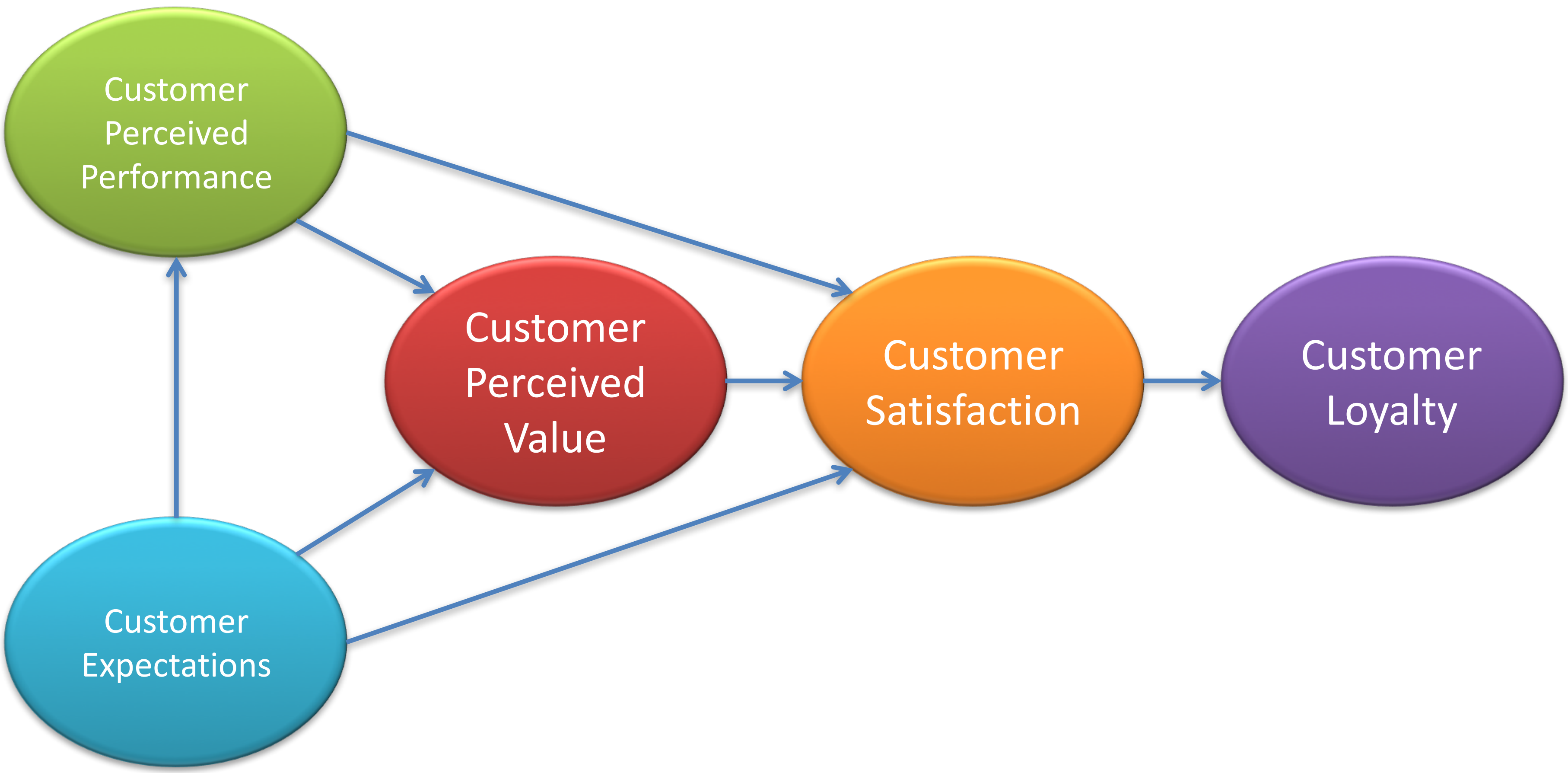
Model of Consumer Behavior



Source: Philip Kotler & Kevin Lane Keller, Marketing Management, 14th ed., Pearson, 2012

Building Customer Value, Satisfaction, and Loyalty

Customer Perceived Value, Customer Satisfaction, and Loyalty



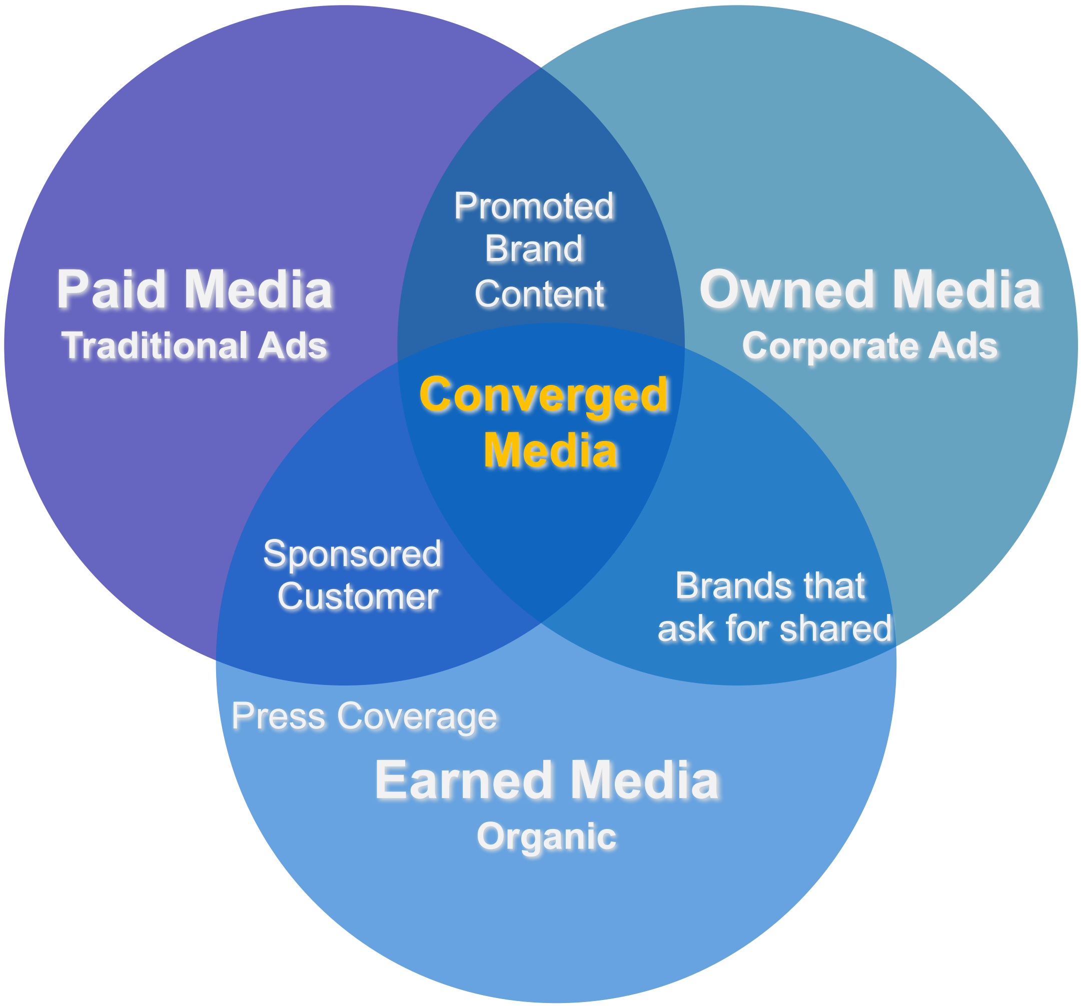
Source: Philip Kotler & Kevin Lane Keller, Marketing Management, 14th ed., Pearson, 2012

Social Media Marketing Analytics

- Social Media Listening
- Search Analytics
- Content Analytics
- Engagement Analytics

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

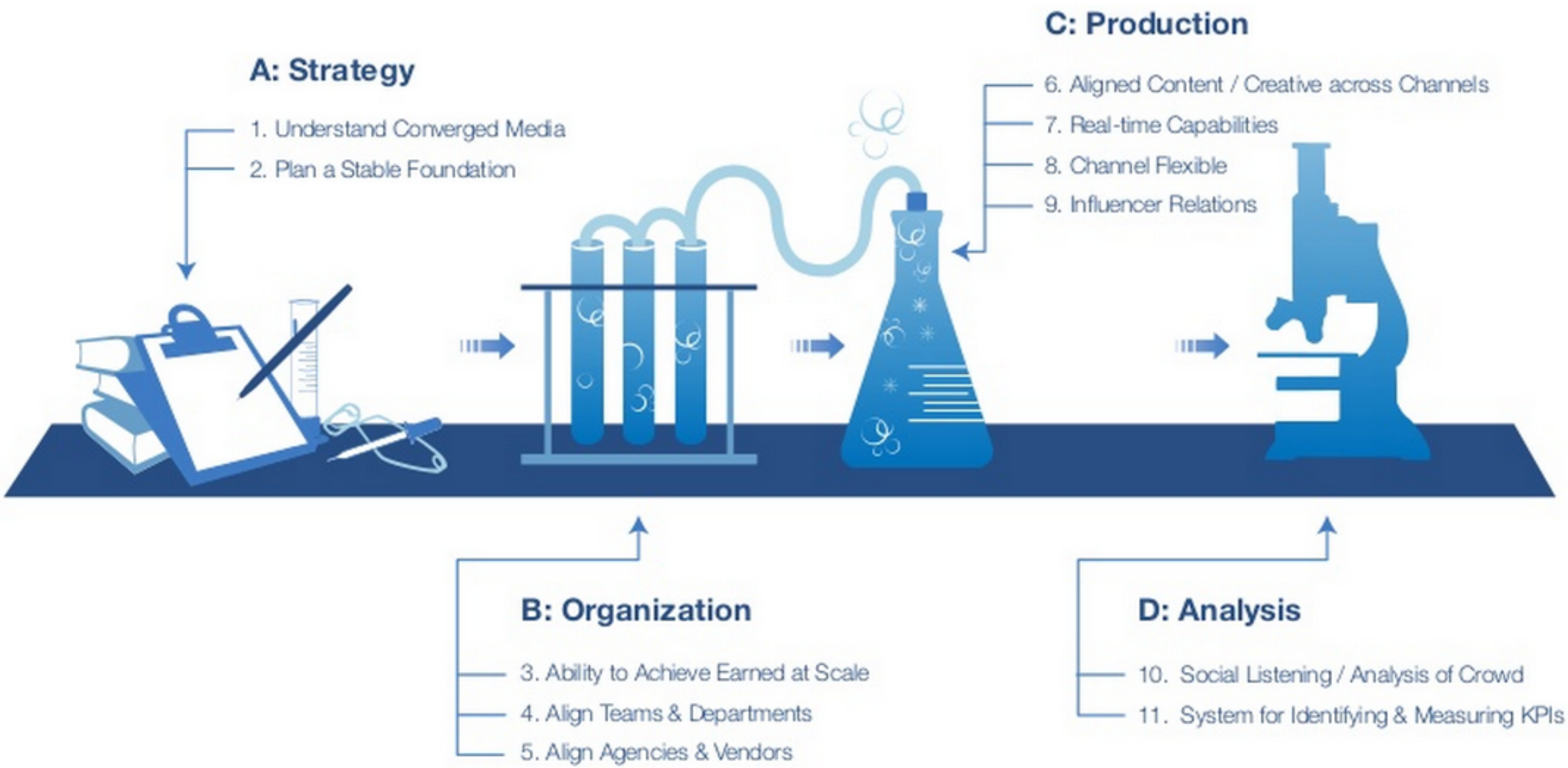
The Convergence of Paid, Owned & Earned Media



Source: "The Converged Media Imperative: How Brands Will Combine Paid, Owned and Earned Media", Altimeter Group, July 19, 2012)
http://www.altimetergroup.com/2012/07/the-converged-media-imperative/

Converged Media
Top 11 Success Criteria

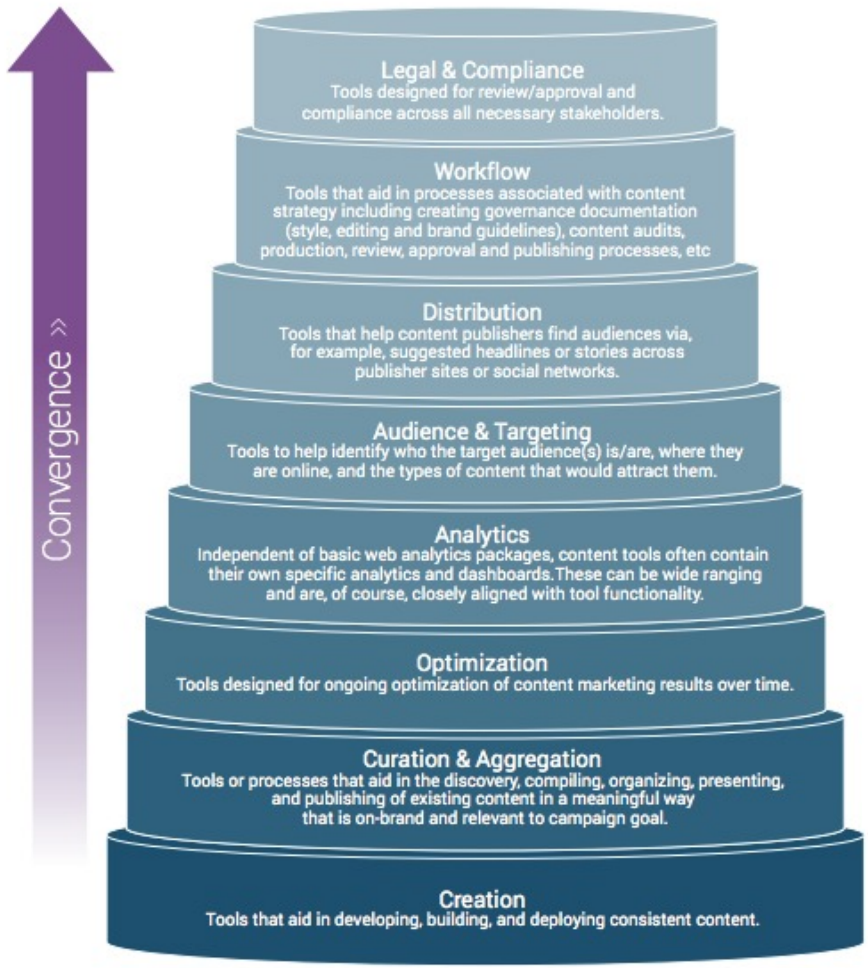
Social Listening / Analysis of Crowd



Source: "The Converged Media Imperative: How Brands Will Combine Paid, Owned and Earned Media", Altimeter Group, July 19, 2012)
http://www.altimetergroup.com/2012/07/the-converged-media-imperative/

Content Tool Stack Hierarchy

Figure 3 Content Tool Stack Hierarchy



Source: Altimeter Group

Source: Rebecca Lieb, "Content marketing in 2015 -- research, not predictions", December 16, 2014
http://www.imediaconnection.com/content/37909.asp

Competitive Intelligence

- Gather competitive intelligence data

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Google Alexa Compete

- Which audience segments are competitors reaching that you are not?
- What keywords are successful for your competitors?
- What sources are driving traffic to your competitors’ websites?

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Competitive Intelligence

- Facebook competitive analysis
- Facebook content analysis
- YouTube competitive analysis
- YouTube channel analysis
- Twitter profile analysis

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Web Analytics (Clickstream)

- Content Analytics
- Mobile Analytics

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Mobile Analytics

- Where is my mobile traffic coming from?
- What content are mobile users most interested in?
- How is my mobile app being used?
  What’s working?
  What isn’t?
- Which mobile platforms work best with my site?
- How does mobile user’s engagement with my site compare to traditional web users’ engagement?

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Identifying a Social Media Listening Tool

- Data Capture
- Spam Prevention
- Integration with Other Data Sources
- Cost
- Mobile Capability
- API Access
- Consistent User Interface
- Workflow Functionality
- Historical Data

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Search Analytics

- Free Tools for Collecting Insights Through
  - Search Data
  - Google Trends
  - YouTube Trends
  - The Google AdWords Keyword Tool
  - Yahoo! Clues
- Paid Tools for Collecting Insights Through Search Data
- The BrightEdge SEO Platform

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Owned Social Metrics

- Facebook page
- Twitter account
- YouTube channel

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Own Social Media Metrics: Facebook

- Total likes
- Reach
  - Organic
  - Paid reach
  - Viral reach
- Engaged users
- People taking about this (PTAT)
- Likes, comments, and shares by post

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Own Social Media Metrics: Twitter

- Followers
- Retweets
- Replies
- Clicks and click-through rate (CTR)
- Impressions

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Own Social Media Metrics: YouTube

- Views
- Subscribers
- Likes/dislikes
- Comments
- Favorites
- Sharing

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Own Social Media Metrics: SlideShare

- Followers
- Views
- Comments
- Shares

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Own Social Media Metrics: Pinterest

- Followers
- Number of boards
- Number of pins
- Likes
- Repins
- Comments

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Own Social Media Metrics: Google+

- Number of people who have an account circled
- +1s
- Comments

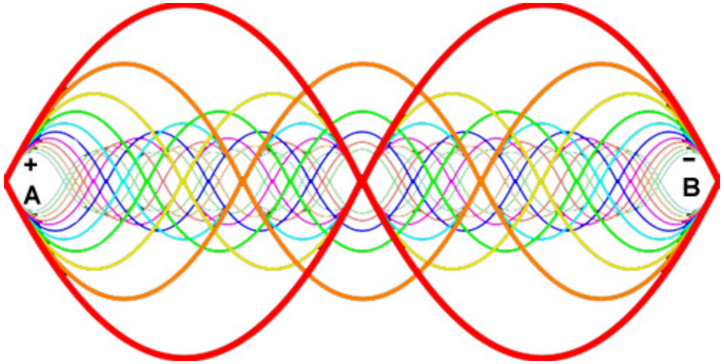
Earned Social Media Metrics

- Earned conversations
- In-network conversations

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Earned Social Media Metrics: Earned conversations

- Share of voice
- Share of conversation
- Sentiment
- Message resonance
- Overall conversation volume



Source: http://www.elvtd.com/elevation/p/beings-of-resonance

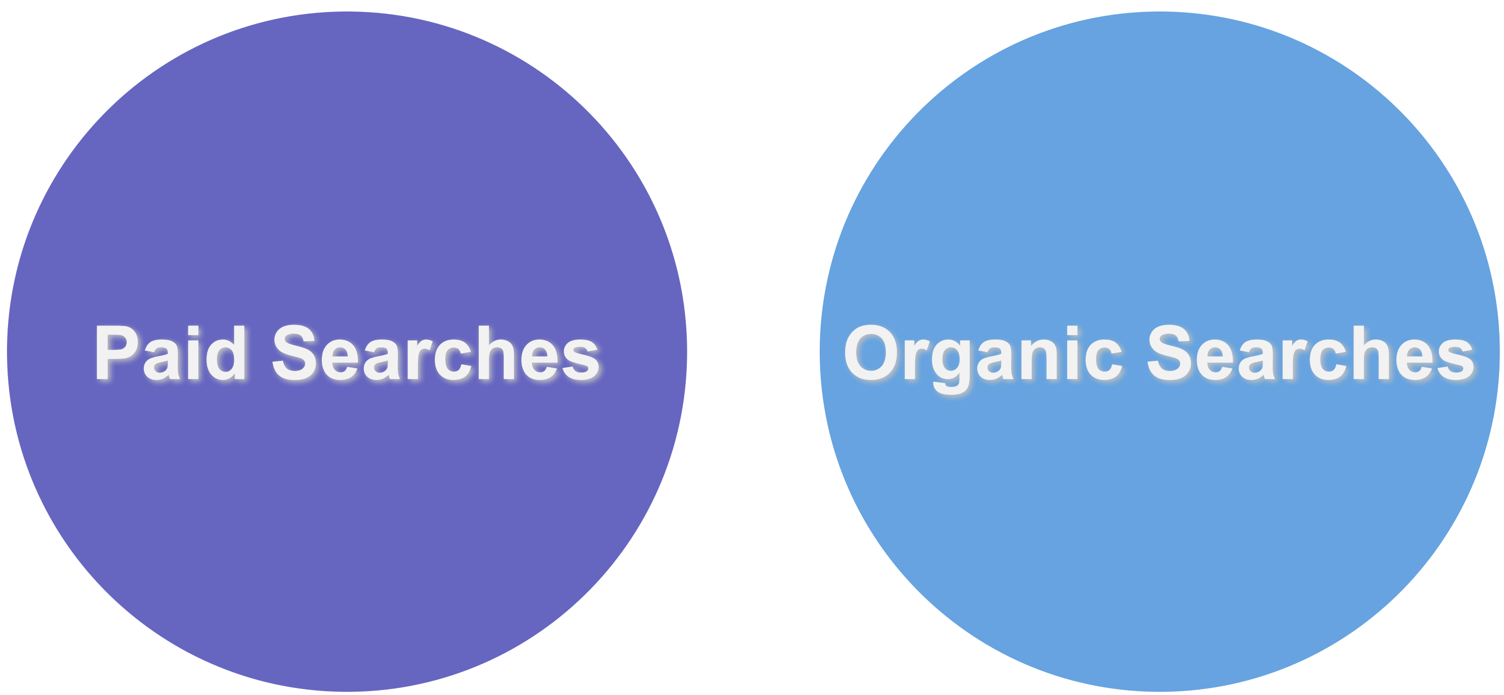
Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Demystifying Web Data

- Visits
- Unique page views
- Bounce rate
- Pages per visit
- Traffic sources
- Conversion

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Searching for the Right Metrics



Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Paid Searches

- Impressions
- Clicks
- Click-through rate (CTR)
- Cost per click (CPC)
- Impression share
- Sales or revenue per click
- Average position

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Organic Searches

- Known and unknown keywords
- Known and unknown branded keywords
- Total visits
- Total conversions from known keywords
- Average search position

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Aligning Digital and Traditional Analytics

- Primary Research
  - Brand reputation
  - Message resonance
  - Executive reputation
  - Advertising performance
- Traditional Media Monitoring
- Traditional CRM Data

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Social Media Listening Evolution

- Location of conversations
- Sentiment
- Key message penetration
- Key influencers

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Social Analytics Lifecycle (5 Stages)

1. Discover

2. Analyze

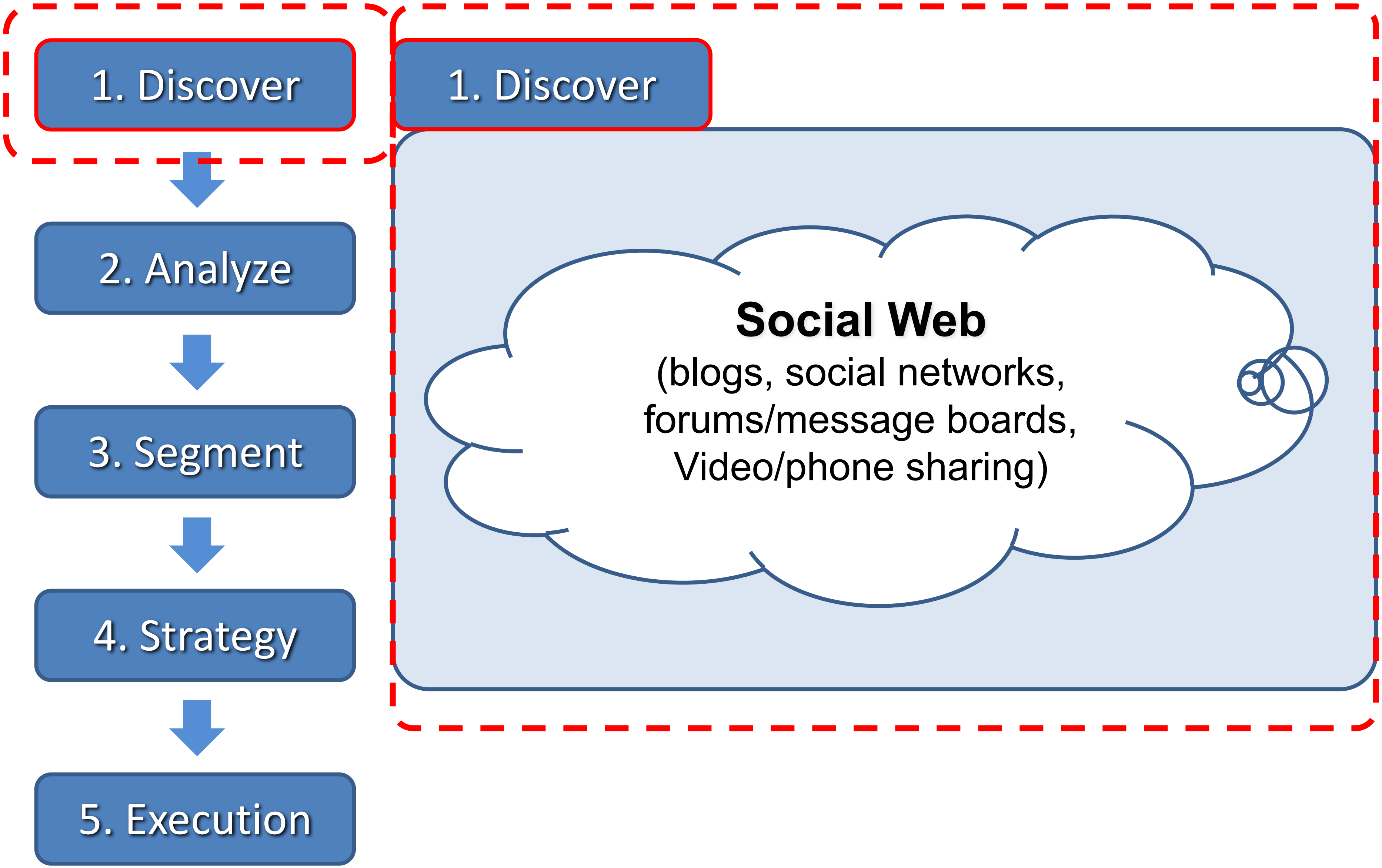
3. Segment

4. Strategy

5. Execution

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Social Analytics Lifecycle (5 Stages)



Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Social Analytics Lifecycle (5 Stages)

1. Discover

2. Analyze

3. Segment

4. Strategy

5. Execution

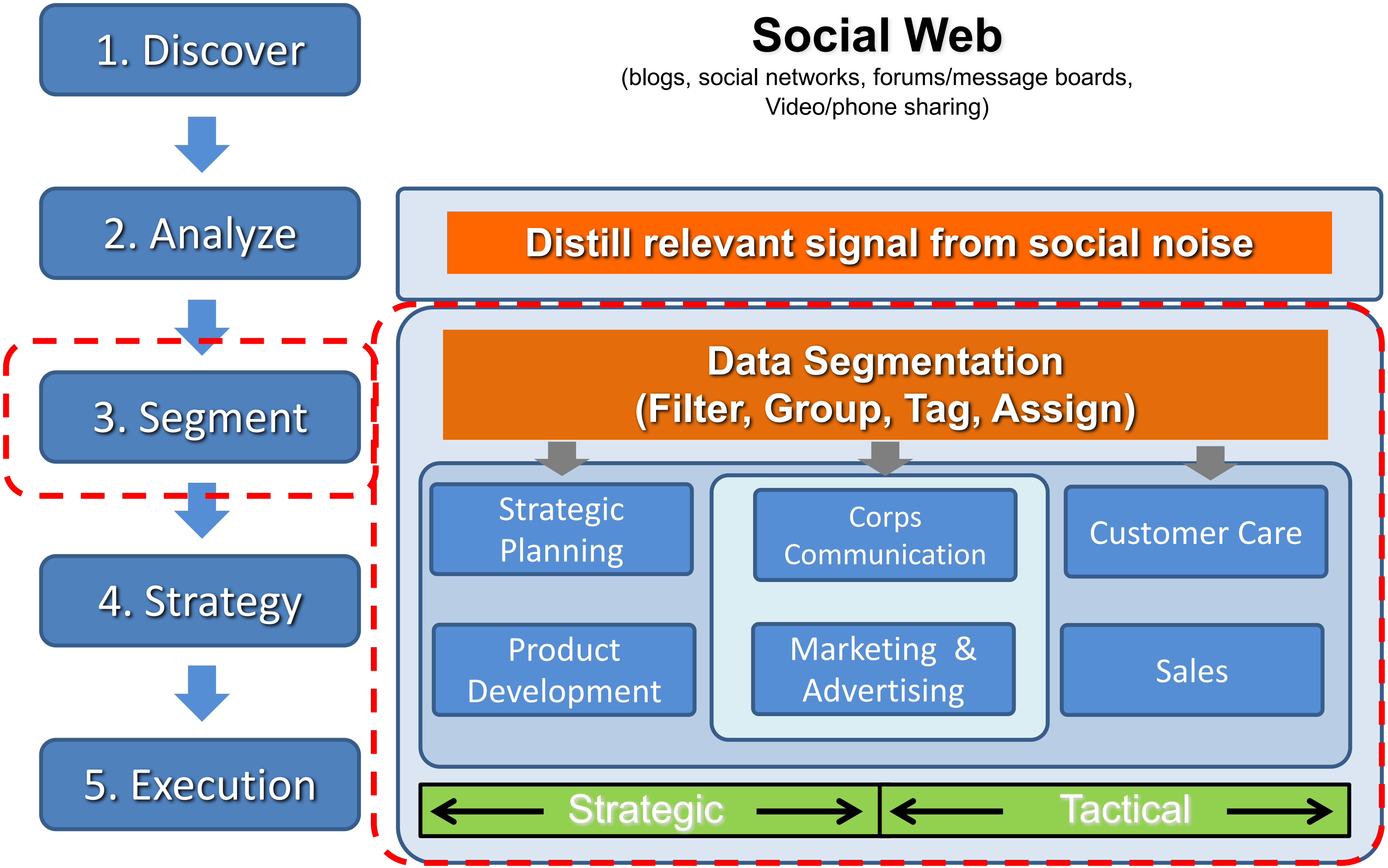
Social Web

(blogs, social networks, forums/message boards, Video/phone sharing)

Distill relevant signal from social noise

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Social Analytics Lifecycle (5 Stages)



Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Social Analytics Lifecycle (5 Stages)

1. Discover
2. Analyze
3. Segment
4. Strategy
5. Execution

Social Web

(blogs, social networks, forums/message boards, Video/phone sharing)

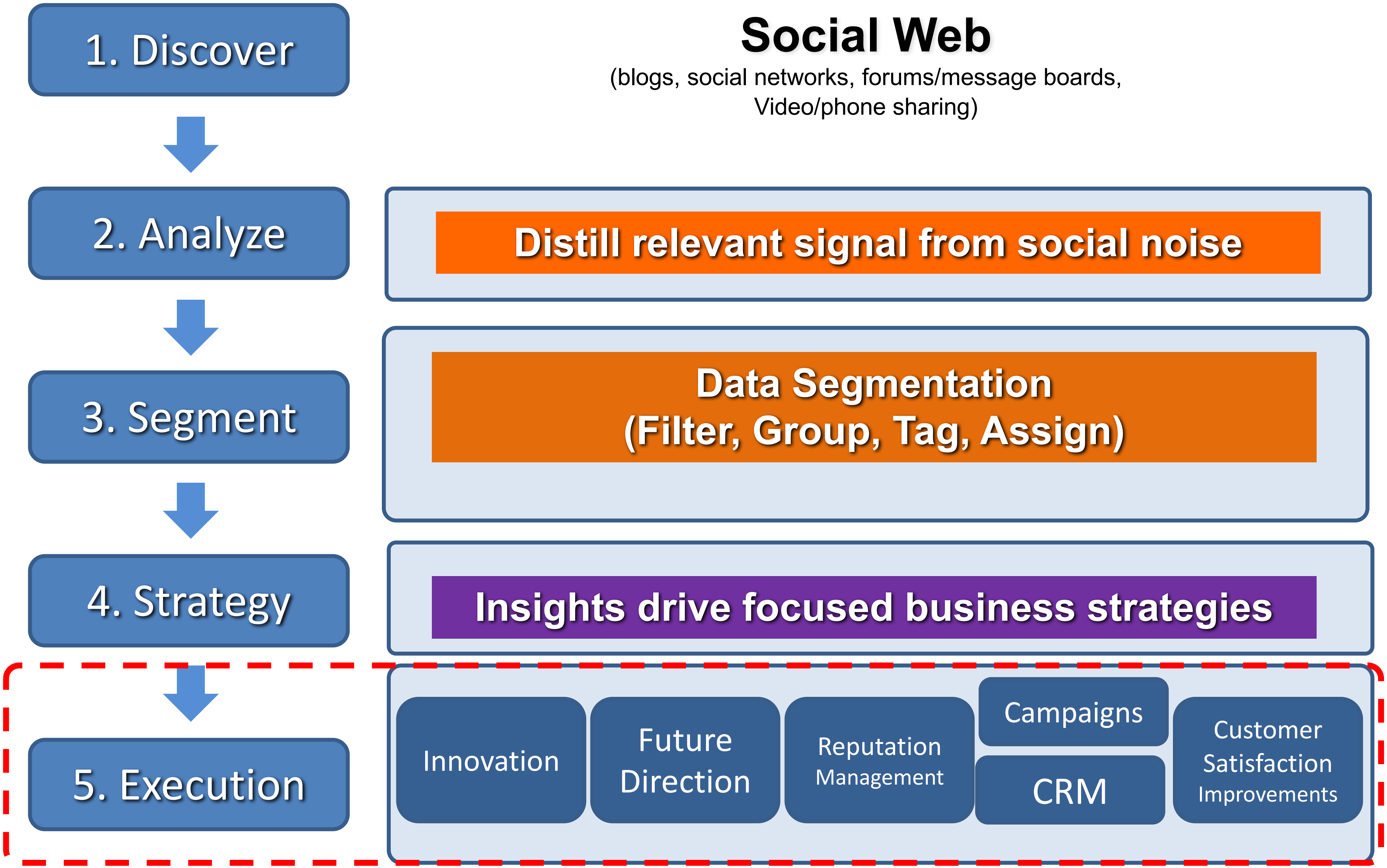
Distill relevant signal from social noise

Data Segmentation (Filter, Group, Tag, Assign)

Insights drive focused business strategies

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Social Analytics Lifecycle (5 Stages)



Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Social Analytics Lifecycle (5 Stages)

1. Discover
2. Analyze
3. Segment
4. Strategy
5. Execution

Social Web
(blogs, social networks, forums/message boards, Video/phone sharing)

Distill relevant signal from social noise

Data Segmentation (Filter, Group, Tag, Assign)

Insights drive focused business strategies

Innovation

Future Direction

Reputation Management

Campaigns

CRM

Customer Satisfaction Improvements

Source: Chuck Hemann and Ken Burbary, Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que. 2013

Social Media

Source: http://hungrywolfmarketing.com/2013/09/09/what-are-your-social-marketing-goals/

Emotions

Love

Joy

Surprise

Anger

Sadness

Fear

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Example of Opinion: review segment on iPhone

"I bought an iPhone a few days ago.

It was such a nice phone.

The touch screen was really cool.

The voice quality was clear too.

However, my mother was mad with me as I did not tell her before I bought it.

She also thought the phone was too expensive, and wanted me to return it to the shop. … "

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Example of Opinion: review segment on iPhone

"(1) I bought an iPhone a few days ago.

(2) It was such a **nice** phone.

(3) The touch screen was really **cool**.

(4) The voice quality was **clear** too.

(5) However, my mother was mad with me as I did not tell her before I bought it.

(6) She also thought the phone was too **expensive**, and wanted me to return it to the shop. … "

+Positive Opinion

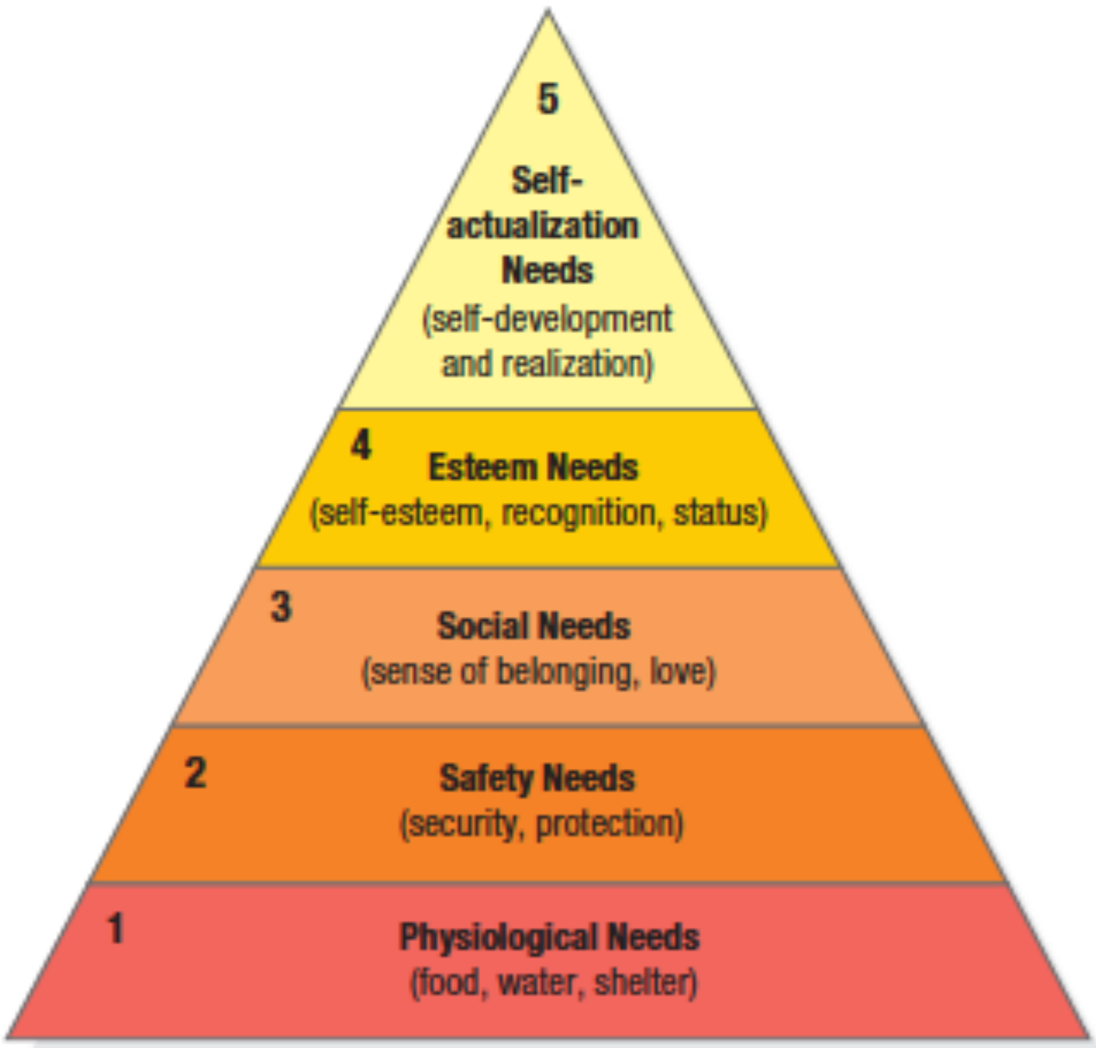
-Negative Opinion

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

How consumers think, feel, and act

Source: Philip Kotler & Kevin Lane Keller, Marketing Management, 14th ed., Pearson, 2012

Maslow’s Hierarchy of Needs



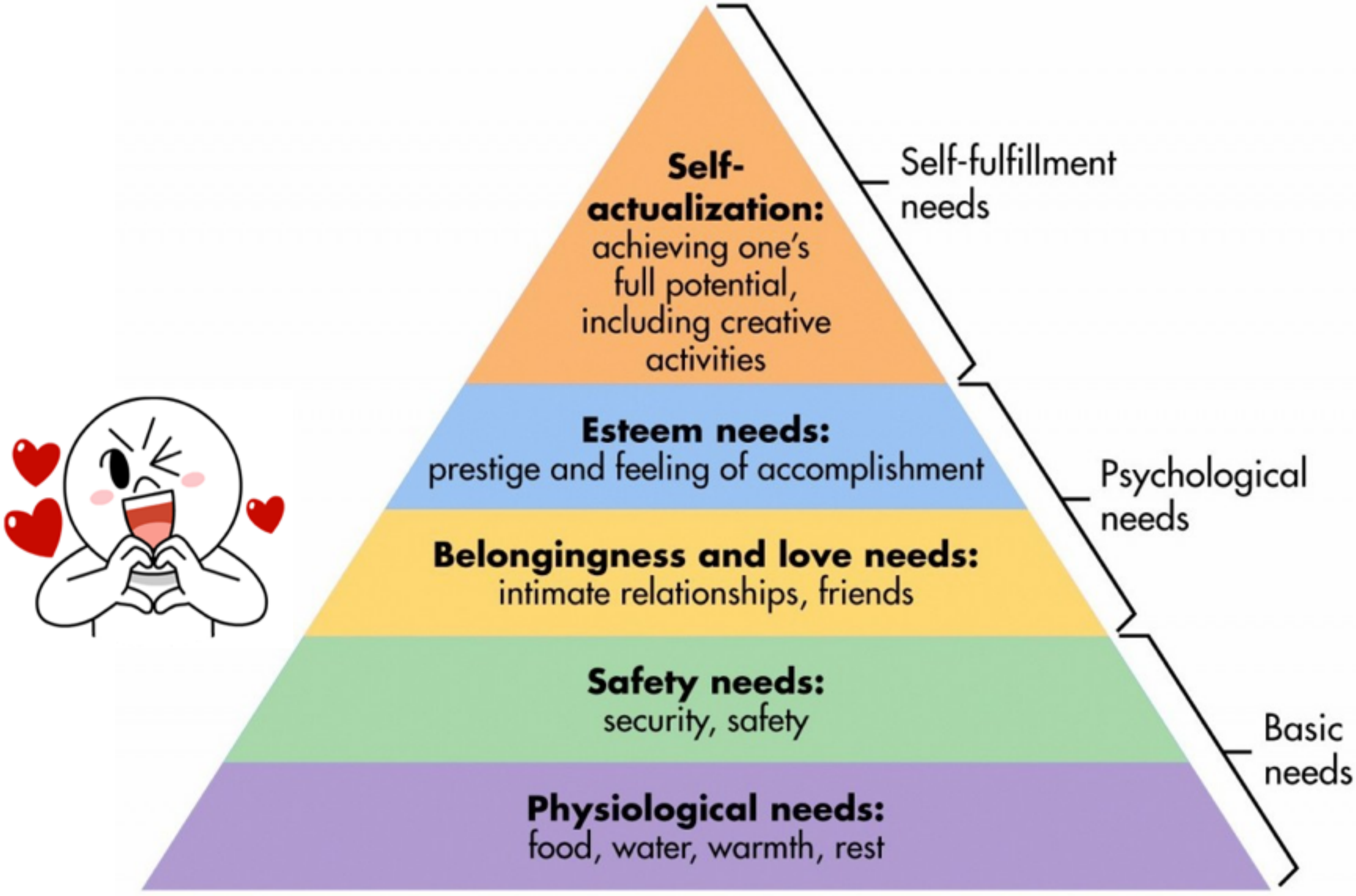
Source: Philip Kotler & Kevin Lane Keller, Marketing Management, 14th ed., Pearson, 2012

Maslow’s hierarchy of human needs

(Maslow, 1943)

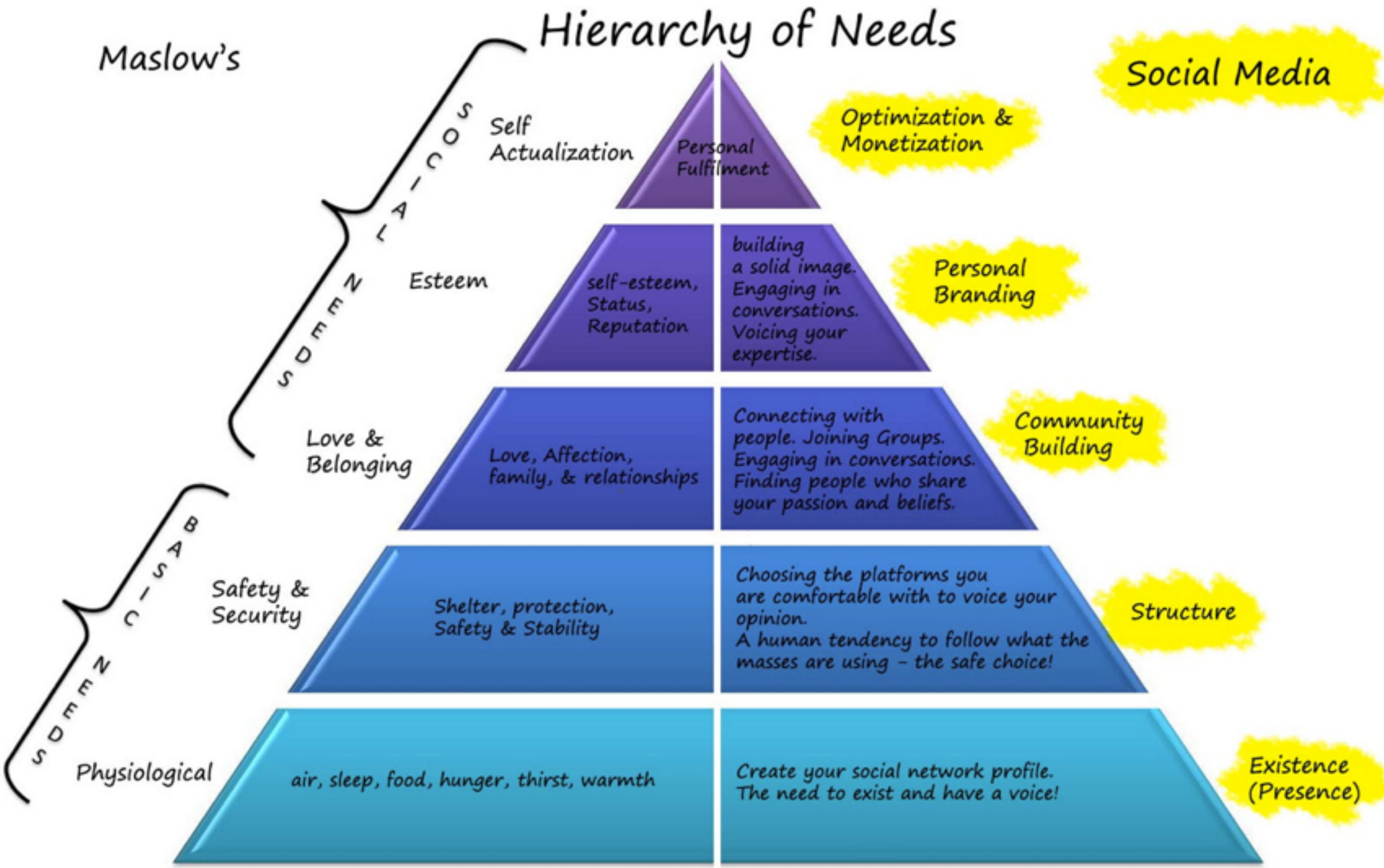
Source: Backer & Saren (2009), Marketing Theory: A Student Text, 2nd Edition, Sage

Maslow's Hierarchy of Needs



Source: http://sixstoriesup.com/social-psyche-what-makes-us-go-social/

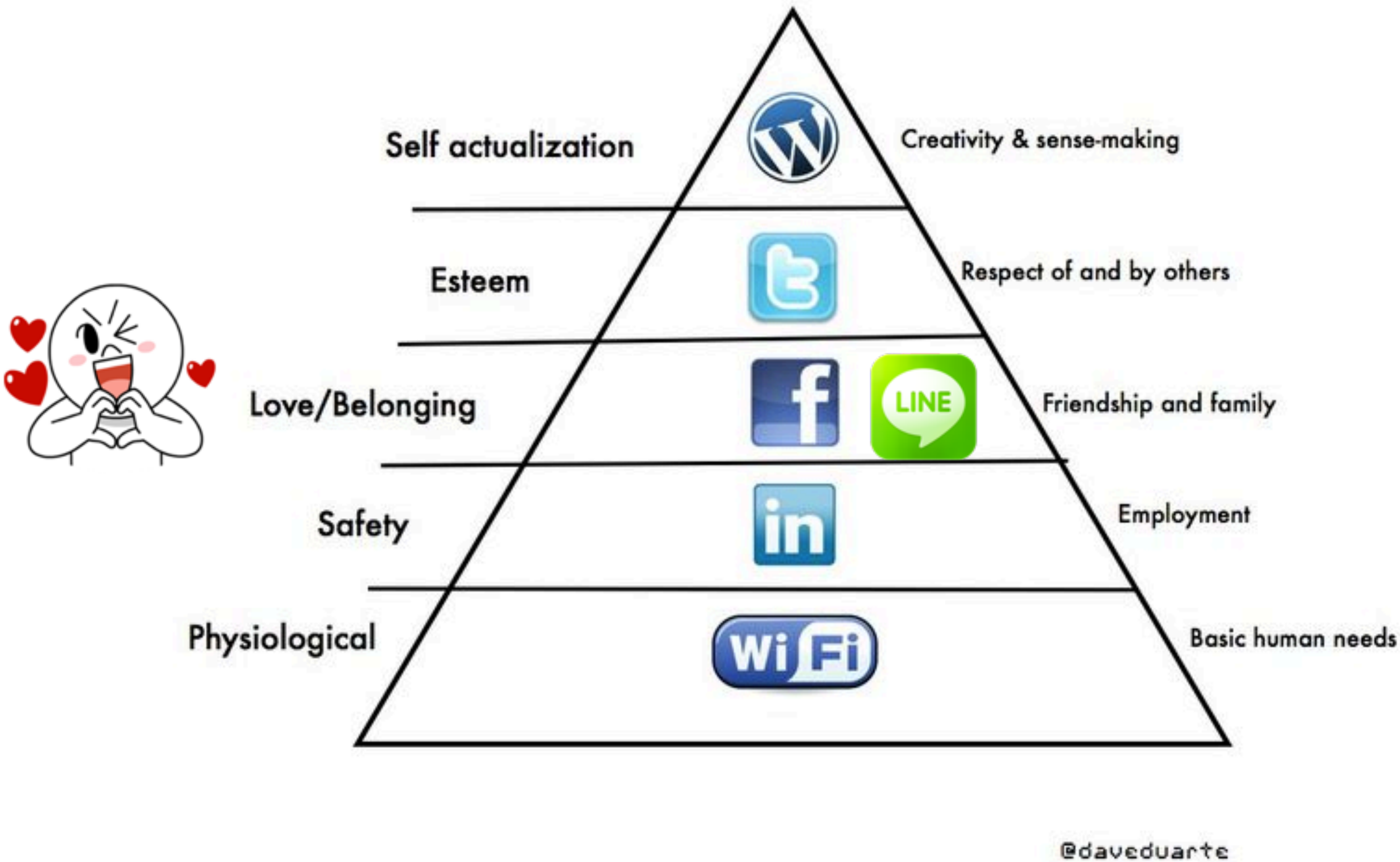
Social Media Hierarchy of Needs



Social Media Hierarchy of Needs - by John Antonios

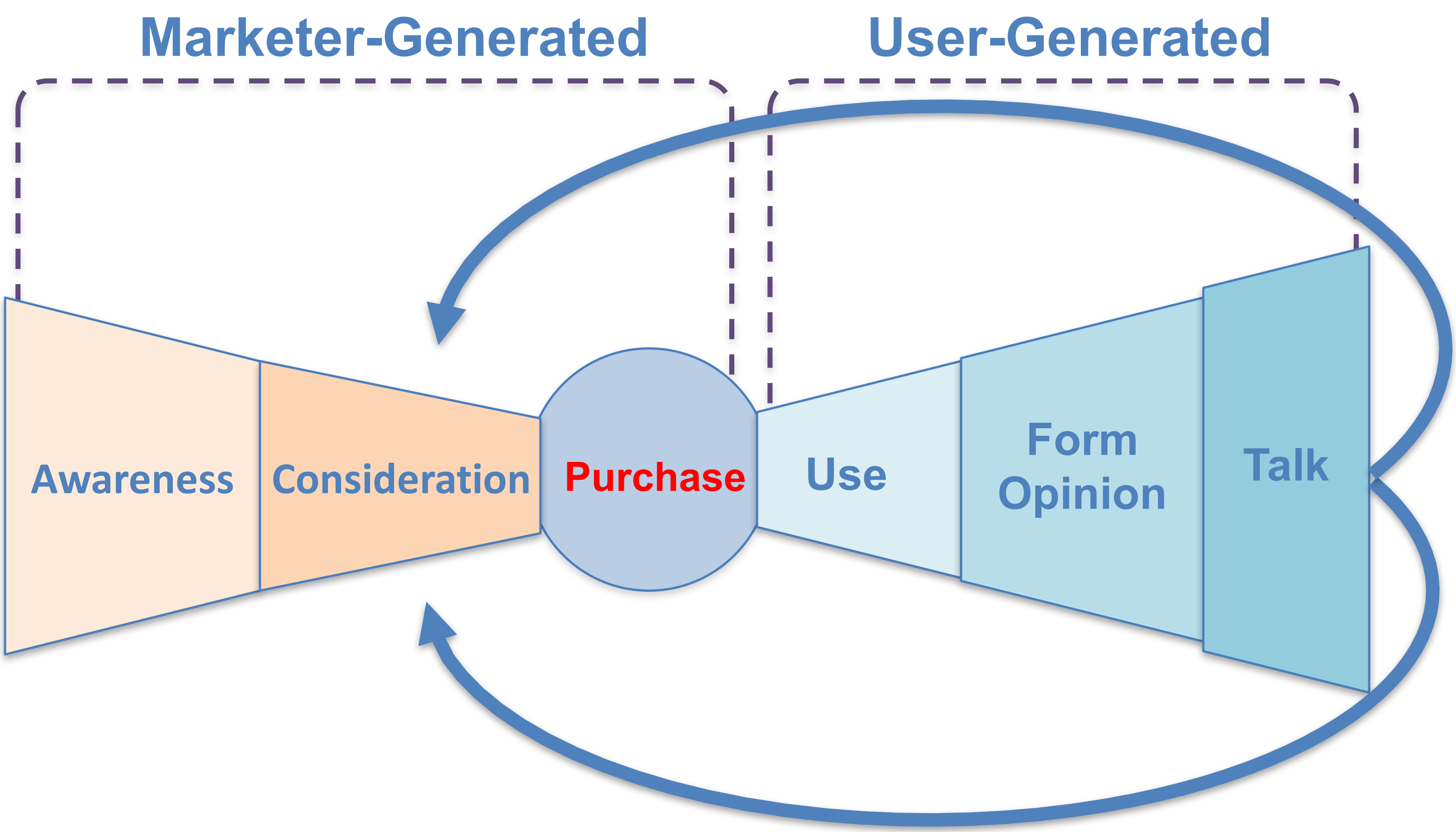
Source: http://2.bp.blogspot.com/_Rta1VZltiMk/TPavcanFtfI/AAAAAAAAACo/OBGnRL5arSU/s1600/social-media-heirarchy-of-needs1.jpg

Social Media Hierarchy of Needs



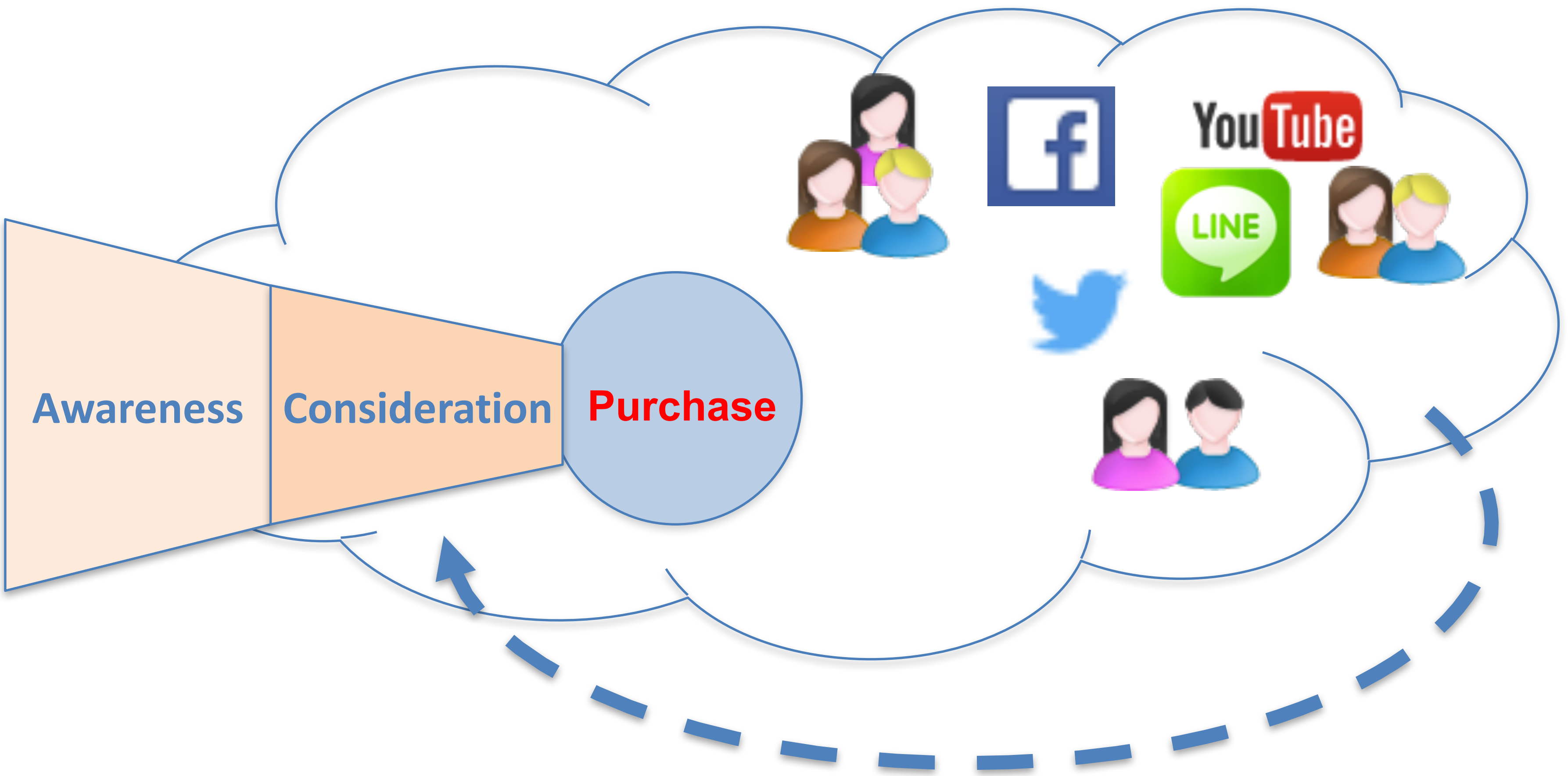
Source: http://www.pinterest.com/pin/186477859309903585/

The Social Feedback Cycle
Consumer Behavior on Social Media



Source: Evans et al. (2010), Social Media Marketing: The Next Generation of Business Engagement

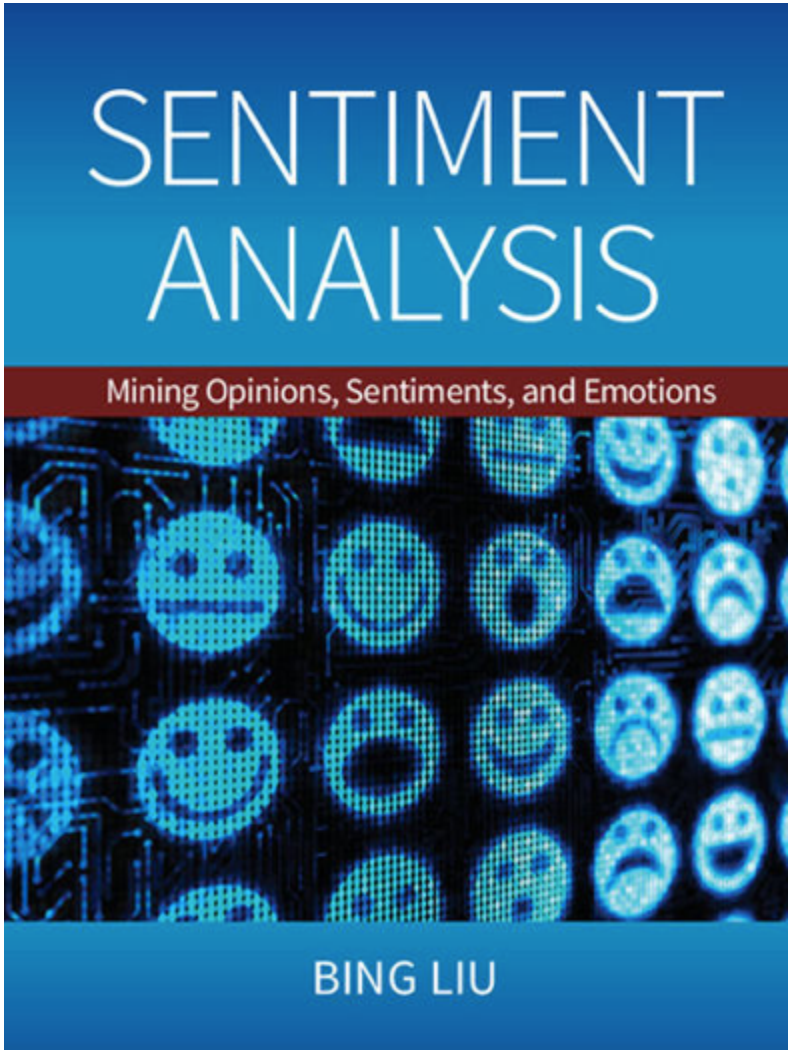
The New Customer Influence Path



Source: Evans et al. (2010), Social Media Marketing: The Next Generation of Business Engagement

Architectures of Sentiment Analytics

Bing Liu (2015),
Sentiment Analysis:
Mining Opinions, Sentiments, and Emotions,
Cambridge University Press



http://www.amazon.com/Sentiment-Analysis-Opinions-Sentiments-Emotions/dp/1107017890

Sentiment Analysis and Opinion Mining

- Computational study of
  opinions,
  sentiments,
  subjectivity,
  evaluations,
  attitudes,
  appraisal,
  affects,
  views,
  emotions,
  ets., expressed in text.
  - Reviews, blogs, discussions, news, comments, feedback, or any other documents

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Research Area of Opinion Mining

- Many names and tasks with difference objective and models
  - Sentiment analysis
  - Opinion mining
  - Sentiment mining
  - Subjectivity analysis
  - Affect analysis
  - Emotion detection
  - Opinion spam detection

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Sentiment Analysis

- Sentiment
  - A thought, view, or attitude, especially one based mainly on emotion instead of reason
- Sentiment Analysis
  - opinion mining
  - use of natural language processing (NLP) and computational techniques to automate the extraction or classification of sentiment from typically unstructured text

Applications of Sentiment Analysis

- Consumer information
  - Product reviews
- Marketing
  - Consumer attitudes
  - Trends
- Politics
  - Politicians want to know voters’ views
  - Voters want to know policitians’ stances and who else supports them
- Social
  - Find like-minded individuals or communities

Sentiment detection

- How to interpret features for sentiment detection?
  - Bag of words (IR)
  - Annotated lexicons (WordNet, SentiWordNet)
  - Syntactic patterns
- Which features to use?
  - Words (unigrams)
  - Phrases/n-grams
  - Sentences

Problem statement of Opinion Mining

- Two aspects of abstraction
  - Opinion definition
    - What is an opinion?
    - What is the structured definition of opinion?
  - Opinion summarization
    - Opinion are subjective
      - An opinion from a single person (unless a VIP) is often not sufficient for action
    - We need opinions from many people, and thus opinion summarization.

What is an opinion?

- Id: **Abc123** on **5-1-2008** "*I bought an iPhone a few days ago. It is such a nice phone. The touch screen is really cool. The voice quality is clear too. It is much better than my old Blackberry, which was a terrible phone and so difficult to type with its tiny keys. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, …*"
- One can look at this review/blog at the
  - Document level
    - Is this review + or -?
  - Sentence level
    - Is each sentence + or -?
  - Entity and feature/aspect level

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Entity and aspect/feature level

- Id: **Abc123** on **5-1-2008** "*I bought an iPhone a few days ago. It is such a nice phone. The touch screen is really cool. The voice quality is clear too. It is much better than my old Blackberry, which was a terrible phone and so difficult to type with its tiny keys. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, …*"
- What do we see?
  - Opinion targets: entities and their features/aspects
  - Sentiments: positive and negative
  - Opinion holders: persons who hold the opinions
  - Time: when opinion are expressed

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Two main types of opinions

- Regular opinions: Sentiment/Opinion expressions on some target entities
  - Direct opinions: sentiment expressions on one object:
    - "The touch screen is really cool."
    - "The picture quality of this camera is great"
  - Indirect opinions: comparisons, relations expressing similarities or differences (objective or subjective) of more than one object
    - "phone X is cheaper than phone Y." (objective)
    - "phone X is better than phone Y." (subjective)
- Comparative opinions: comparisons of more than one entity.
  - "iPhone is better than Blackberry."

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Subjective and Objective

- Objective
  - An objective sentence expresses some factual information about the world.
  - "I returned the phone yesterday."
  - Objective sentences can implicitly indicate opinions
    - "The earphone broke in two days."
- Subjective
  - A subjective sentence expresses some personal feelings or beliefs.
  - "The voice on my phone was not so clear"
  - Not every subjective sentence contains an opinion
    - "I wanted a phone with good voice quality"
- → Subjective analysis

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Sentiment Analysis
vs.
Subjectivity Analysis

Sentiment Analysis	Subjectivity Analysis
Positive	Subjective
Negative	Subjective
Neutral	Objective

A (regular) opinion

- Opinion (a restricted definition)
  - An opinion (regular opinion) is simply a positive or negative sentiment, view, attitude, emotion, or appraisal about an entity or an aspect of the entity from an opinion holder.
- Sentiment orientation of an opinion
  - Positive, negative, or neutral (no opinion)
  - Also called:
    - Opinion orientation
    - Semantic orientation
    - Sentiment polarity

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Entity and aspect

- Definition of Entity:
  - An *entity e* is a product, person, event, organization, or topic.
  - e is represented as
    - A hierarchy of components, sub-components.
    - Each node represents a components and is associated with a set of attributes of the components
- An opinion can be expressed on any node or attribute of the node
- Aspects(features)
  - represent both components and attribute

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Opinion Definition

- An opinion is a quintuple
  (e_j, a_{jk}, so_{ijkl}, h_i, t_l)
  where
  - e_j is a target entity.
  - a_{jk} is an aspect/feature of the entity e_j .
  - so_{ijkl} is the sentiment value of the opinion from the opinion holder on feature of entity at time.
    so_{ijkl} is +ve, -ve, or neu, or more granular ratings
  - h_i is an opinion holder.
  - t_l is the time when the opinion is expressed.
- (e_j, a_{jk}) is also called opinion target

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Terminologies

- Entity: object
- Aspect: feature, attribute, facet
- Opinion holder: opinion source

- Topic: entity, aspect

- Product features, political issues

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Subjectivity and Emotion

- Sentence subjectivity
  - An objective sentence presents some factual information, while a subjective sentence expresses some personal feelings, views, emotions, or beliefs.
- Emotion
  - Emotions are people's subjective feelings and thoughts.

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Classification Based on Supervised Learning

- Sentiment classification
  - Supervised learning Problem
  - Three classes
    - *Positive*
    - *Negative*
    - *Neutral*

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Opinion words in Sentiment classification

- topic-based classification
  - topic-related words are important
    - e.g., *politics, sciences, sports*
- Sentiment classification
  - topic-related words are unimportant
  - **opinion words** (also called **sentiment words)**
    - **that indicate positive or negative opinions** are important,
      e.g., *great, excellent, amazing, horrible, bad, worst*

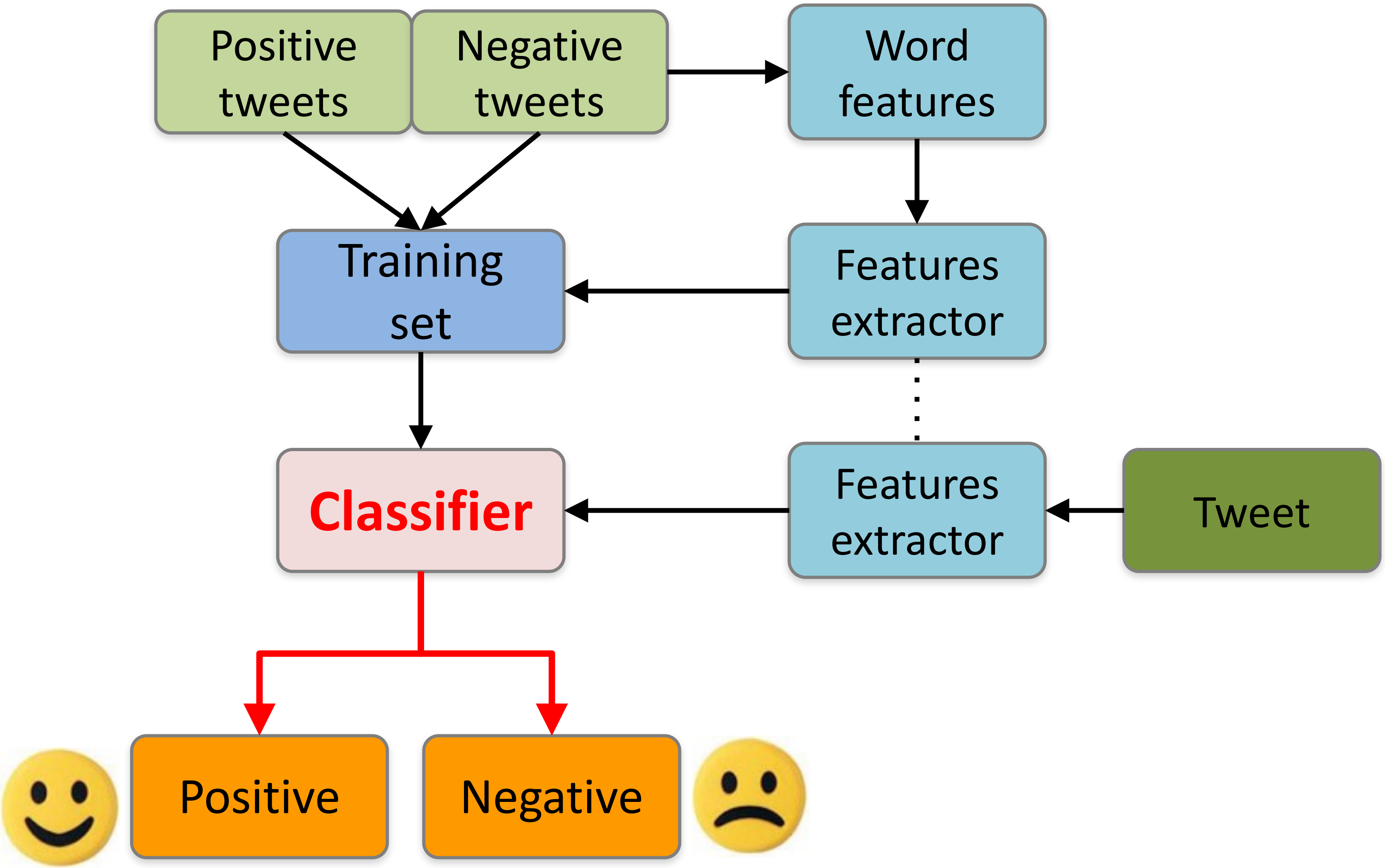
Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Features in Opinion Mining

- *Terms and their frequency*
  - *TF-IDF*
- *Part of speech (POS)*
  - Adjectives
- *Opinion words and phrases*
  - *beautiful, wonderful, good, and amazing are positive opinion words*
  - *bad, poor, and terrible are negative opinion words.*
  - opinion phrases and idioms,
    e.g., *cost someone an arm and a leg*
- *Rules of opinions*
- *Negations*
- *Syntactic dependency*

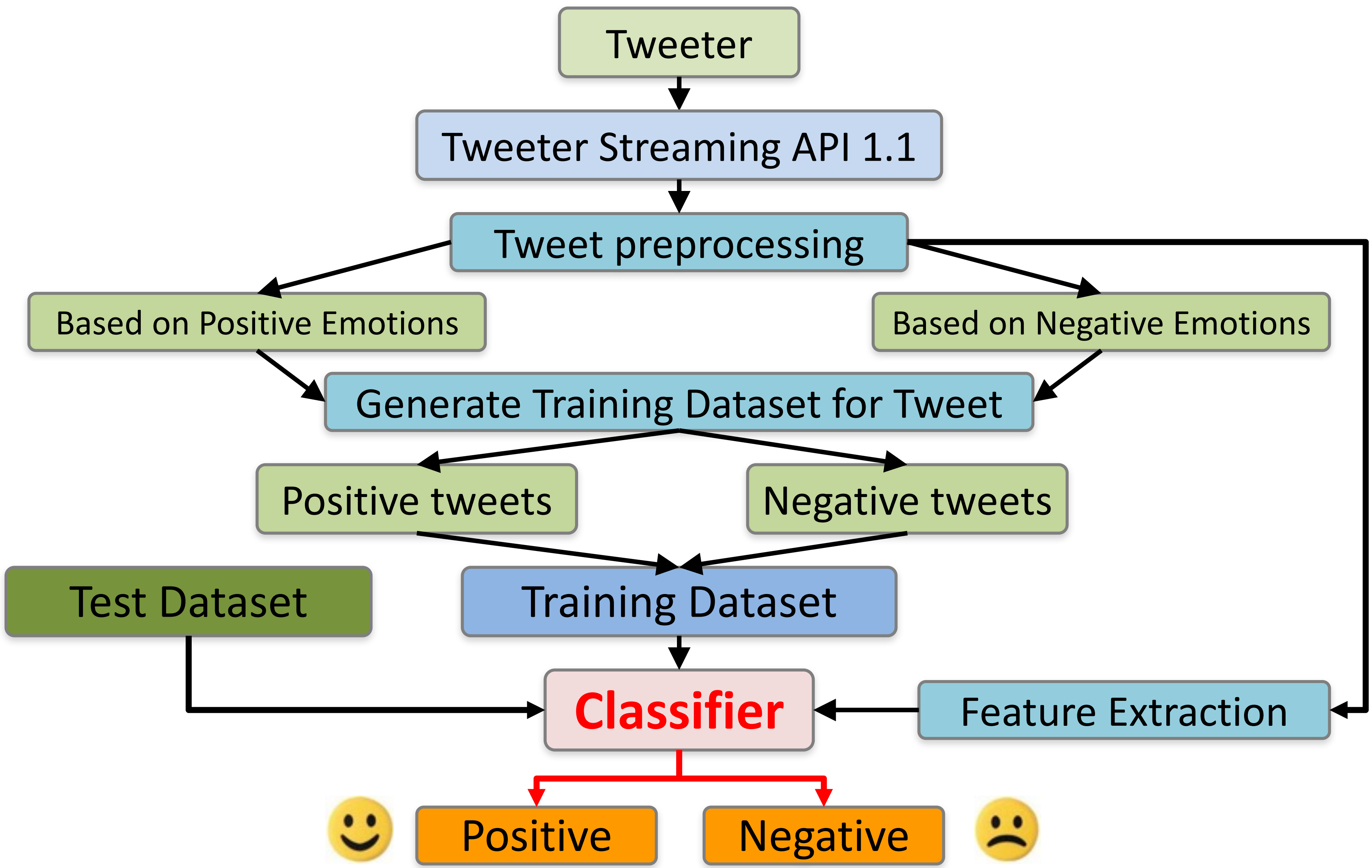
Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Sentiment Analysis Architecture



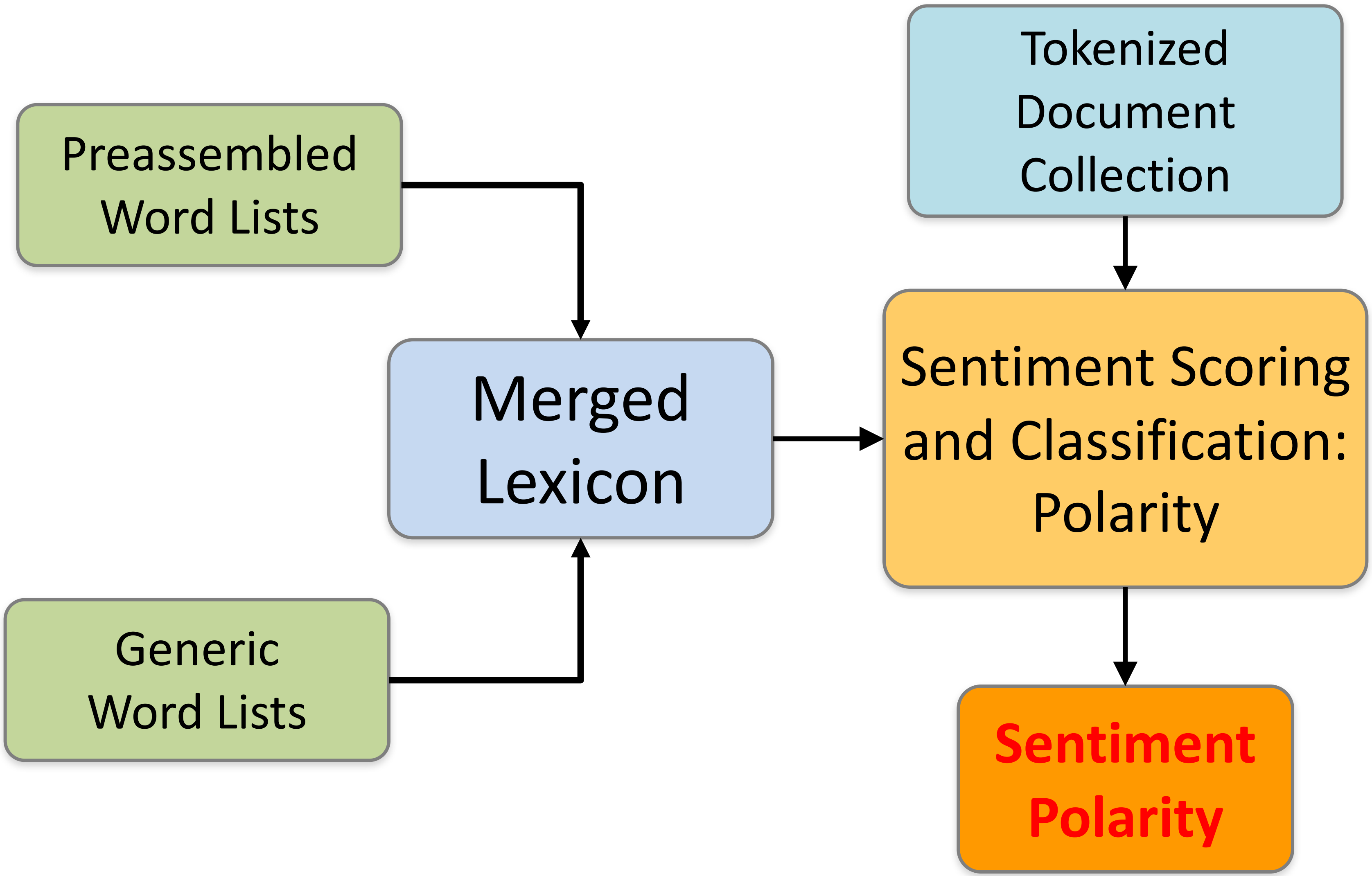
Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, Vol 139, No. 11, 2016. pp.5-15

Sentiment Classification Based on Emoticons



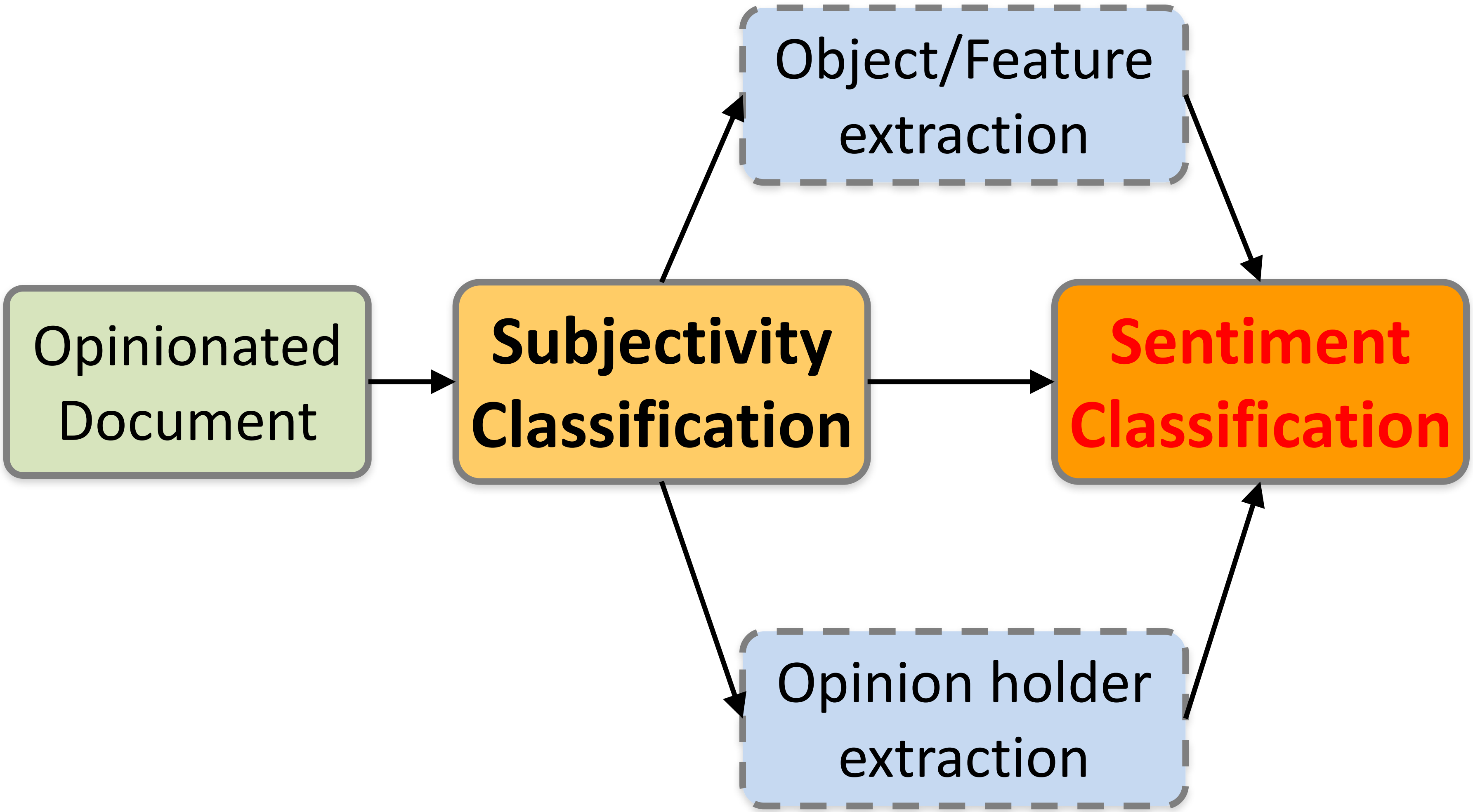
Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, Vol 139, No. 11, 2016. pp.5-15

Lexicon-Based Model



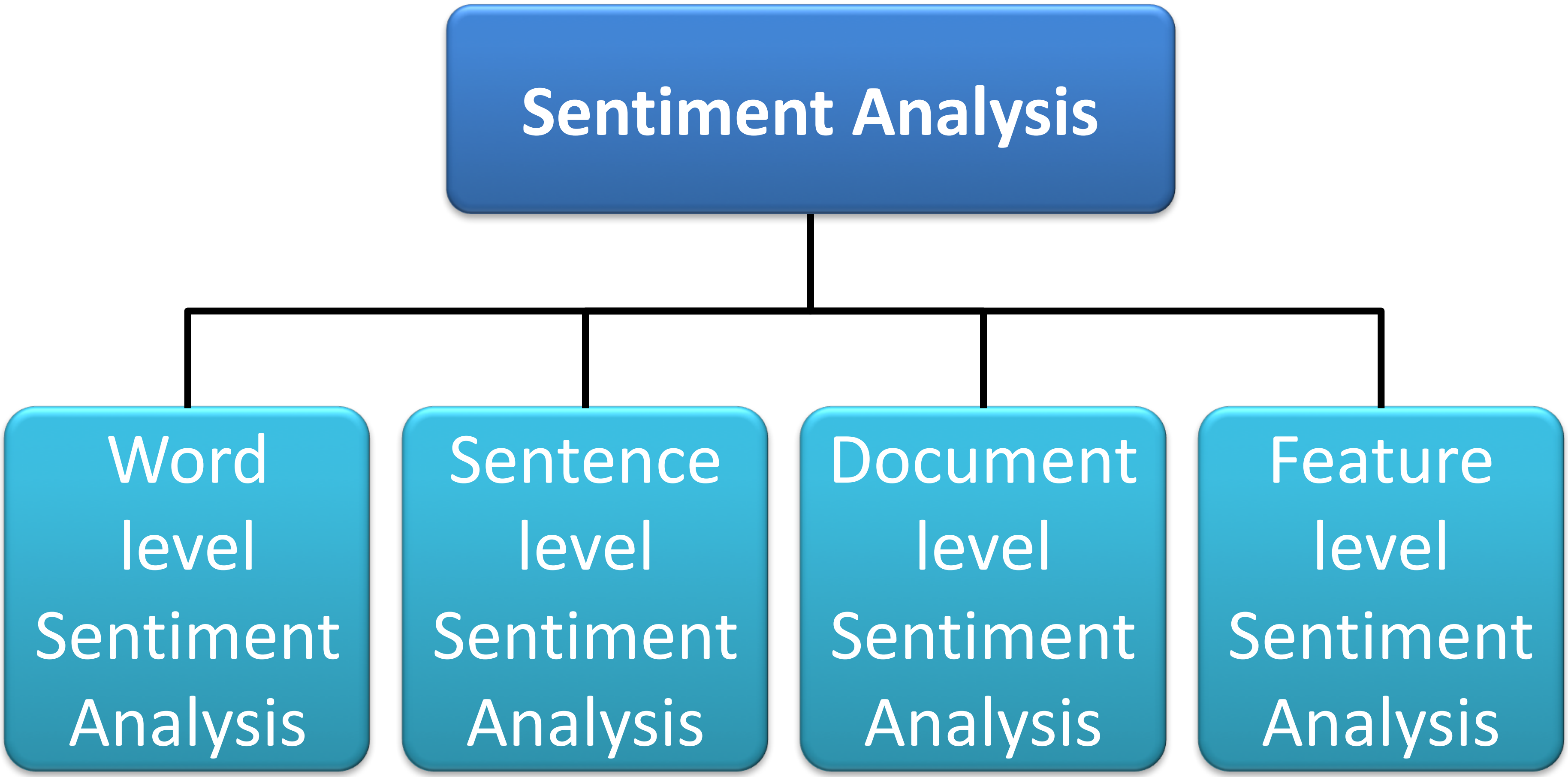
Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, Vol 139, No. 11, 2016. pp.5-15

Sentiment Analysis Tasks



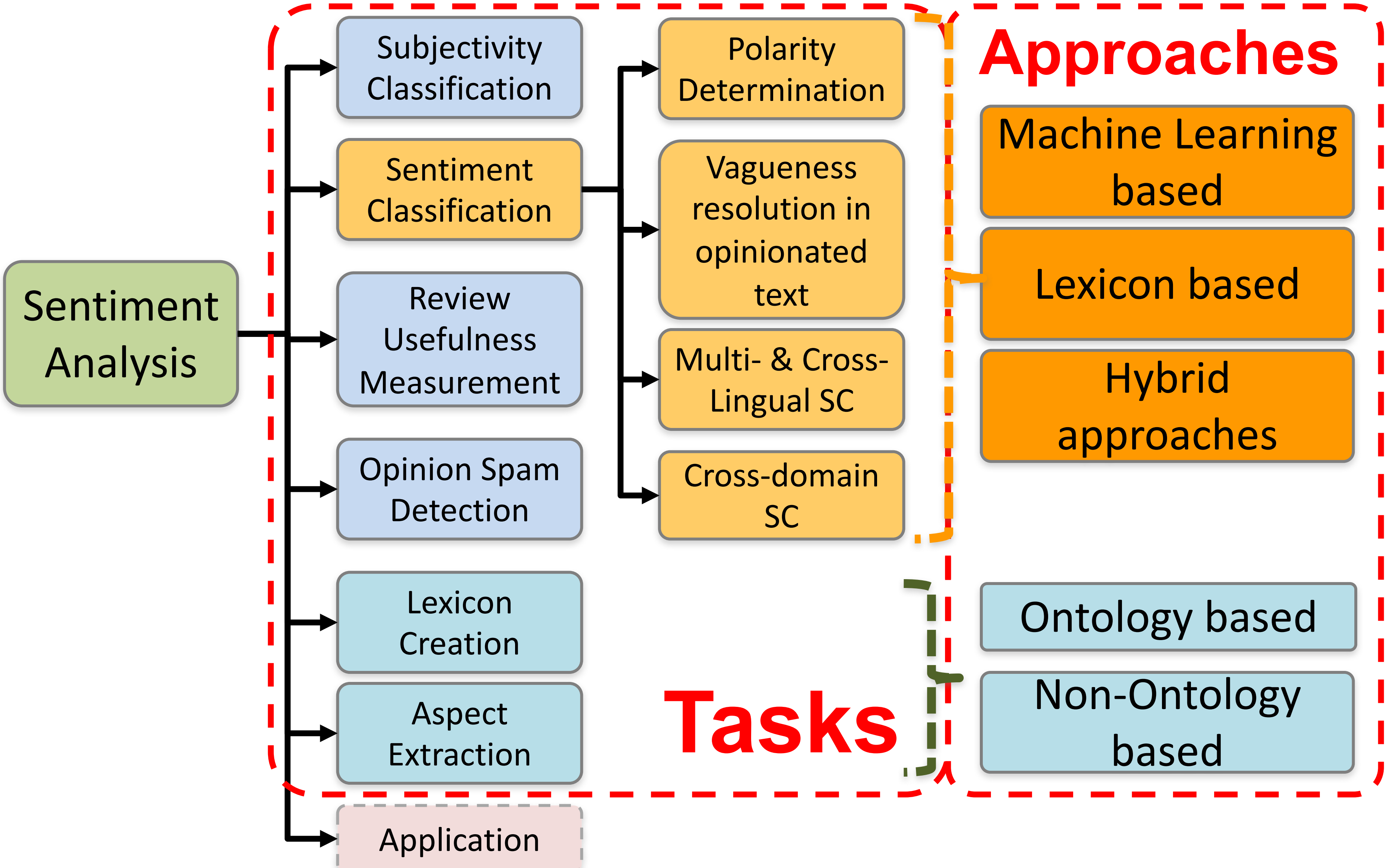
Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, Vol 139, No. 11, 2016. pp.5-15

Levels of Sentiment Analysis



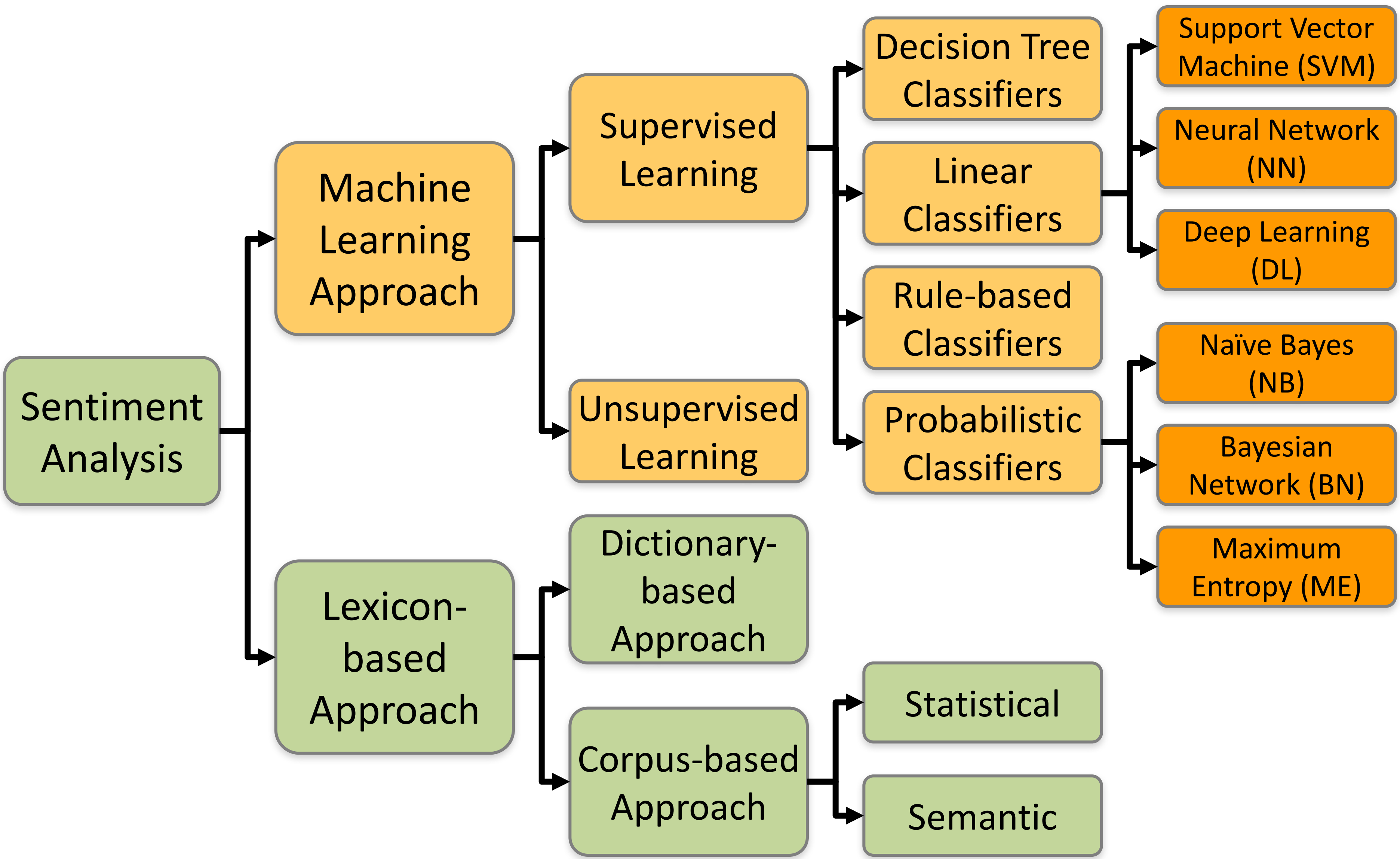
Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, Vol 139, No. 11, 2016. pp.5-15

Sentiment Analysis



Source: Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.

Sentiment Classification Techniques



Source: Jesus Serrano-Guerrero, Jose A. Olivas, Francisco P. Romero, and Enrique Herrera-Viedma (2015), "Sentiment analysis: A review and comparative analysis of web services," Information Sciences, 311, pp. 18-38.

A Brief Summary of Sentiment Analysis Methods

Study	Analysis Task	Sentiment Identification		Sentiment Aggregation		Nature of Measure
		Method	Level	Method	Level	
Hu and Li, 2011	Polarity	ML (Probabilistic model)	Snippet			Valence
Li and Wu, 2010	Polarity	Lexicon/Rule	Phrase	Sum	Snippet	Valence
Thelwall et al., 2010	Polarity	Lexicon/Rule	Sentence	Max & Min	Snippet	Range
Boiy and Moens, 2009	Both	ML (Cascade ensemble)	Sentence			Valence
Chung 2009	Polarity	Lexicon	Phrase	Average	Sentence	Valence
Wilson, Wiebe, and Hoffmann, 2009	Both	ML (SVM, AdaBoost, Rule, etc.)	Phrase			Valence
Zhang et al., 2009	Polarity	Lexicon/Rule	Sentence	Weighted average	Snippet	Valence
Abbasi, Chen, and Salem, 2008	Polarity	ML (GA + feature selection)	Snippet			Valence
Subrahmanian and Reforgiato, 2008	Polarity	Lexicon/Rule	Phrase	Rule	Snippet	Valence
Tan and Zhang 2008	Polarity	ML (SVM, Winnow, NB, etc.)	Snippet			Valence
Airoldi, Bai, and Padman, 2007	Polarity	ML (Markov Blanket)	Snippet			Valence
Das and Chen, 2007	Polarity	ML (Bayesian, Discriminate, etc.)	Snippet	Average	Daily	Valence
Liu et al., 2007	Polarity	ML (PLSA)	Snippet			Valence
Kennedy and Inkpen, 2006	Polarity	Lexicon/Rule, ML (SVM)	Phrase	Count	Snippet	Valence
Mishne 2006	Polarity	Lexicon	Phrase	Average	Snippet	Valence
Liu et al., 2005	Polarity	Lexicon/Rule	Phrase	Distribution	Object	Range
Mishne 2005	Polarity	ML (SVM)	Snippet			Valence
Popescu and Etzioni 2005	Polarity	Lexicon/Rule	Phrase			Valence
Efron 2004	Polarity	ML (SVN, NB)	Snippet			Valence
Wilson, Wiebe, and Hwa, 2004	Both	ML (SVM, AdaBoost, Rule, etc.)	Sentence			Valence
Nigam and Hurst 2004	Polarity	Lexicon/Rule	Chunk	Rule	Sentence	Valence
Dave, Lawrence, and Pennock, 2003	Polarity	ML (SVM, Rainbow, etc.)	Snippet			Valence
Nasukawa and Yi 2003	Polarity	Lexicon/Rule	Phrase	Rule	Sentence	Valence
Yi et al., 2003	Polarity	Lexicon/Rule	Phrase	Rule	Sentence	Valence
Yu and Hatzivassiloglou 2003	Both	ML (NB) + Lexicon/Rule	Phrase	Average	Sentence	Valence
Pang, Lee, and Vaithyanathan 2002	Polarity	ML (SVM, MaxEnt, NB)	Snippet			Valence
Subasic and Huettner 2001	Polarity	Lexicon/Fuzzy logic	Phrase	Average	Snippet	Valence
Turney 2001	Polarity	Lexicon/Rule	Phrase	Average	Snippet	Valence

(Both = Subjectivity and Polarity; ML= Machine Learning; Lexicon/Rule= Lexicon enhanced by linguistic rules)

Source: Zhang, Z., Li, X., and Chen, Y. (2012), "Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews," ACM Trans. Manage. Inf. Syst. (3:1) 2012, pp 1-23.,

Word-of-Mouth (WOM)

- “This book is the best written documentary thus far, yet sadly, there is no soft cover edition.”

- “This book is the best written documentary thus far, yet sadly, there is no soft cover edition.”

Source: Zhang, Z., Li, X., and Chen, Y. (2012), "Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews," ACM Trans. Manage. Inf. Syst. (3:1) 2012, pp 1-23.,

This
book
is
the
best
written
documentary
thus
far
,
yet
sadly
,
there
is
no
soft
cover
edition
.

Word	POS
This	DT
book	NN
is	VBZ
the	DT
best	JJS
written	VBN
documentary	NN
thus	RB
far	RB
,	,
yet	RB
sadly	RB
,	,
there	EX
is	VBZ
no	DT
soft	JJ
cover	NN
edition	NN
.	.

Source: Zhang, Z., Li, X., and Chen, Y. (2012), "Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews," ACM Trans. Manage. Inf. Syst. (3:1) 2012, pp 1-23.,

Conversion of text representation

Word Vector (WV)	Polarity Score Vector (PSV) *pscore*	*nscore*		Microstate Sequence (MS)
This	0	0	Neutral (0)	0
book	0	0	Neutral (0)	0
is	0	0	Neutral (0)	0
the	0	0	Neutral (0)	0
best	0.75	0	Positive (0.75)	1
written	0	0	Neutral (0)	0
documentary	0	0	Neutral (0)	0
thus	0.375	0	Positive (0.375)	1
far	0.375	0	Positive (0.375)	1
,				
yet	0	0.125	Negative (0.125)	-1
sadly	0.25	0.5	Negative (0.25)	-1
,				
there	0	0	Neutral (0)	0
is	0	0	Neutral (0)	0
no	0	0.75	Negative (0.75)	-1
soft	0	0	Neutral (0)	0
cover	0	0	Neutral (0)	0
edition	0	0	Neutral (0)	0
.				

SentiWordNet Lookup → Microstate Mapping → Probability Mapping

Probability Distribution (P)

P("1")=3/17

P("-1")=3/17

P("0")=11/17

Source: Zhang, Z., Li, X., and Chen, Y. (2012), "Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews," ACM Trans. Manage. Inf. Syst. (3:1) 2012, pp 1-23.,

Example of SentiWordNet

POS ID PosScore NegScore SynsetTerms Gloss

a 00217728 0.75 0 beautiful#1 delighting the senses or exciting intellectual or emotional admiration; "a beautiful child"; "beautiful country"; "a beautiful painting"; "a beautiful theory"; "a beautiful party“

a 00227507 0.75 0 best#1 (superlative of `good') having the most positive qualities; "the best film of the year"; "the best solution"; "the best time for planting"; "wore his best suit“

r 00042614 0 0.625 unhappily#2 sadly#1 in an unfortunate way; "sadly he died before he could see his grandchild“

r 00093270 0 0.875 woefully#1 sadly#3 lamentably#1 deplorably#1 in an unfortunate or deplorable manner; "he was sadly neglected"; "it was woefully inadequate“

r 00404501 0 0.25 sadly#2 with sadness; in a sad manner; "`She died last night,' he said sadly"

SenticNet

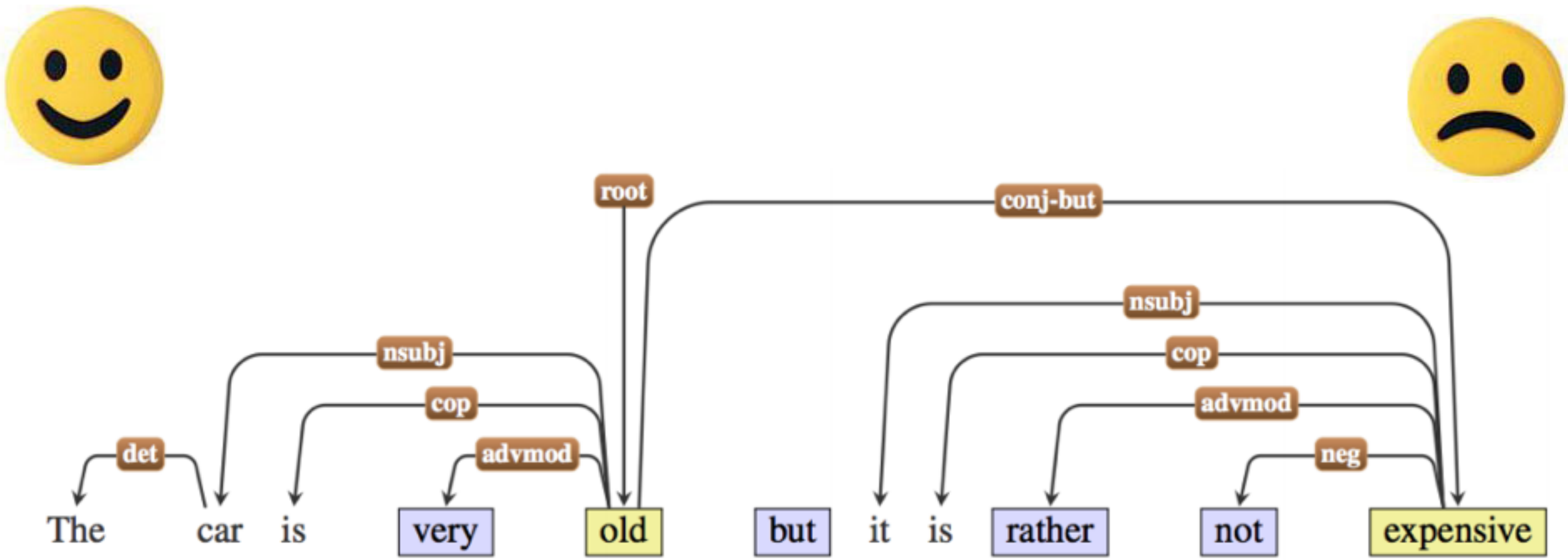
The car is very old but it is rather not expensive.

The car is very old but it is rather not expensive.

The car is <u>very</u> old <u>but</u> it is <u>rather</u> <u>not</u> expensive.

Source: Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In *the 26th International Conference on Computational Linguistics (COLING), Osaka*. 2016.

Polarity Detection with SenticNet

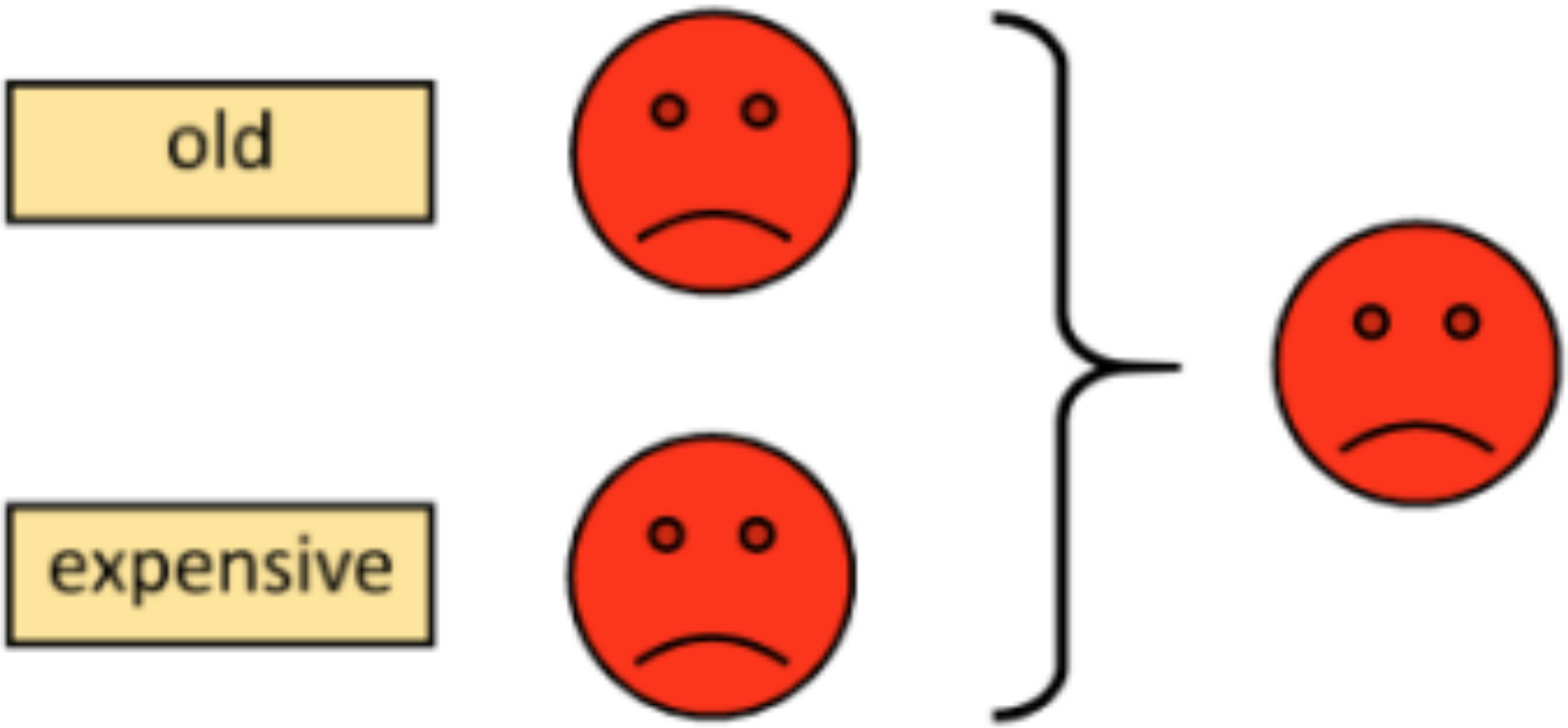


The car is very old but it is rather not expensive.

The car is very old but it is rather not expensive.

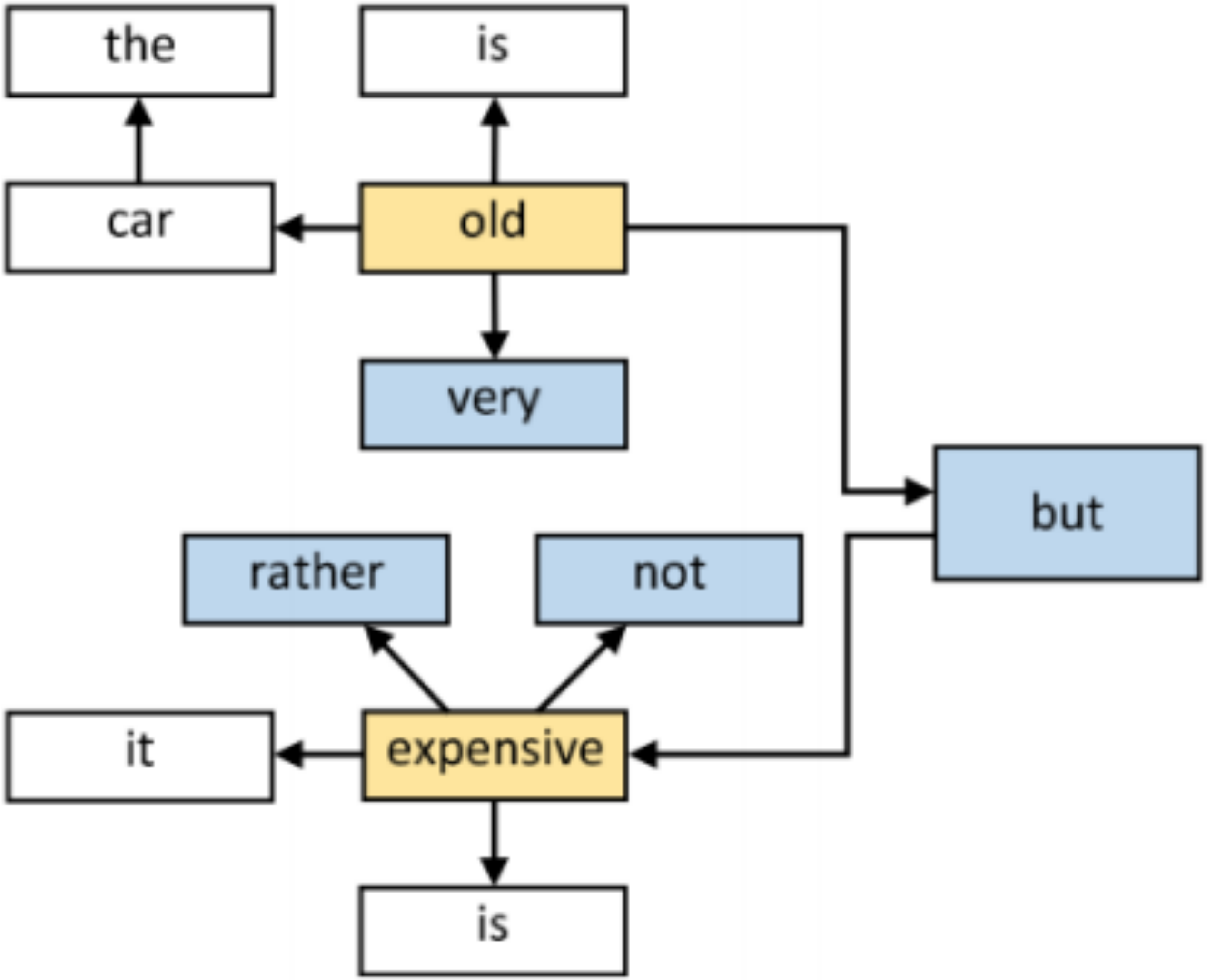
Source: Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In *the 26th International Conference on Computational Linguistics (COLING), Osaka*. 2016.

Polarity Detection with SenticNet



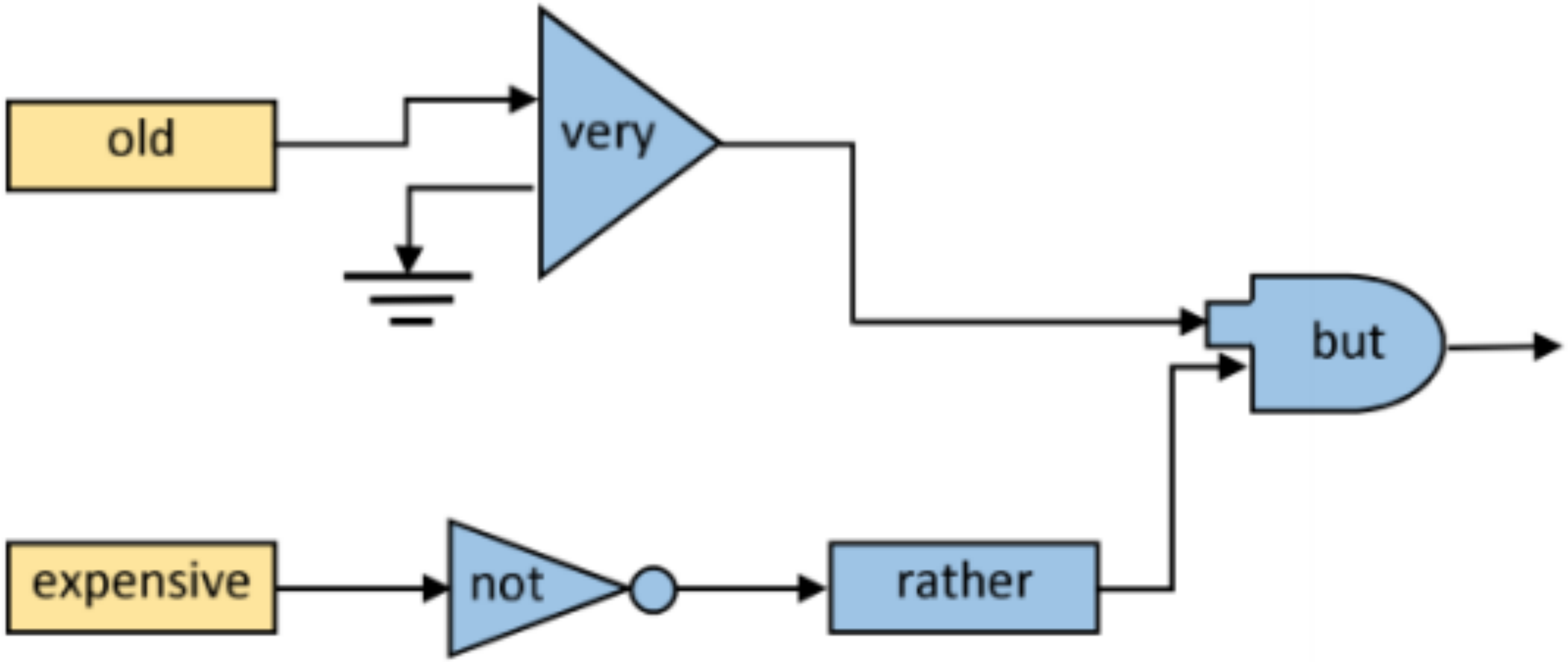
Source: Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In *the 26th International Conference on Computational Linguistics (COLING), Osaka*. 2016.

Polarity Detection with SenticNet



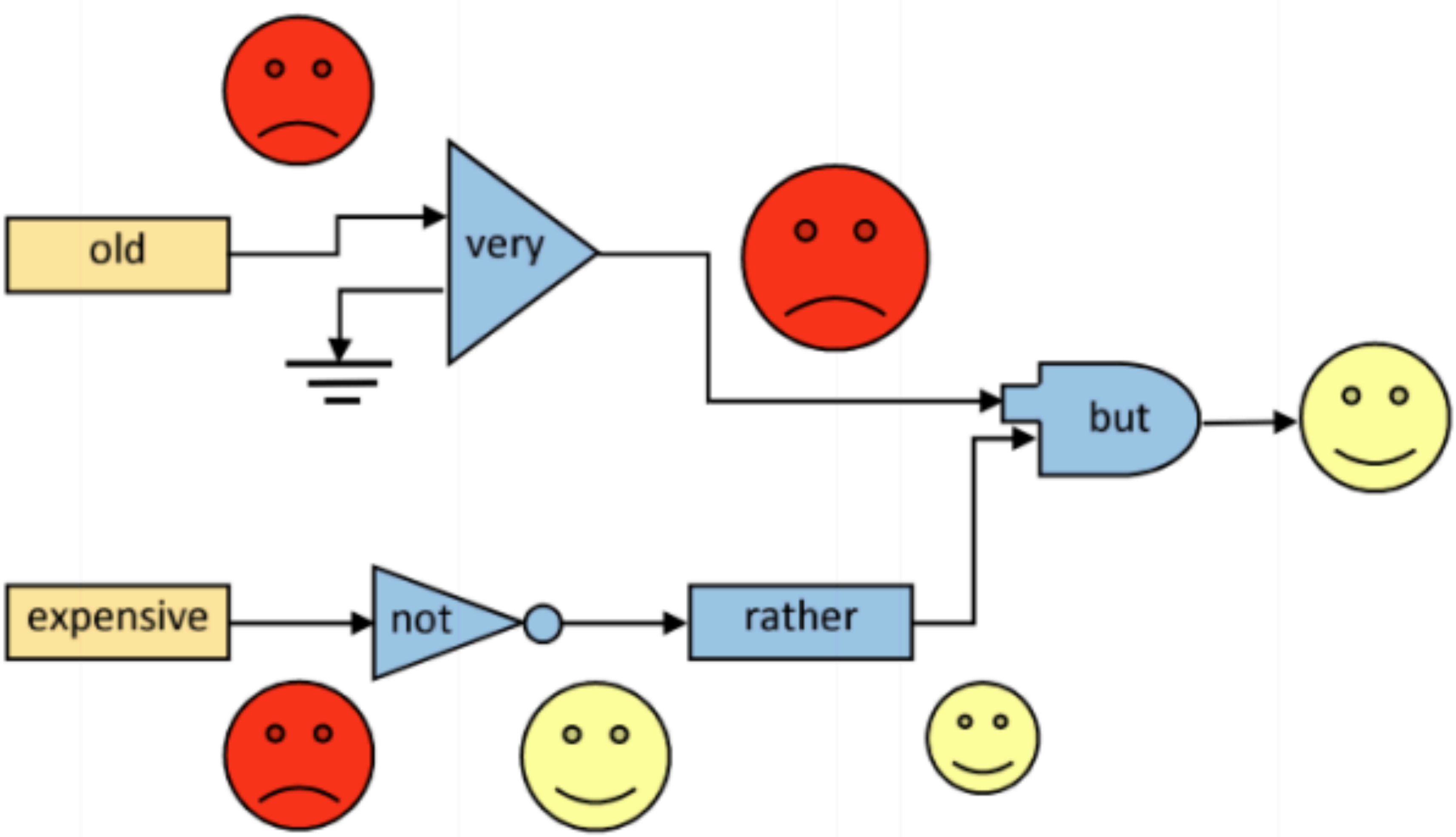
Source: Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In *the 26th International Conference on Computational Linguistics (COLING), Osaka*. 2016.

Polarity Detection with SenticNet



Source: Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In *the 26th International Conference on Computational Linguistics (COLING), Osaka*. 2016.

Polarity Detection with SenticNet



Source: Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In *the 26th International Conference on Computational Linguistics (COLING), Osaka*. 2016.

Evaluation of Text Mining and Sentiment Analysis

- Evaluation of Information Retrieval
- Evaluation of Classification Model (Prediction)
  - Accuracy
  - Precision
  - Recall
  - F-score

Deep Learning for Sentiment Analytics

Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank

Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank

Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng and Christopher Potts
Stanford University, Stanford, CA 94305, USA
richard@socher.org,{aperelyg,jcchuang,ang}@cs.stanford.edu
{jeaneis,manning,cgpotts}@stanford.edu

Abstract

Semantic word spaces have been very useful but cannot express the meaning of longer phrases in a principled way. Further progress towards understanding compositionality in tasks such as sentiment detection requires richer supervised training and evaluation resources and more powerful models of composition. To remedy this, we introduce a Sentiment Treebank. It includes fine grained sentiment labels for 215,154 phrases in the parse trees of 11,855 sentences and presents new challenges for sentiment compositionality. To address them, we introduce the Recursive Neural Tensor Network. When trained on the new treebank, this model out-

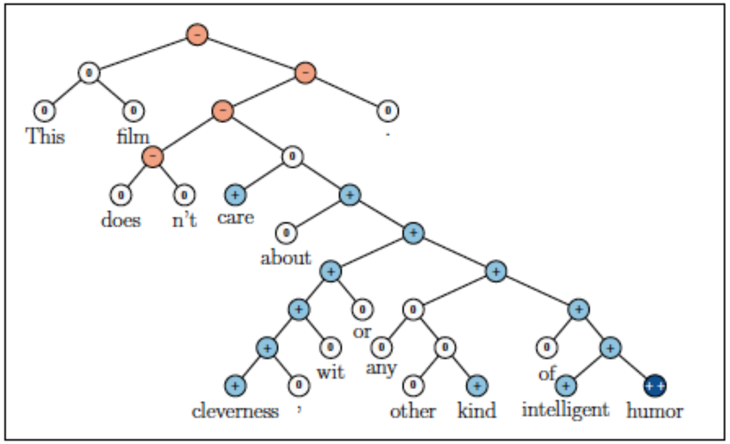
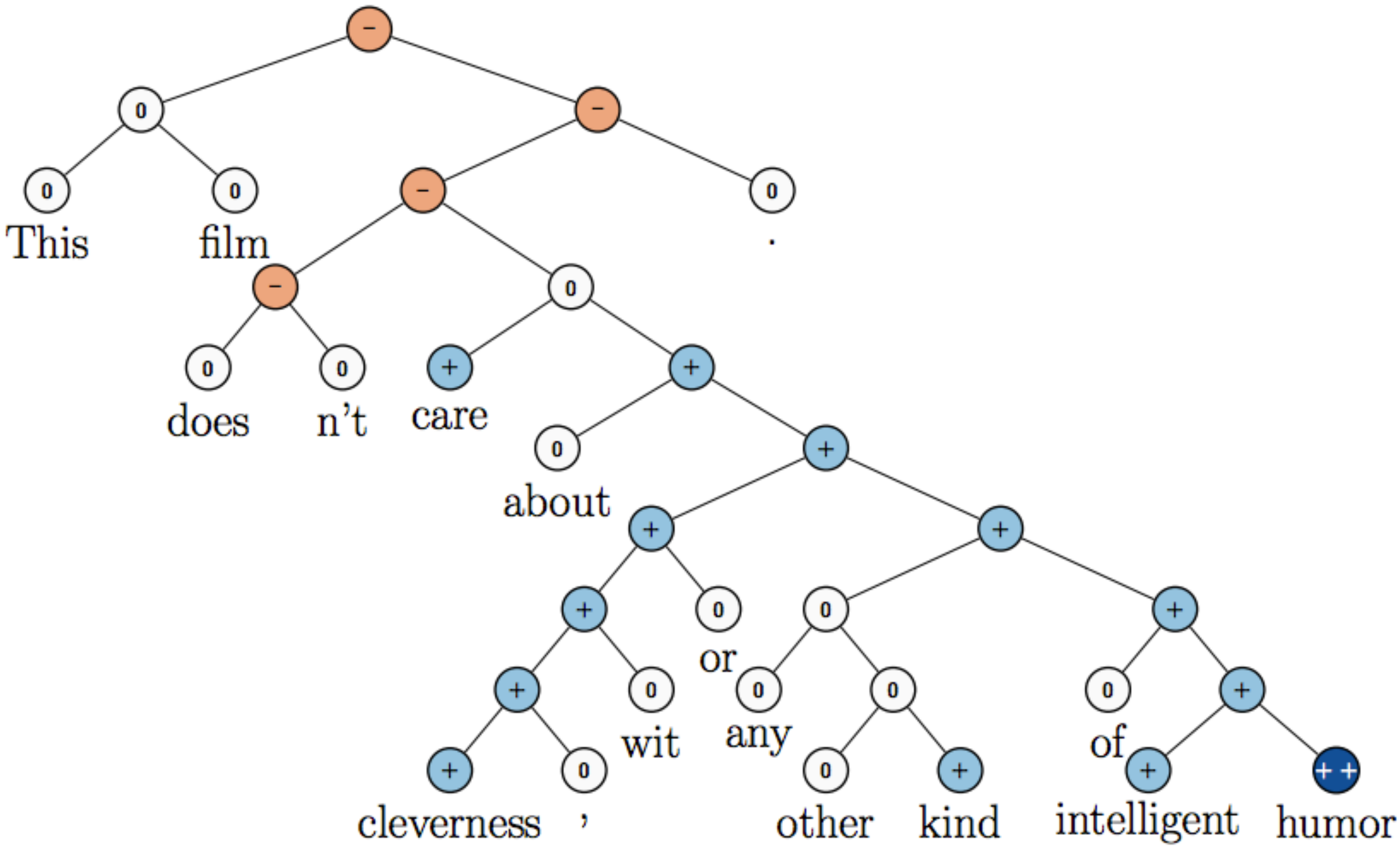


Figure 1: Example of the Recursive Neural Tensor Network accurately predicting 5 sentiment classes, very negative to very positive (– –, –, 0, +, + +), at every node of a parse tree and capturing the negation and its scope in this sentence.

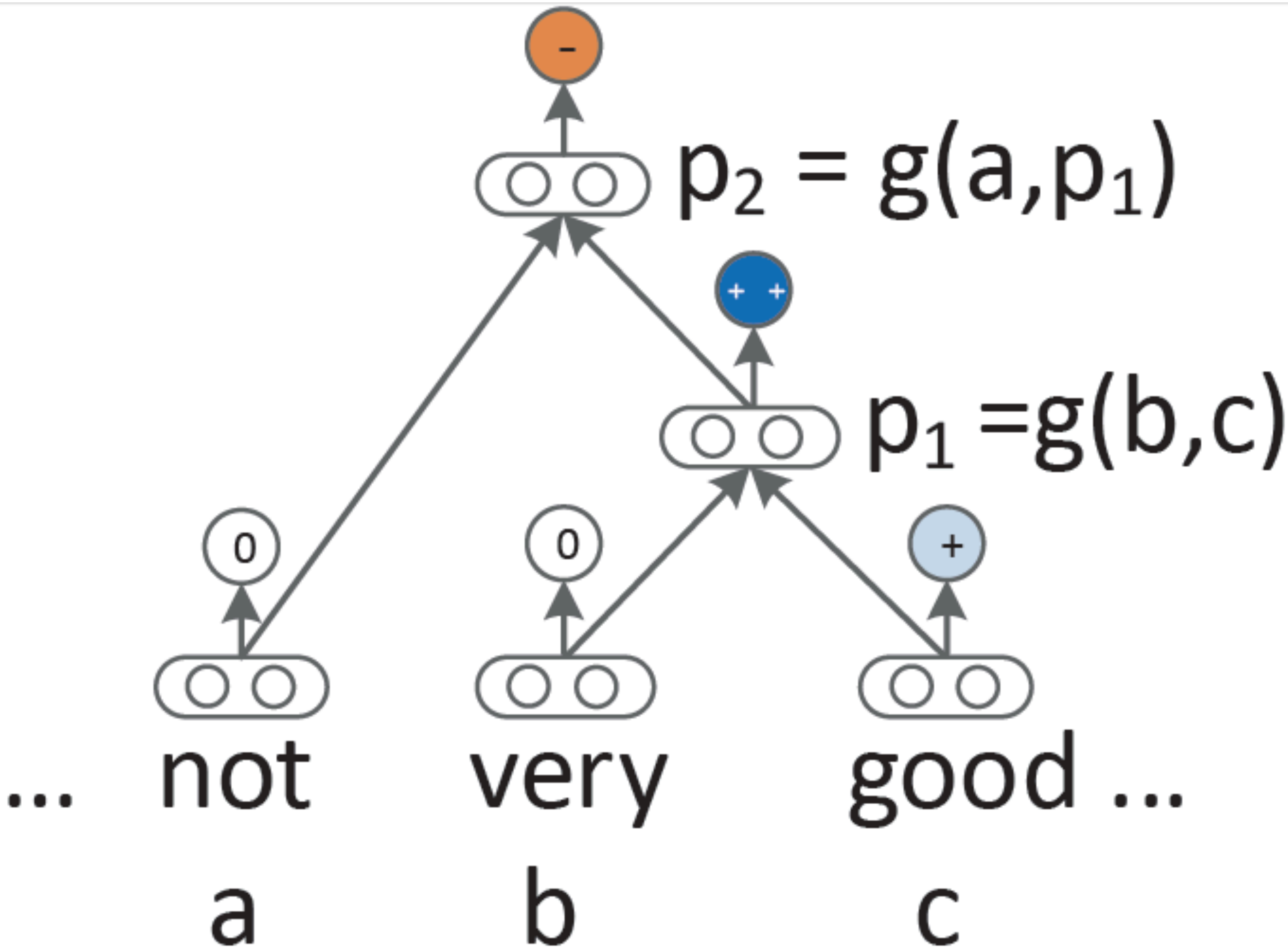
Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

Recursive Neural Tensor Network (RNTN)



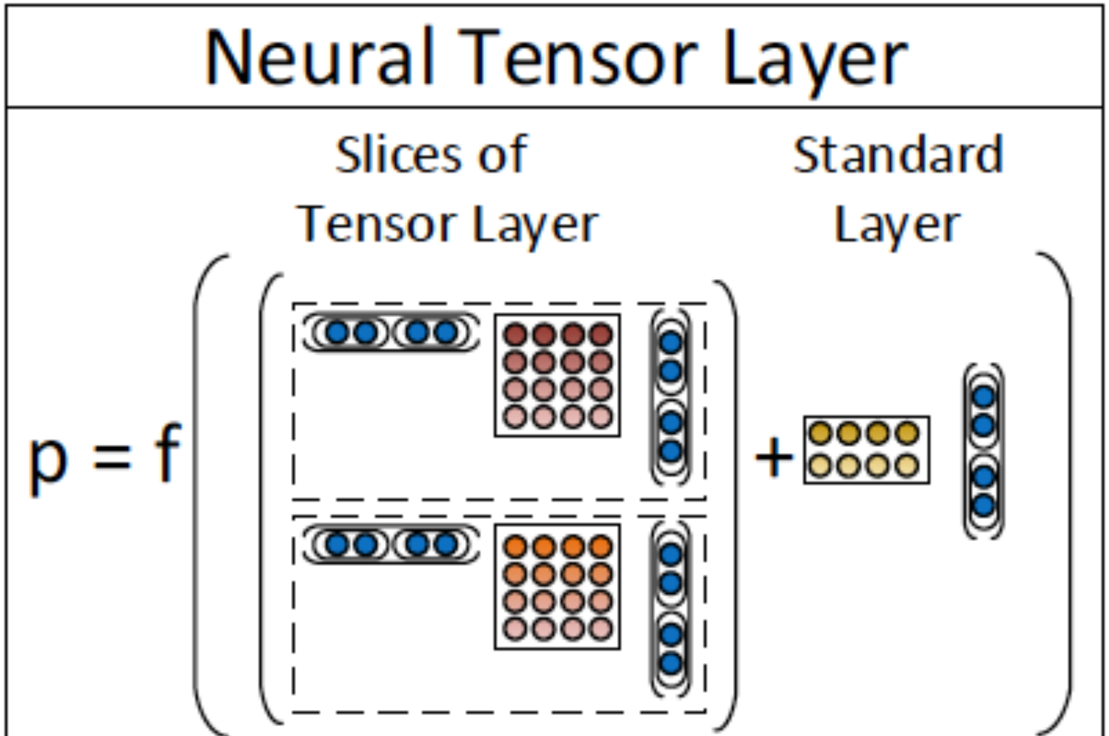
Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

Recursive Neural Network (RNN) models for sentiment



Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

Recursive Neural Tensor Network (RNTN)



$$p = f\left(\begin{bmatrix} b \\ c \end{bmatrix}^T V^{[1:2]} \begin{bmatrix} b \\ c \end{bmatrix} + W \begin{bmatrix} b \\ c \end{bmatrix} \right)$$

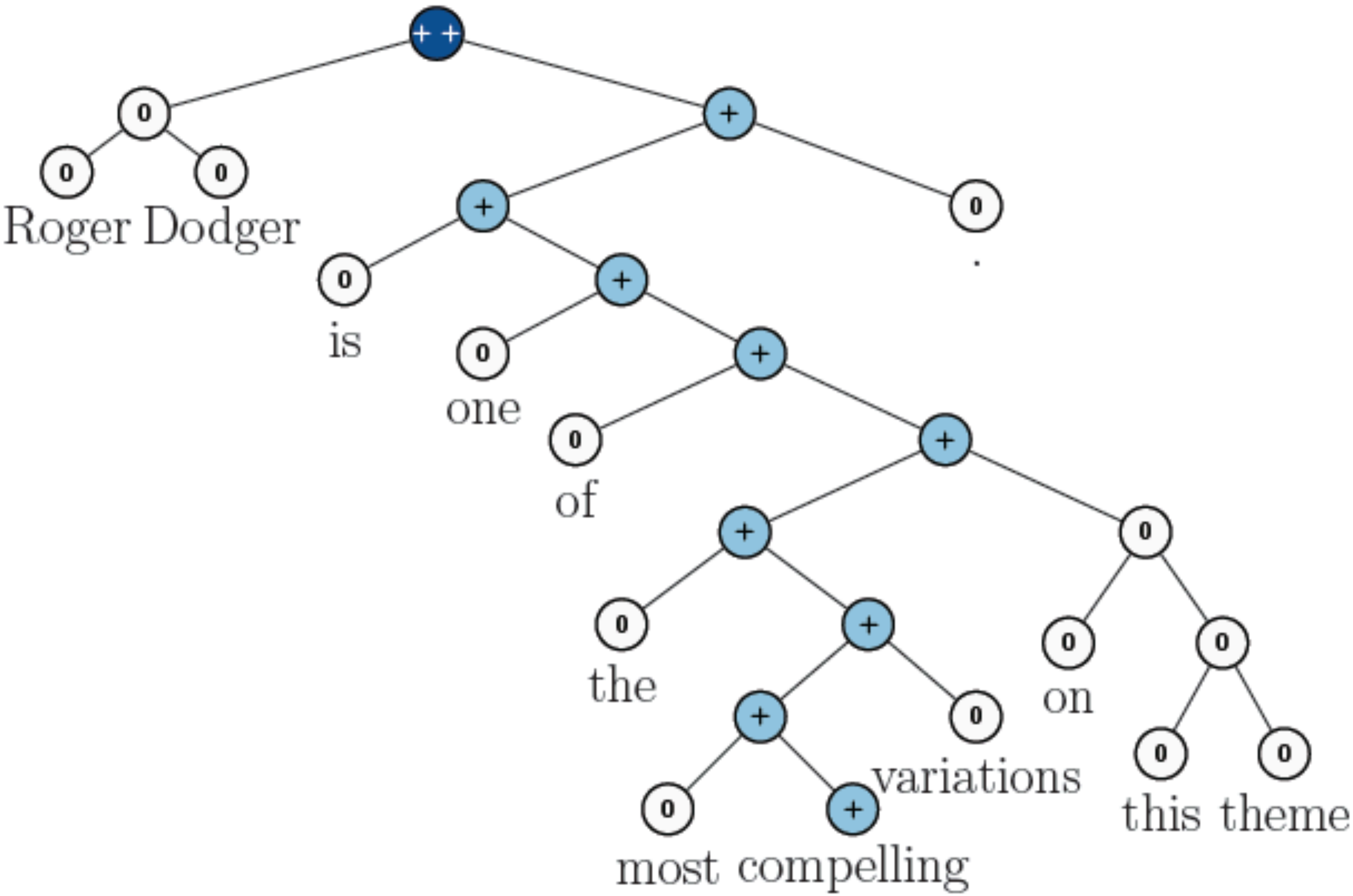
Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

Roger Dodger is one of the most compelling variations on this theme.

Roger Dodger is one of the least compelling variations on this theme.

Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

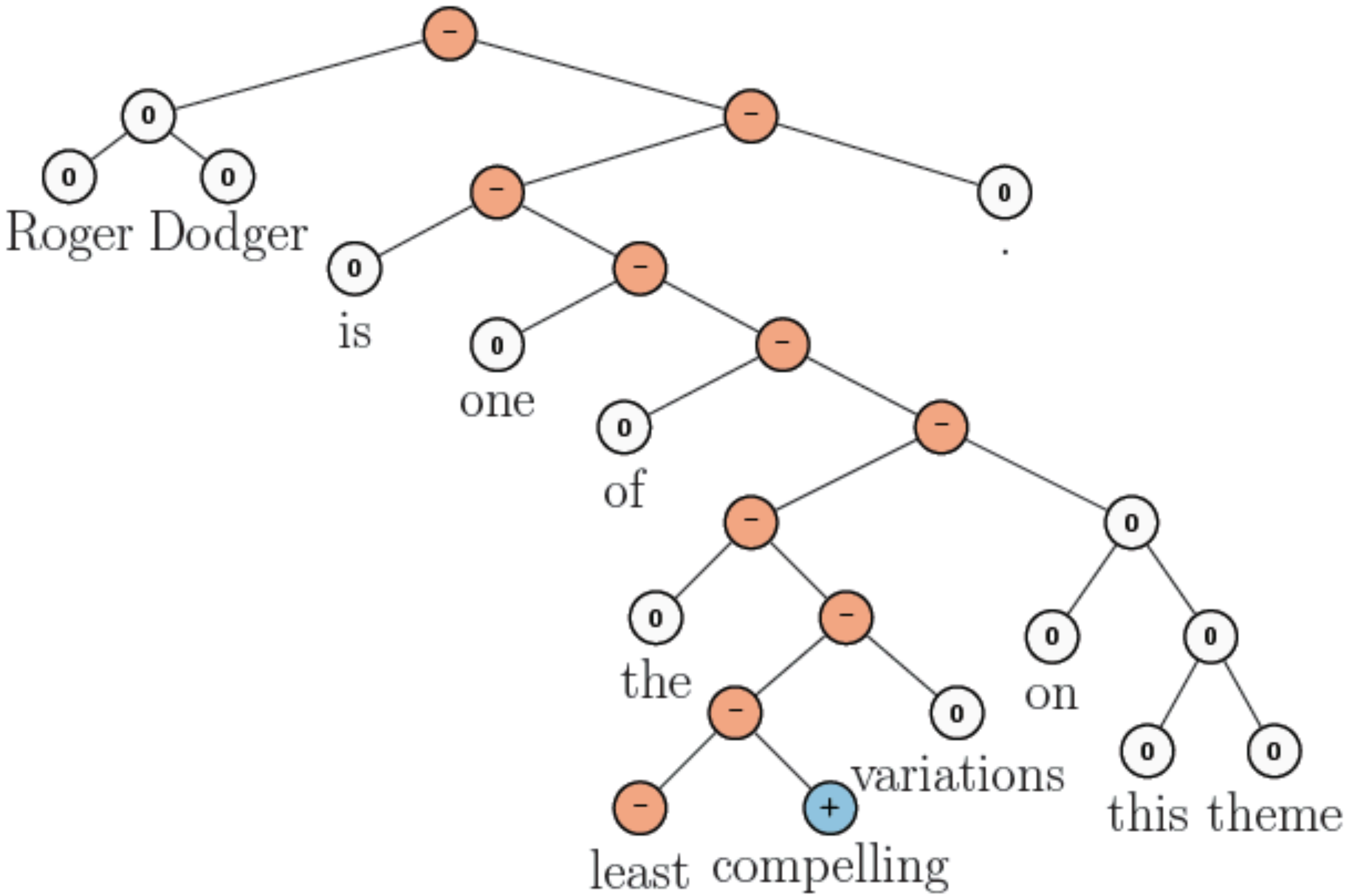
RNTN for Sentiment Analysis



Roger Dodger is one of the most compelling variations on this theme.

Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

RNTN for Sentiment Analysis



Roger Dodger is one of the least compelling variations on this theme.

Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

Accuracy for fine grained (5-class) and binary predictions at the sentence level (root) and for all nodes

Model	Fine-grained		Positive/Negative	
	All	Root	All	Root
NB	67.2	41.0	82.6	81.8
SVM	64.3	40.7	84.6	79.4
BiNB	71.0	41.9	82.7	83.1
VecAvg	73.3	32.7	85.1	80.1
RNN	79.0	43.2	86.1	82.4
MV-RNN	78.7	44.4	86.8	82.9
RNTN	**80.7**	**45.7**	**87.6**	**85.4**

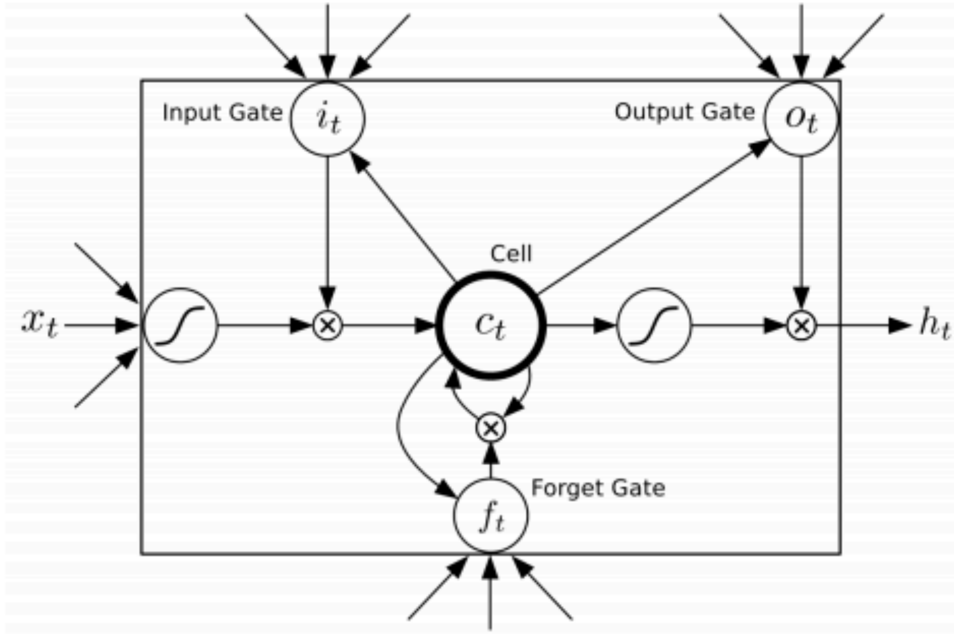
Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

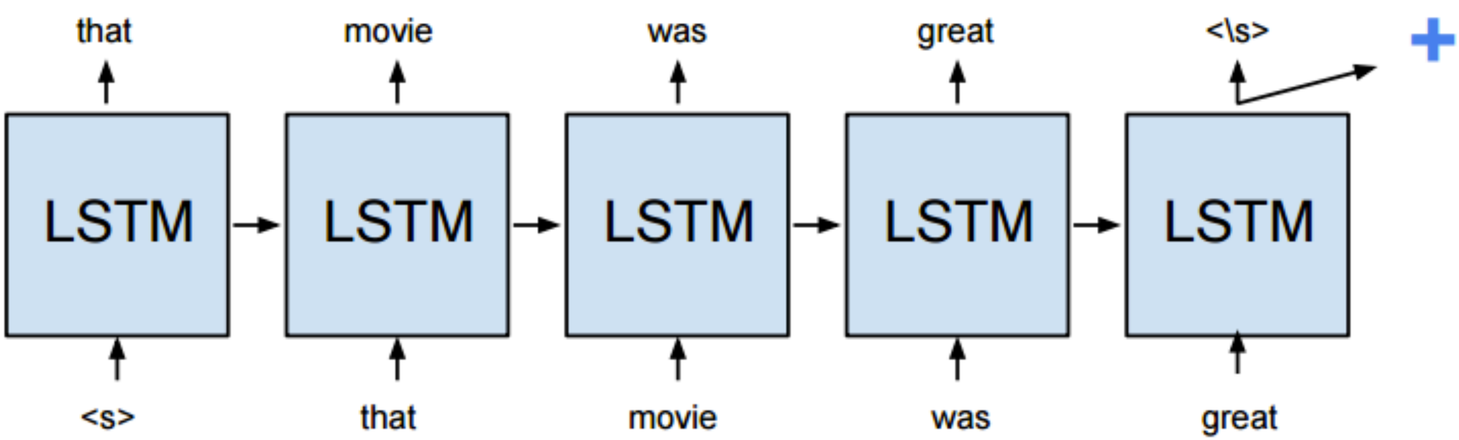
Accuracy of negation detection

Model	Accuracy	
	Negated Positive	Negated Negative
biNB	19.0	27.3
RNN	33.3	45.5
MV-RNN	52.4	54.6
RNTN	**71.4**	**81.8**

Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

Long Short-Term Memory (LSTM)





Source: https://cs224d.stanford.edu/reports/HongJames.pdf

Deep Learning for Sentiment Analysis CNN RNTN LSTM

Model	Fine (5-class)	Binary
DCNN (Blunsom, et al. 2014)	0.485	0.868
RNTN (Socher, et al. 2013)	0.457	0.854
CNN-non-static (Kim, 2014)	0.480	0.872
CNN-multi-channel (Kim, 2014)	0.474	0.881
DRNN w. pretrained word-embeddings (Irsoy and Cardie, 2014)	0.498	0.866
Paragraph Vector (Le and Mikolov. 2014)	0.487	0.878
Dependency Tree-LSTM (Tai, et al, 2015)	0.484	0.857
Constituency Tree-LSTM (Tai, et al, 2015)	0.439	0.820
Constituency Tree-LSTM (Glove vectors) (Tai, et al, 2015)	0.510	0.880
Paragraph Vector	0.391	0.798
LSTM	0.456	0.843
Deep Recursive-NN	0.469	0.847

Source: https://cs224d.stanford.edu/reports/HongJames.pdf

Performance Comparison of Sentiment Analysis Methods

	Method	Data Set	Acc.	Author
Machine Learning	SVM	Movie reviews	86.40%	Pang, Lee[23]
	CoTraining SVM	Twitter	82.52%	Liu[14]
	Deep learning	Stanford Sentiment Treebank	80.70%	Richard[18]
Lexical based	Corpus	Product reviews	74.00%	Turkey
	Dictionary	Amazon's Mechanical Turk	---	Taboada[20]
Cross-lingual	Ensemble	Amazon	81.00%	Wan,X[16]
	Co-Train	Amazon, ITI68	81.30%	Wan,X.[16]
	EWGA	IMDb movie review	>90%	Abbasi,A.
	CLMM	MPQA,NTCIR,ISI	83.02%	Mengi
Cross-domain	Active Learning	Book, DVD, Electronics, Kitchen	80% (avg)	Li, S
	Thesaurus			Bollegala[22]
	SFA			Pan S J[15]

Resources of Opinion Mining

Datasets of Opinion Mining

- Blog06
  - 25GB TREC test collection
  - http://ir.dcs.gla.ac.uk/test collections/access to data.html
- Cornell movie-review datasets
  - http://www.cs.cornell.edu/people/pabo/movie-review-data/
- Customer review datasets
  - http://www.cs.uic.edu/~liub/FBS/CustomerReviewData.zip
- Multiple-aspect restaurant reviews
  - http://people.csail.mit.edu/bsnyder/naacl07
- NTCIR multilingual corpus
  - NTCIR Multilingual Opinion-Analysis Task (MOAT)

Source: Bo Pang and Lillian Lee (2008), "Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval

Lexical Resources of Opinion Mining

- SentiWordnet
  - http://sentiwordnet.isti.cnr.it/
- General Inquirer
  - http://www.wjh.harvard.edu/~inquirer/
- OpinionFinder's Subjectivity Lexicon
  - http://www.cs.pitt.edu/mpqa/
- NTU Sentiment Dictionary (NTUSD)
  - http://nlg18.csie.ntu.edu.tw:8080/opinion/
- Hownet Sentiment
  - http://www.keenage.com/html/c_bulletin_2007.htm

Sentiment Analysis Dictionary

- NTUSD: SD\NTUSD.rar
- HOWNET: SD\HOWNET.rar
- SentiWordNet: SD\SentiWordNet3.rar
- TYCCL Antonym Negation: SD\TYCCL\TYCCL.rar
- DLUTSD : SD\DLUTSD.zip
- IMTKU iCosmeSD: SD\iCosmeSD2014.rar
- IMTKU iMFinanceSD: SD\iMFinanceSD.zip
- IMTKU Antonym: SD\IMTKUAntonym.txt

Example of SentiWordNet

POS ID PosScore NegScore SynsetTerms Gloss

a 00217728 0.75 0 beautiful#1 delighting the senses or exciting intellectual or emotional admiration; "a beautiful child"; "beautiful country"; "a beautiful painting"; "a beautiful theory"; "a beautiful party“

a 00227507 0.75 0 best#1 (superlative of `good') having the most positive qualities; "the best film of the year"; "the best solution"; "the best time for planting"; "wore his best suit“

r 00042614 0 0.625 unhappily#2 sadly#1 in an unfortunate way; "sadly he died before he could see his grandchild“

r 00093270 0 0.875 woefully#1 sadly#3 lamentably#1 deplorably#1 in an unfortunate or deplorable manner; "he was sadly neglected"; "it was woefully inadequate“

r 00404501 0 0.25 sadly#2 with sadness; in a sad manner; "`She died last night,' he said sadly"

《知網》情感分析用詞語集（beta版）

- “中英文情感分析用詞語集”
  - 包含詞語約 17887
- “中文情感分析用詞語集”
  - 包含詞語約 9193
- “英文情感分析用詞語集”
  - 包含詞語 8945

Source: http://www.keenage.com/html/c_bulletin_2007.htm

中文情感分析用詞語集

中文正面情感詞語	836
中文負面情感詞語	1254
中文正面評價詞語	3730
中文負面評價詞語	3116
中文程度級別詞語	219
中文主張詞語	38
Total	9193

Source: http://www.keenage.com/html/c_bulletin_2007.htm

中文情感分析用詞語集

- “正面情感”詞語
  - 如：
    愛，讚賞，快樂，感同身受，好奇，喝彩，魂牽夢縈，嘉許 ...
- “負面情感”詞語
  - 如：
    哀傷，半信半疑，鄙視，不滿意，不是滋味兒，後悔，大失所望 ...

Source: http://www.keenage.com/html/c_bulletin_2007.htm

中文情感分析用詞語集

- “正面評價”詞語
  - 如：
    不可或缺，部優，才高八斗，沉魚落雁，催人奮進，動聽，對勁兒 ...
- “負面評價”詞語
  - 如：
    醜，苦，超標，華而不實，荒涼，混濁，畸輕畸重，價高，空洞無物 ...

Source: http://www.keenage.com/html/c_bulletin_2007.htm

中文情感分析用詞語集

- “程度級別”詞語
  - 1. “極其|extreme / 最|most”
    - 非常，極，極度，無以倫比，最為
  - 2. “很|very”
    - 多麼，分外，格外，著實
  - …
- “主張”詞語
  - 1. {perception|感知}
    - 感覺，覺得，預感
  - 2. {regard|認為}
    - 認為，以為，主張

Source: http://www.keenage.com/html/c_bulletin_2007.htm

Fake Review

Opinion Spam Detection

Opinion Spam Detection

- Opinion Spam Detection: Detecting Fake Reviews and Reviewers
  - Spam Review
  - Fake Review
  - Bogus Review
  - Deceptive review
  - Opinion Spammer
  - Review Spammer
  - Fake Reviewer
  - Shill (Stooge or Plant)

Source: http://www.cs.uic.edu/~liub/FBS/fake-reviews.html

Opinion Spamming

- Opinion Spamming
  - "illegal" activities
    - e.g., writing fake reviews, also called shilling
  - try to mislead readers or automated opinion mining and sentiment analysis systems by giving undeserving positive opinions to some target entities in order to promote the entities and/or by giving false negative opinions to some other entities in order to damage their reputations.

Source: http://www.cs.uic.edu/~liub/FBS/fake-reviews.html

Forms of Opinion spam

- fake reviews (also called bogus reviews)
- fake comments
- fake blogs
- fake social network postings
- deceptions
- deceptive messages

Source: http://www.cs.uic.edu/~liub/FBS/fake-reviews.html

Fake Review Detection

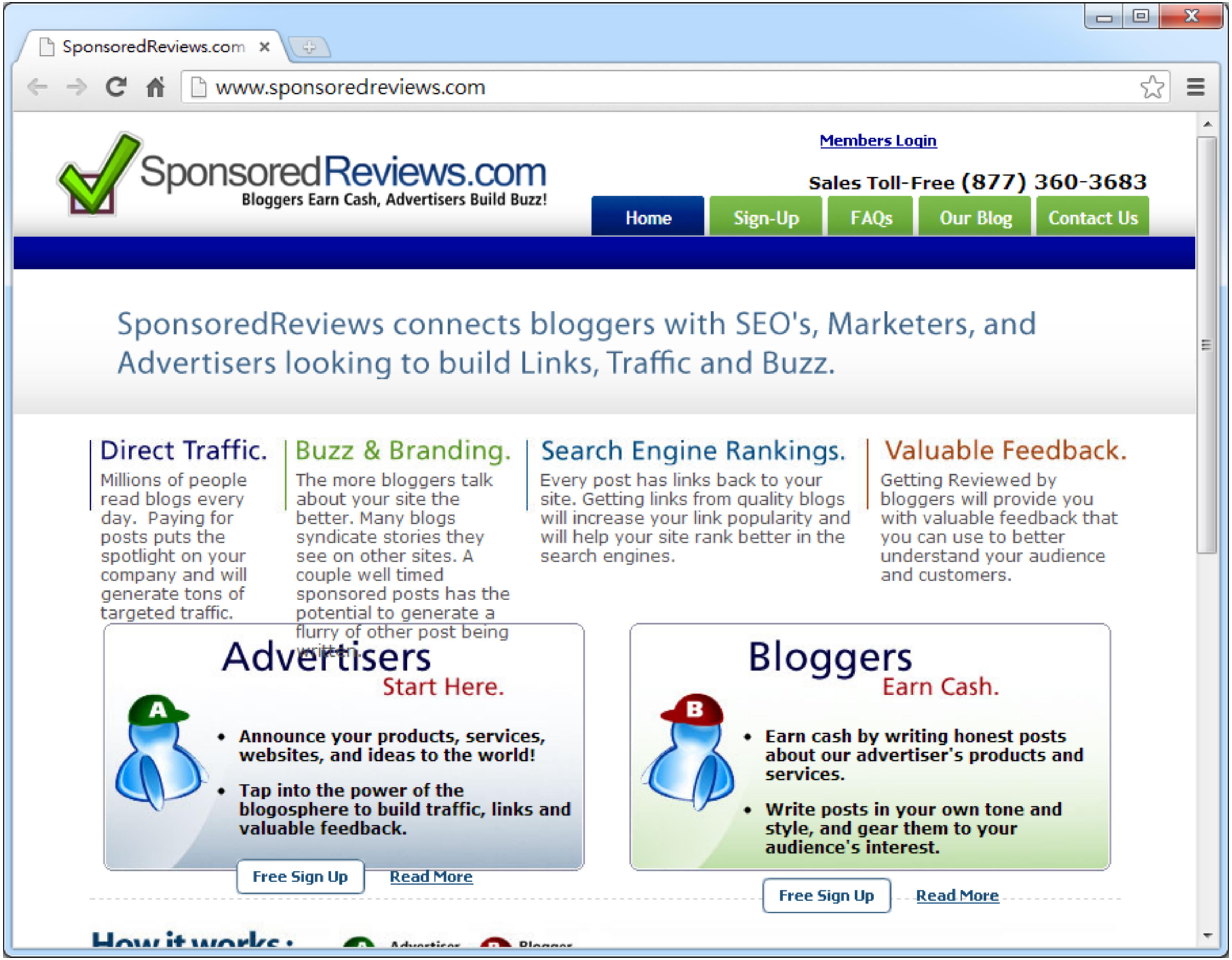
- Methods
  - supervised learning
  - pattern discovery
  - graph-based methods
  - relational modeling
- Signals
  - Review content
  - Reviewer abnormal behaviors
  - Product related features
  - Relationships

Source: http://www.cs.uic.edu/~liub/FBS/fake-reviews.html

Professional Fake Review Writing Services (some Reputation Management companies)

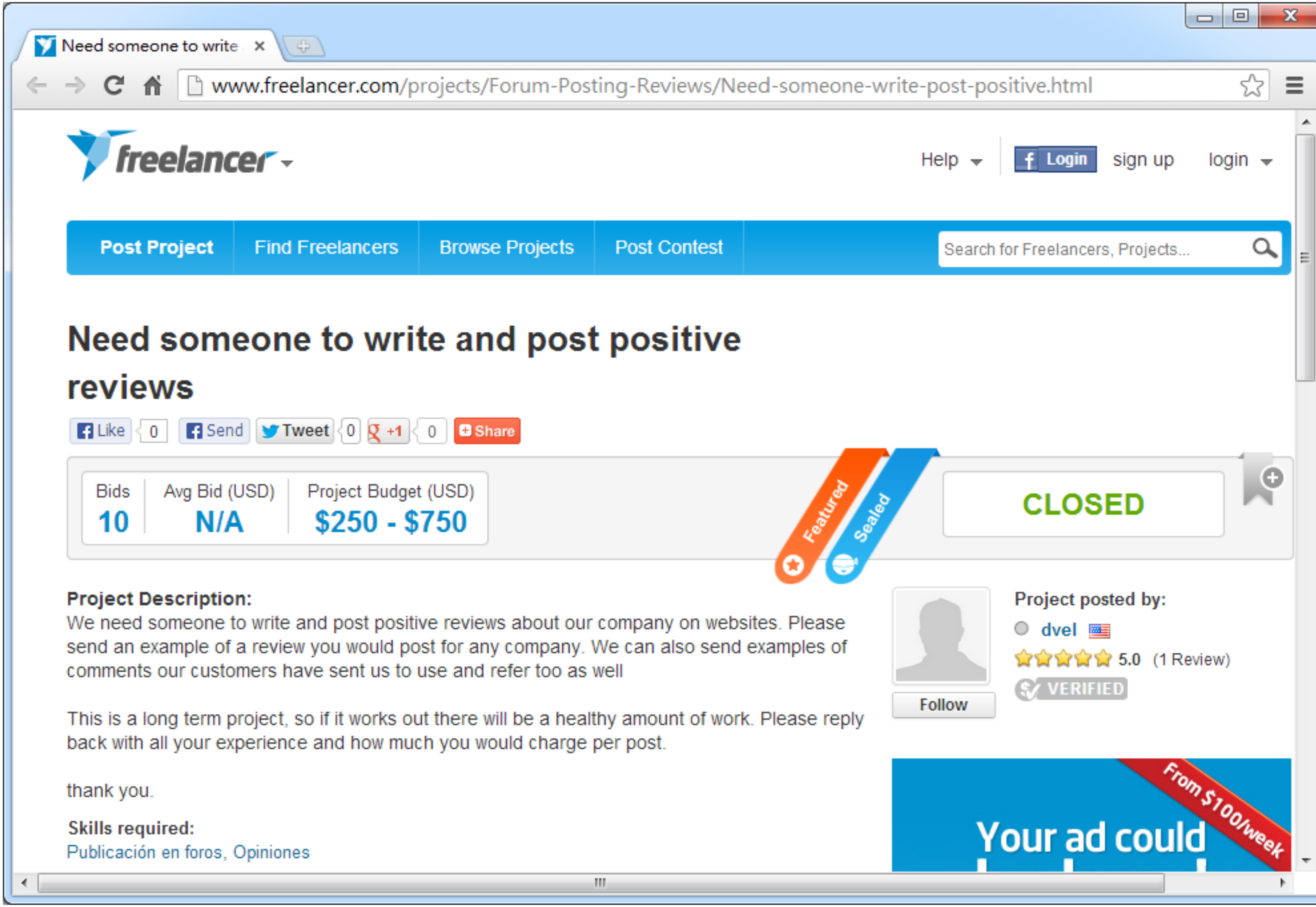
- Post positive reviews
- Sponsored reviews
- Pay per post
- Need someone to write positive reviews about our company (budget: $250-$750 USD)
- Fake review writer
- Product review writer for hire
- Hire a content writer
- Fake Amazon book reviews (hiring book reviewers)
- People are just having fun (not serious)

Source: http://www.cs.uic.edu/~liub/FBS/fake-reviews.html



Source: http://www.sponsoredreviews.com/

Source: https://payperpost.com/



Source:http://www.freelancer.com/projects/Forum-Posting-Reviews/Need-someone-write-post-positive.html

Deceptive Review Spam Detection Techniques

- Supervised learning techniques
  - labeled data
- Unsupervised learning techniques
  - unlabeled data
- Semi-supervised learning techniques
  - minimum labeled data

Source: Rout, Jitendra Kumar, Smriti Singh, Sanjay Kumar Jena, and Sambit Bakshi. "Deceptive review detection using labeled and unlabeled data." *Multimedia Tools and Applications* (2017): 1-25.

Comparison of deceptive review spam detection techniques based on labeled data

Authors	Key concept	Features	Learner	Result
Jindal and Liu (2008)	Text duplication	Review, reviewer and product centric	Logistic regression	78 %(accuracy)
Lai et al. (2010)	Text similarity	Review text	SVM	81 %(precision)
Algur et al. (2010)	Product feature similarity	Product features	Cosine similarity	43.6 %(precision)
Ott et al. (2011)	Content similarity	LIWC + Bigram	SVM	89.6(accuracy)
Ott et al. (2013)	Content review(Taking negative reviews only)	n-gram	SVM	86 %(accuracy)
Mukherjee et al. (2013)	Content similarity	Behavioral+ Bigrams	SVM	86.1 %(accuracy)
Shojaee et al. (2013)	Stylometric	Lexical and syntactical	SVM	84 %(F-score)
Long et al. (2014)	Ontology	Ontological features	Conditional filtering	75 %(precision)
Rout et al. (2017)	Content similarity	Sentiment Score,	SVM	88.71 %(accuracy)
Rout et al. (2017)	and sentiment	Lingustic features	Naive Bayes	91.9 %(accuracy)
Rout et al. (2017)	polarity	and unigram	Decision Tree	92.11 %(accuracy)

Source: Rout, Jitendra Kumar, Smriti Singh, Sanjay Kumar Jena, and Sambit Bakshi.
"Deceptive review detection using labeled and unlabeled data." *Multimedia Tools and Applications* (2017): 1-25.

Comparison of deceptive review spam detection techniques based on unlabeled data

Author	Key concept	Dataset	Features	Approach
Wu et al. (2010)	Distortion used to separate out true positives from false positives.	Irish tripAdvisor Data	Proportion of positive singletons (PPS) and concentration of positive singletons(CPS)	Clustering
Raymond et al. (2011)	Semantic content overlapping among reviews	Amazon review dataset	cosine similarity measure	Clustering
Mukherjee et al. (2013)	Difference in behavioral distributions of spammers and non-spammers	Amazon review dataset	Author features, Review features	Unsupervised clustering in bayesian setting
Akoglu et al. (2013)	Network effect among reviewer and products	Software marketplace (SWM) dataset	Honesty and goodness of products, Review scores	Graph clustering (Cross-Association Clustering)
Rout et al. (2017)	Difference in behavioral patterns of reviews	Amazon cell phone reviews dataset	Review data, Reviewer data and product information	Clustering

Source: Rout, Jitendra Kumar, Smriti Singh, Sanjay Kumar Jena, and Sambit Bakshi.
"Deceptive review detection using labeled and unlabeled data." *Multimedia Tools and Applications* (2017): 1-25.

Comparison of deceptive review spam detection techniques based on minimum labeled data

Author	Key concept	Dataset	Approach	Features	Result
Li et al. (2011)	Review spammer consistently writes spam	Product reviews obtained from Epinions	Co-training algorithm	Review related features (Content, sentiment, product, data features) and reviewer related features (Profile and Behavioral Features)	0.631 (F-Score)
Fusilier et al. (2013)	Learning from positive example and set of unlabeled data	Ott's hotel review dataset	PU-learning	n-gram	0.84 (F-score)
Ren et al. (2014)	Based on some truthful reviews and a lot of unlabeled reviews to build an accurate classifier	Ott's hotel review dataset	Mixing population and individual property PU learning(MPIPUL)	Similarity weights	83.91 % (Accuracy)

Source: Rout, Jitendra Kumar, Smriti Singh, Sanjay Kumar Jena, and Sambit Bakshi. "Deceptive review detection using labeled and unlabeled data." *Multimedia Tools and Applications* (2017): 1-25.

References

- Chuck Hemann and Ken Burbary (2013), Digital Marketing Analytics: Making Sense of Consumer Data in a Digital World, Que.
- Bing Liu (2015), Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Cambridge University Press
- Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” 2nd Edition, Springer. http://www.cs.uic.edu/~liub/WebMiningBook.html
- Bing Liu (2013), Opinion Spam Detection: Detecting Fake Reviews and Reviewers, http://www.cs.uic.edu/~liub/FBS/fake-reviews.html
- Bo Pang and Lillian Lee (2008), "Opinion mining and sentiment analysis,” Foundations and Trends in Information Retrieval 2(1-2), pp. 1–135, 2008.
- Wiltrud Kessler (2012), Introduction to Sentiment Analysis, http://www.ims.uni-stuttgart.de/~kesslewd/lehre/sentimentanalysis12s/introduction_sentimentanalysis.pdf
- Z. Zhang, X. Li, and Y. Chen (2012), "Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews," ACM Trans. Manage. Inf. Syst. (3:1) 2012, pp 1-23.
- Efraim Turban, Ramesh Sharda, Dursun Delen (2011), Decision Support and Business Intelligence Systems, Ninth Edition, 2011, Pearson.
- Guandong Xu, Yanchun Zhang, Lin Li (2011), Web Mining and Social Networking: Techniques and Applications, 2011, Springer

References

- Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In the 26th International Conference on Computational Linguistics (COLING), Osaka. 2016.
- Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts (2013), "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank," In Proceedings of the conference on empirical methods in natural language processing (EMNLP), vol. 1631, p. 1642 http://nlp.stanford.edu/~socherr/EMNLP2013_RNTN.pdf
- Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.
- Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, vol 139, no. 11, 2016. pp.5-15.
- Jesus Serrano-Guerrero, Jose A. Olivas, Francisco P. Romero, and Enrique Herrera-Viedma (2015), "Sentiment analysis: A review and comparative analysis of web services," Information Sciences, 311, pp. 18-38.
- Steven Struhl (2015), Practical Text Analytics: Interpreting Text and Unstructured Data for Business Intelligence (Marketing Science), Kogan Page
- Jitendra Kumar Rout, Smriti Singh, Sanjay Kumar Jena, and Sambit Bakshi (2017), "Deceptive review detection using labeled and unlabeled data." Multimedia Tools and Applications (2017): 1-25.