CMPE 561:Natural Language Processing

RESEARCH PROJECT

SENTIMENT ANALYSIS

KADRIYE YASEMIN USTA
OĞUZHAN MURAT ÇAKMAK

Agenda

Sentiment

Sentiment := <Holder, Target, **Polarity**, Auxiliary>

- **Holder:** who expresses the sentiment

- **Target:** what/whom the sentiment is expressed to

- **Polarity:** the nature of the sentiment (positive, negative, or neutral)

- **Auxiliary:** strength, summary, confidence, time

The games in iPhone 4s are pretty funny!

Feature/Aspect

Target

Positive

Polarity

Thumb up or down

Positive or negative

Love or hate

01

Introduction

< 5 >

Sentiment Analysis

Computational study of opinions, sentiments, appraisal, and emotions expressed in text
Ex: Reviews, blogs, discussions, microblogs, social networks

Basic tasks:

📌 Holder detection

📌 Target recognition

📌 **Sentiment (Polarity) classification**

📌 Opinion summarization

📌 Opinion spam detection

01

Introduction

< 6 >

Application Domains

- **Politics/political science**: to find public opinions about political candidates and issues

- **Law/policy making:** to examine public opinions about law candidates and issues

- **Sociology:** to understand Idea propagation through groups

- **Psychology:** to investigate dream sentiment analysis

Types of Sentiment Classification

📌 Lexicon Based Sentiment Classification

    📌 Dictionary-based Approach

    📌 Corpora-based Approach

        📌 Statistical and Semantic

📌 Machine Learning Based Sentiment Classification

    📌 Unsupervised Learning

    📌 Supervised Learning

   📌 Decision Tree,Linear,Rule-based,Probabilistic
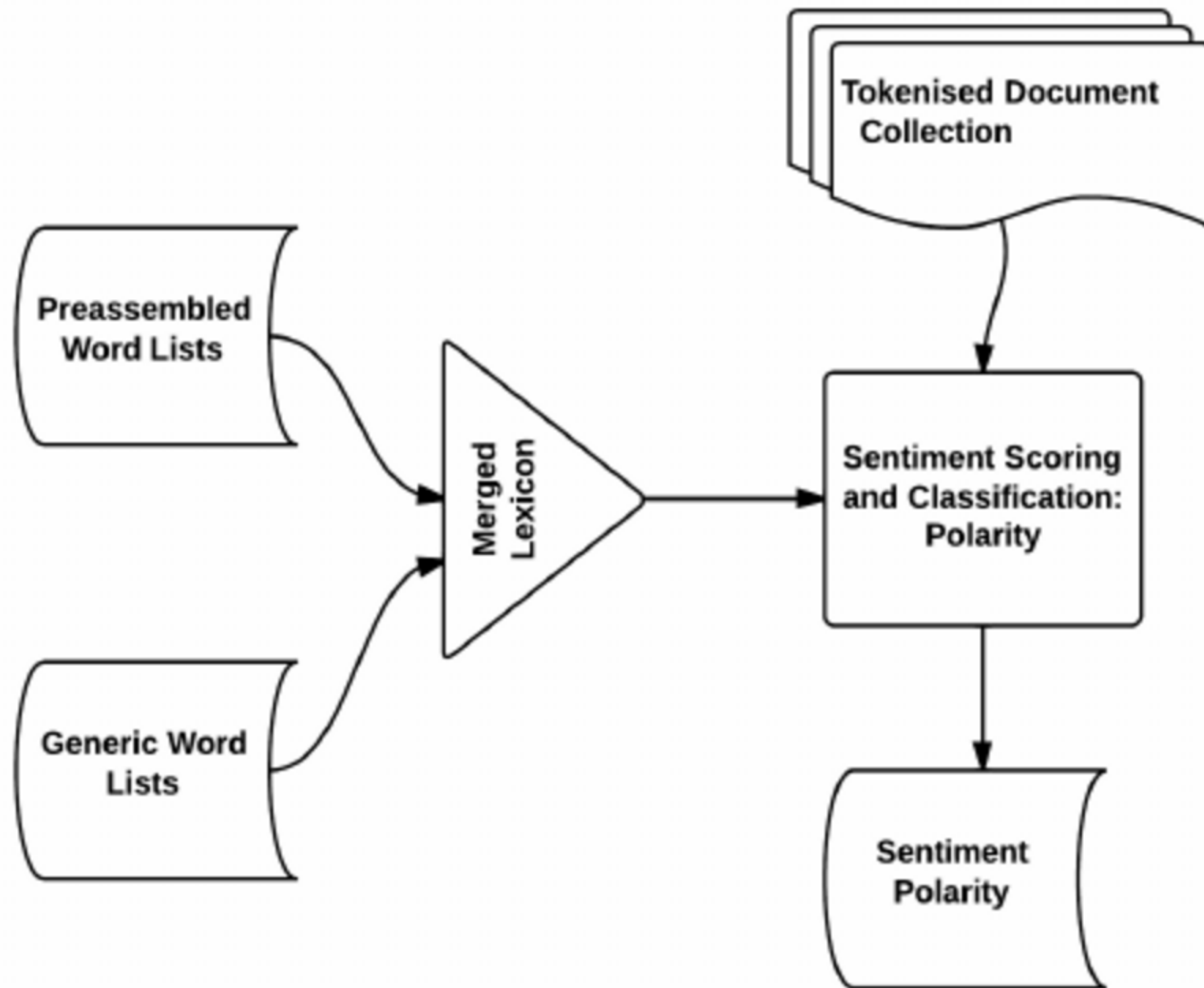
Lexicon-based Sentiment Classification

- Use the dominant polarity of the opinion words (ADJ, ADV and VERB) in the sentence to determine its polarity

- If positive/negative opinion prevails, the opinion sentence is regarded as positive/negative

Optimization Methods:
Lexicon + Grammar Rule + Inference
Lexicon + Counting
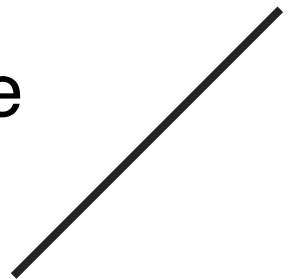
General Work-Flow Diagram of Lexicon-based Classification
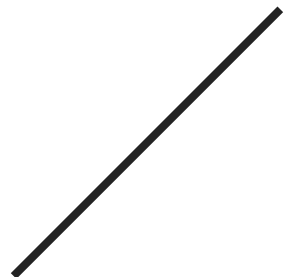
02

Sentiment Classification

< 10 >

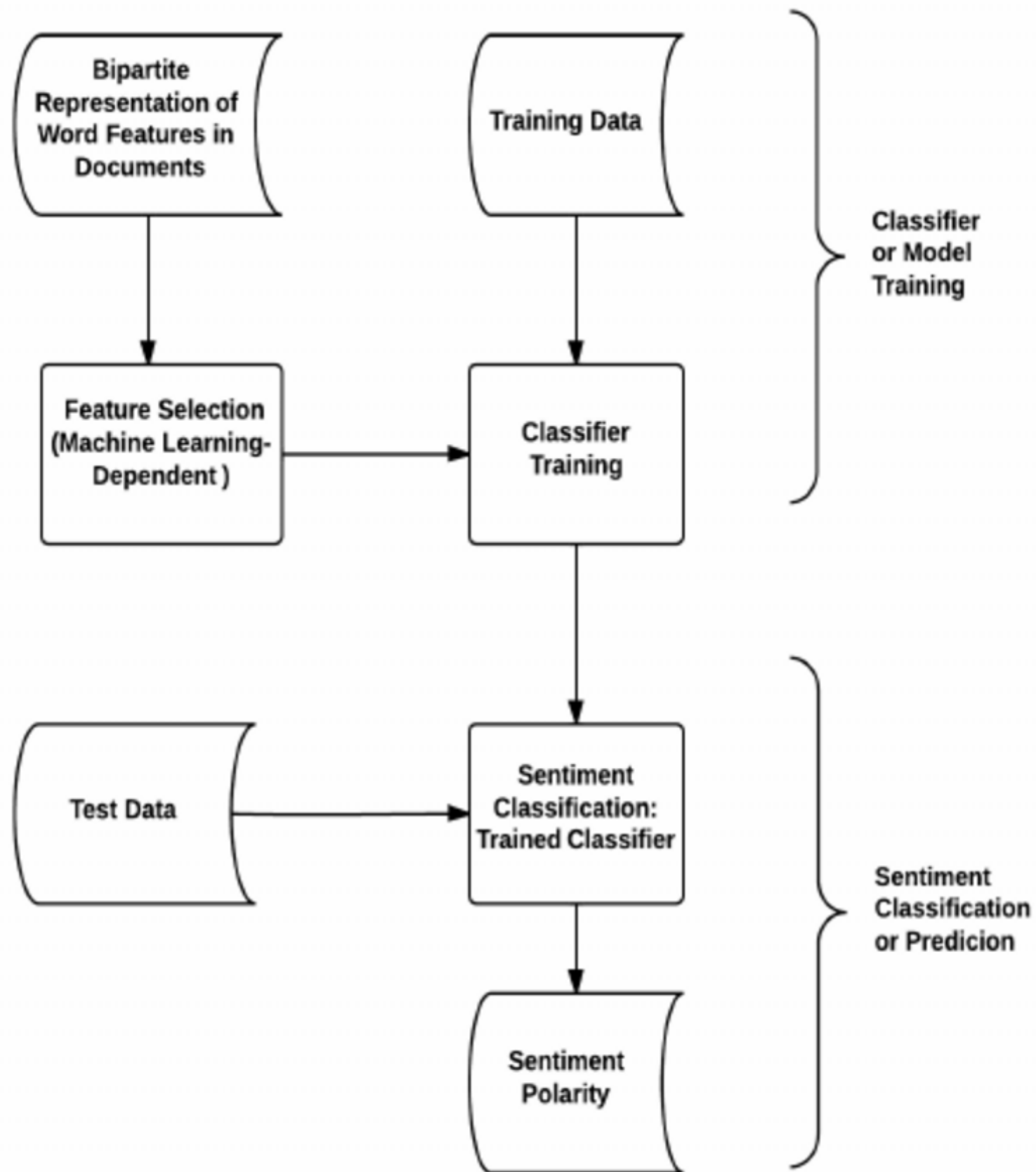# Machine Learning-based Sentiment Classification

📌Treat sentiment classification simply as a special case of topic-based categorization

Ex : With the two "topics" being <u>positive</u> sentiment and <u>negative</u> sentiment

Common Method
Data + Feature + Model

General Work-Flow Diagram of Machine Learning-based Classification

Machine Learning-based vs Lexicon-based Classification

- No explicit result
  - The success depends on details, data, domains, scenarios
- Observation
  - Lexicon-based: simple, intuitive, and understandable
  - ML-based: data-driven, uninterruptible, and large-scaled
- Briefly,
  - Sentence: Lexicon-based is better
  - Document: ML-based is better

Current Studies

- *Turney(2002)* an unsupervised learning algorithm for classifying reviews as recommended or not recommended

- Starts with tagging phrases, then calculates PMI(Point-wise Mutual Information)

Current Studies

- *Pang and Lee(2008)* classify documents not by topic, by overall sentiment

Current Studies

- *Socher et al.(2013)* uses Deep Learning

- Introduce a sentiment treebank

- Recursive Neural Tensor Network(RNTN)

- Very good results

Tools

General Inquirer :
A computer-assisted approach for content analysis of textual data.

- Classifies as positive and negative

04

Tools&Corpora

< 16 >

Tools

LIWC(Linguistic Inquiry and Word Count)

- The program includes the primary text analysis module along with a group of built-in dictionaries.

- Reads computer-readable form.

- Contains 6400 words, word stems, and selected emoticons

- Affective(negative, positive) and cognitive(tentative, inhibition)

Tools

WordNet is a lexical database.

- Include sets of synonyms called senses, short definitions, use cases, and various relations among the words.

04

Tools&Corpora

< 18 >

Tools

SentiWordNet is a strategy for examination of definition related to synnets of the WordNet data set.

- The synnet has three numerical score including objective, negative, and positive polarity.

Tools

SentiTurkNet is the first Turkish polarity resource.

- Three scores: positivity, negativity and objectivity

- Results for Turkish better than SentiWordNet

Corpora

- **Pang & Lee Data Sets**
  Polarity(positive or negative)
  Subjective Rate(from 1 to 5)
  Subjective or Objective

- **Blitzer et al. Multi-domain sentiment analysis**
  Amazon data
  Classify as Pos or Neg

- **MPQA**
  Private states(beliefs, emotions, sentiments, speculations

Current Challenges

- Sentiments words and phrases(lexicon) is necessary but not sufficient for high performance sentiment analysis

05

Challenges

< 22 >

Current Challenges

- A positive or negative opinion word may have inverse introductions in diverse application areas.

  - "This camera <u>sucks</u>"

  - "This vacuum cleaner really <u>sucks</u>"

Current Challenges

- Sarcastic sentences with or without sentiments words are hard to handle

- "What a great car! It stopped working in two days."

Current Challenges

- A sentence containing opinion words may not express any sentiment.

- e.g "Can you tell me which Sony camera is <u>good</u>?"

- "Does anyone know how to repair this <u>terrible printer</u>?"

Current Challenges

- Many sentences without sentiment words can likewise suggest opinions.

- "This washer uses a lot of water"

05
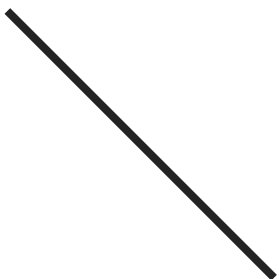
Challenges

< 26 >

Sentiment analysis is a difficult task

The difficulty increases with the nuance and complexity of opinions expressed

There are many different methods for sentiment analysis
- Lexicon-based
- Machine Learning-based

06

Conclusions

< 27 >

THANKS FOR LISTENING