WikiSent : Weakly Supervised Sentiment Analysis Through Extractive Summarization With Wikipedia

Subhabrata Mukherjee, Pushpak Bhattacharyya

Dept. of Computer Science and Engineering,

IIT Bombay

European Conference on Machine Learning

ECML PKDD 2012,

Bristol, U.K., 24-28 Sept, 2012

World Knowledge

- Extensive world knowledge required to perform analysis of reviews
- Distinguish between **What's** and **About's** of the review
- Filter out concepts irrelevant to the reviewer opinion about the movie
- Retain only the objects of interest and corresponding opinions

A Negative Review

- best remembered for his understated performance as dr. hannibal lecter in michael mann's forensics thriller , manhunter , scottish character actor brian cox brings something special to every movie he works on .
- usually playing a bit role in some studio schlock (he dies halfway through the long kiss goodnight) , he's only occasionally given something meaty and substantial to do .
- if you want to see some brilliant acting , check out his work as a dogged police inspector opposite frances mcdormand in ken loach's hidden agenda .
- cox plays the role of big john harrigan in the disturbing new indie flick l . i . e . , which lot 47 picked up at sundance when other distributors were scared to budge
- big john feels the love that dares not speak its name , but he expresses it through seeking out adolescents and bringing them back to his pad .
- what bothered some audience members was the presentation of big john in an oddly empathetic light .
- he's an even-tempered , funny , robust old man who actually listens to the kids' problems (as opposed to their parents and friends , both caught up in the high-wire act of their own confused lives .)
- he'll have sex-for-pay with them only after an elaborate courtship , charming them with temptations from the grown-up world .

Facets of a Movie Review

- General Perception about the Crew
- Objective Facts about the Crew and Movies
- Past Performance of the Crew and Movies
- Expectations from the Movie or Crew
- Movie Plot
- Opinion about the Characters in the Movie
- Characteristics of a movie or genre
- **Opinion about the Movie and Crew**
- Unrelated Category

Facets of a Movie Review

- General Perception about the Crew
- Objective Facts about the Crew and Movies
- Past Performance of the Crew and Movies
- Expectations from the Movie or Crew
- Movie Plot
- Opinion about the Characters in the Movie
- Characteristics of a movie or genre
- **Opinion about the Movie and Crew**
- Unrelated Category

- *John Travolta is considered by many to be a has-been, or a one-hit wonder*

…

- *Leornardo DeCaprio is an awesome actor.*

Facets of a Movie Review

- General Perception about the Crew
- Objective Facts about the Crew and Movies
- Past Performance of the Crew and Movies
- Expectations from the Movie or Crew
- Movie Plot
- Opinion about the Characters in the Movie
- Characteristics of a movie or genre
- **Opinion about the Movie and Crew**
- Unrelated Category

- *Born into a family of thespians -- parents Roger Winslet and Sally Bridges-Winslet were both stage actors, maternal grandparents Oliver and Linda*
- *Bridges ran the Reading Repertory Theatre, and uncle Robert Bridges was a fixture in London's West End theatre district – Kate Winslet came into her talent at an early age.*

Facets of a Movie Review

- General Perception about the Crew
- Objective Facts about the Crew and Movies
- Past Performance of the Crew and Movies
- Expectations from the Movie or Crew
- Movie Plot
- Opinion about the Characters in the Movie
- Characteristics of a movie or genre
- **Opinion about the Movie and Crew**
- Unrelated Category

- *Born into a family of thespians -- parents Roger Winslet and Sally Bridges-Winslet*

- *The role that transformed Winslet from art house attraction to international star was Rose DeWitt Bukater, the passionate, rosy-cheeked aristocrat in James Cameron's Titanic (1997).*

age.

Facets of a Movie Review

- General Perception about the Crew
- Objective Facts about the Crew and Movies
- Past Performance of the Crew and Movies
- Expectations from the Movie or Crew
- Movie Plot
- Opinion about the Characters in the Movie
- Characteristics of a movie or genre
- **Opinion about the Movie and Crew**
- Unrelated Category

- *Born into a family of thespians -- parents Roger Winslet and Sally Bridges-Winslet*

- *I cancelled the date with my girlfriend just to watch my favorite star featuring in this movie.*

age.

Facets of a Movie Review

- General Perception about the Crew
- Objective Facts about the Crew and Movies
- Past Performance of the Crew and Movies
- Expectations from the Movie or Crew
- Movie Plot
- Opinion about the Characters in the Movie
- Characteristics of a movie or genre
- **Opinion about the Movie and Crew**
- Unrelated Category

- *Born into a family of thespians -- parents Roger Winslet and Sally Bridges-Winslet*
- *L.I.E. stands for Long Island Expressway, which slices through the strip malls and middle-class homes of suburbia. Filmmaker Michael Cuesta uses it as a (pretty transparent) metaphor of dangerous escape for his 15-year old protagonist, Howie (Paul Franklin Dano).*

age.

Facets of a Movie Review

- General Perception about the Crew
- Objective Facts about the Crew and Movies
- Past Performance of the Crew and Movies
- Expectations from the Movie or Crew
- Movie Plot
- Opinion about the Characters in the Movie
- Characteristics of a movie or genre
- **Opinion about the Movie and Crew**
- Unrelated Category

- *Born into a family of thespians -- parents Roger Winslet and Sally Bridges-Winslet*
- *He's an even-tempered, funny, robust old man who actually listens to the kids' problems (as opposed to their parents and friends, both caught up in the high-wire act of their own confused lives.).*

age.

Facets of a Movie Review

- General Perception about the Crew
- Objective Facts about the Crew and Movies
- Past Performance of the Crew and Movies
- Expectations from the Movie or Crew
- Movie Plot
- Opinion about the Characters in the Movie
- Characteristics of a movie or genre
- **Opinion about the Movie and Crew**
- Unrelated Category

- *Born into a family of thespians -- parents Roger Winslet and Sally Bridges-Winslet*

- *Horror movies are supposed to be scary.*
- *There is an axiom that directors who have a big hit with their debut have a big bomb with their second film.*

age.

Facets of a Movie Review

- General Perception about the Crew
- Objective Facts about the Crew and Movies
- Past Performance of the Crew and Movies
- Expectations from the Movie or Crew
- Movie Plot
- Opinion about the Characters in the Movie
- Characteristics of a movie or genre
- **Opinion about the Movie and Crew**
- Unrelated Category

- *Born into a family of thespians -- parents Roger Winslet and Sally Bridges-Winslet*

- *While the movie is brutal, the violence is neither very graphic nor gratuitous. It may scare the little ones, but the teen-age audience for which it is aimed will appreciate the man-eating chomping that runs through the film.*

age.

Facets of a Movie Review

- General Perception about the Crew
- Objective Facts about the Crew and Movies
- Past Performance of the Crew and Movies
- Expectations from the Movie or Crew
- Movie Plot
- Opinion about the Characters in the Movie
- Characteristics of a movie or genre
- **Opinion about the Movie and Crew**
- Unrelated Category

- *Born into a family of thespians -- parents Roger Winslet and Sally Bridges-Winslet*

- *So my grandson gives me passes to this new picture One Night at McCool's because the free screening is the same night as that horrible show with those poor prisoners trapped on the island who eat the bugs. "Go," he says, "it's just like Rush-o-Man."*

age.

Wikipedia

- Extensive World Knowledge required to perform this analysis
- Wikipedia is used to create a *topic-specific, extractive summary* of a review
- The extract is classified with a Lexicon, instead of the entire review

Feature Extraction from Wikipedia

Feature Extraction from Wikipedia

METADATA

Harry Potter and the Deathly Hallows – Part 1 is a 2010 fantasy film[5] directed by David Yates and the first of two films based on the novel Harry Potter and the Deathly Hallows by J. K. Rowling. It is the seventh instalment in the Harry Potter film series, written by Steve Kloves and produced by David Heyman, David Barron and Rowling. The story follows Harry Potter on a quest to find and destroy Lord Voldemort's secret to immortality – the Horcruxes. The film stars Daniel Radcliffe as Harry Potter, alongside Rupert Grint and Emma Watson as Harry's best friends Ron Weasley and Hermione Granger. It is the sequel to Harry Potter and the Half-Blood Prince and is followed by the concluding film, Harry Potter and the Deathly Hallows – Part 2.

Principal photography began on 19 February 2009 and was completed on 12 June 2010.[6] Part 1 was released in 2D cinemas and IMAX formats worldwide on 19 November 2010.

Feature Extraction from Wikipedia

PLOT

Further information: Harry Potter and the Deathly Hallows Novel Plot

Minister Rufus Scrimgeour addresses the wizarding media stating that the Ministry of Magic will remain strong as Lord Voldemort gains power throughout the wizarding and Muggle worlds. Severus Snape arrives at Malfoy Manor to inform Lord Voldemort and his Death Eaters of Harry's departure from No. 4 Privet Drive. Voldemort commandeers Lucius Malfoy's wand, as Voldemort's own wand cannot be used to kill Harry; their wands are "twins".

Meanwhile, the Order of the Phoenix arrive at Privet Drive and escort Harry to safety using Polyjuice Potion to create six decoy Harrys. During their flight to the Burrow, they are ambushed by Death Eaters, who kill Mad-Eye Moody and Hedwig, and injure George Weasley. They arrive at the Burrow, where Harry has a vision of Ollivander being tormented by Voldemort, who claims that the wand-maker had lied to him by informing him of the only way to kill Harry: obtaining another's wand.

Feature Extraction from Wikipedia

CREW

Directed by : David Yates
Produced by : David Heyman,David Barro, J. K. Rowling
Screenplay by : Steve Kloves Based on Harry Potter and the Deathly Hallows by J. K. Rowling
Starring : Daniel Radcliffe, Rupert Grint, Emma Watson
Music by : Alexandre Desplat
Themes: John Williams
Cinematography : Eduardo Serra
Editing by: Mark Day
Studio :Heyday Films
Distributed by: Warner Bros. Pictures

that the wand-maker had lied to him by informing him of the only way to kill Harry: obtaining another's wand.

Feature Extraction from Wikipedia

CREW

Directed by : David Yates
Produced by : David Heyman,David Barro, J. K. Rowling
Screenplay by : Steve Kloves Based on Harry Potter and the Deathly Hallows by J. K. Rowling
Starring : Daniel Radcliffe, Rupert Grint, Emma Watson
Music by : Alexandre Desplat
Themes: John Williams
Cinematography : Eduardo Serra
Editing by: Mark Day
Studio :Heyday Films
Distributed by: Warner Bros. Pictures

Narcissa Malfoy.
Ralph Fiennes as Lord Voldemort, the film's main antagonist

Feature Extraction from Wikipedia

DOMAIN SPECIFIC FEATURE LIST

Movie, Staffing, casting, Writing, Theory, Writing, Rewriting, Screenplay, Format, Treatments, Scriptments, Synopsis, Logline, Pitching, Certification,scripts, Budget, Ideas, Funding, budgeting, Funding, Plans, Grants, Pitching, Tax, Contracts, law, Copyright, Pre-production, Budgeting, Scheduling, Pre-production, film , stock, Story, boarding, plot, Casting , Directors, Location, Scouting, …..

Ralph Fiennes as Lord Voldemort, the film's main antagonist

Feature List Creation

Feature List Creation

- Metadata and Plot sentences are POS-tagged
  - Nouns retrieved
  - Stemmed
  - Added to the **Plot** list
  - Character information also added to **Plot** list

Feature List Creation

- Metadata and Plot sentences are POS-tagged
  - Nouns retrieved
  - Stemmed
  - Added to the **Plot** list
  - Character information also added to **Plot** list

- Domain specific feature list extracted from Wiki articles on films and movies
  - Added to **MovieFeature** list

Feature List Creation

- Metadata and Plot sentences are POS-tagged
  - Nouns retrieved
  - Stemmed
  - Added to the **Plot** list
  - Character information also added to **Plot** list
- Domain specific feature list extracted from Wiki articles on films and movies
  - Added to **MovieFeature** list
- Crew information added to **Crew** list

Feature List Creation

- Metadata and Plot sentences are POS-tagged
  - Nouns retrieved
  - Stemmed
  - Added to the **Plot** list
  - Character information also added to **Plot** list
- Domain specific feature list extracted from Wiki articles on films and movies
  - Added to **MovieFeature** list
- Crew information added to **Crew** list
- Laptop and Printer domain data used to filter frequent occurring concepts
  - Non-overlapping domains
  - Frequently occurring terms in all the domains added t**o FreqWords** list
    The **FreqWords** are pruned from the other feature lists

Why only Nouns?

- To restrict genre-specific concepts to be entities
- *Harry acted as if nothing has happened* vs. *Kate Winslet acted awesome in the movie.*
- Here *act* is present as a Verb in both the sentences
  - First sentence belongs to the *plot* (*Category 5*)
  - Second sentence depicts the reviewer opinion (*Category 8*)
- Difference lies in the presence of different subjects of interest with the Verb
- Our focus is to capture the *subjects* and *objects* in the sentence which give direct clues about the *category* of the reviewer statement, so that feature lists are as pure as possible

Extractive Summary

- Given a review R with n sentences S_i, determine if each sentence S_i is to be *accepted* or *rejected* based on a *relevancy factor* in judging this movie

- $$Rel_{factor_i} = 2\sum_j 1_{w_{ij} \in Crew\ or\ MovieTitle} + \sum_j 1_{w_{ij} \in MovieFeature} - \sum_j 1_{w_{ij} \in Plot, w_{ij} \notin Crew, w_{ij} \notin MovieTitle}$$

- $$Acc_{factor_i} = 1\ \boldsymbol{if}\ Rel_{factor_i} \geq 0\ and\ \exists w_{ij} \in S_i\ \boldsymbol{s.t.}\ w_{ij} \in Crew\ or\ MovieFeature\ or\ Movie\ Title$$
$$= 0\ otherwise$$

Semi-Supervised Learning of Parameters

- *Equations 2* can be re-written as:

- $Rel_{factor_i} = \alpha \times X_{i,1} + \beta \times X_{i,2} - \gamma \times X_{i,3}$

- $Acc_{factor_i} = Rel_{factor_i} - \theta$

$$= \alpha \times X_{i,1} + \beta \times X_{i,2} - \gamma \times X_{i,3} - \theta$$

$$= \alpha \times X_{i,1} + \beta \times X_{i,2} - \gamma \times X_{i,3} - \theta \times X_{i,4} \text{ (where } X_{i,4} = 1)$$

- Let Y_i be the binary label information corresponding to each sentence in the development set, where $Y_i=1$ if $Acc_{factor_i} \geq 0$ and -1 otherwise.

- $Y_i = \boldsymbol{W}.X_i$ where,
$\boldsymbol{W} = [\alpha\ \beta\ -\gamma\ -\theta]^T$ *and* $X_i = [X_{i,1}\ X_{i,2}\ X_{i,3}\ X_{i,4}]$
$or, \boldsymbol{Y} = \boldsymbol{W^T}.\boldsymbol{X}$

Algorithm

Algorithm

Input : Review R

Output: OpinionSummary

Step 1: Extract the Crew list from Wikipedia

Step 2: Extract the Plot list from Wikipedia

Step 3: Extract the MovieFeature list from Wikipedia

Step 4: Extract the FreqWords list as the common frequently occurring concepts in Mobile Phone, Printer and Movie domains.

Let $OpinionSummary = \emptyset$

for i=1..n

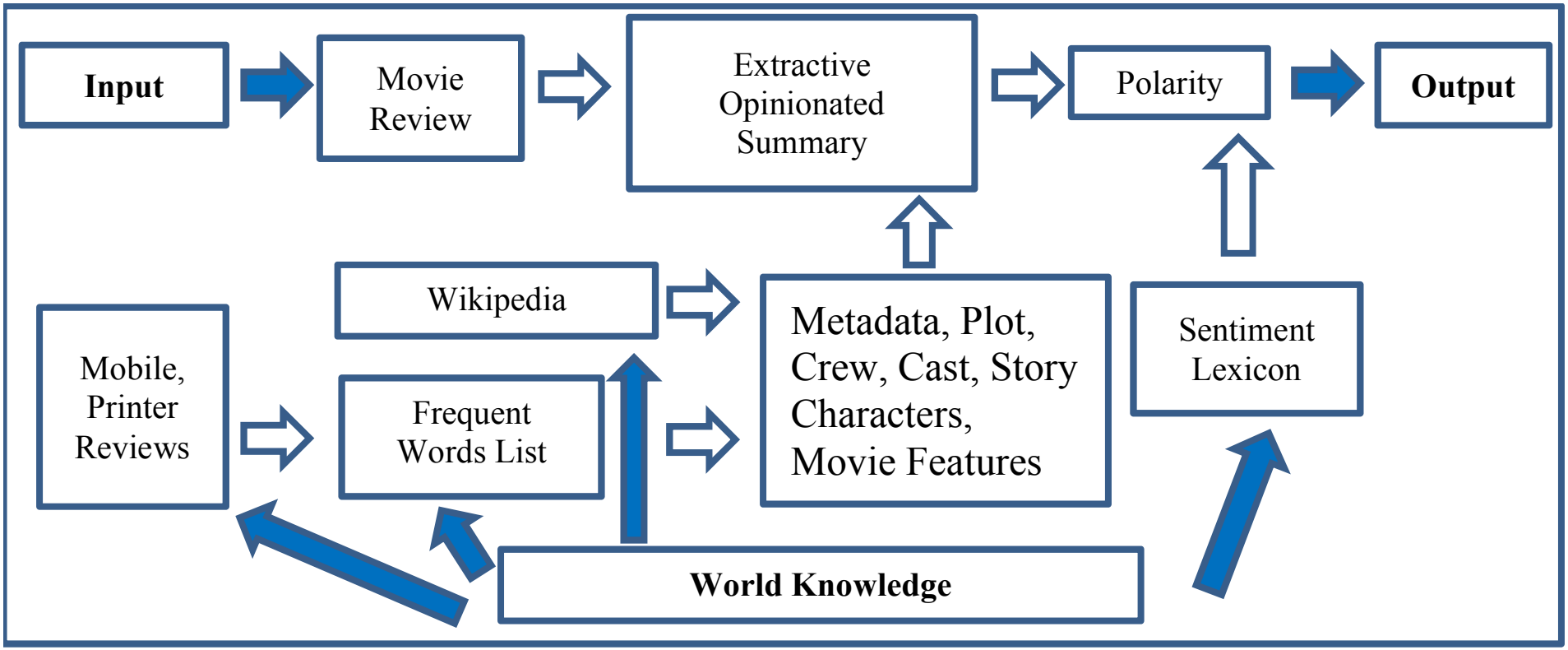
  if $Acc_{factor_i} == 1$

    add S_i *to* $OpinionSummary$

  end if

end for

WikiSent


Input
Movie Review
Extractive Opinionated Summary
Polarity
Output
Wikipedia
Metadata, Plot, Crew, Cast, Story Characters, Movie Features
Sentiment Lexicon
Mobile, Printer Reviews
Frequent Words List
World Knowledge

Algorithm Demonstration

- In Sentence [1], **Brian Cox** is the only keyword present and it belongs to the *Cast* list. Rel_{factor_i}=2*1+1*0-0*0=2 > -1 and the sentence is accepted.
- In [2], there is no keyword from the lists and it is rejected.
- [3] has the keyword **acting** from MovieFeature and is accepted.
- [4] has the keywords **Cox**, **L.I.E** from *Cast*, *MovieTitle*, **John Harrigan** from *Character* list and **distributor** from *MovieFeature* list. Rel_{factor_i}=2*2+1-1=4 > -1 and is accepted.
- [5] has only the keyword **Big John** from *Character* and Rel_{factor_i}=0+0-1=-1 and is rejected.
- [6] has the keyword **audience** from *MovieFeature* and **Big John** from *Character* and its Rel_{factor_i}=0+1-1=0 > -1 and is accepted.
- [7] has the keywords **temper**, **friend** from *Plot* and Rel_{factor_i}=0+0-2=-2 and is rejected.
- [8] has the keywords **sex**, **charm** from *Plot* and Rel_{factor_i}=0+0-2=-2 and is rejected.

Algorithm Demonstration

- In Sentence [1], **Brian Cox** is the only keyword present and it belongs to the *Cast* list. Rel_{factor_i}=2*1+1*0-0*0=2 > -1 and the sentence is accepted.
- In [2], there is no keyword from the lists and it is rejected.
- [3] has the keyword **acting** from MovieFeature and is accepted.
- [4] has the keywords **Cox**, **L.I.E** from *Cast*, *MovieTitle*, **John Harrigan** from *Character* list and **distributor** from *MovieFeature* list. Rel_{factor_i}=2*2+1-1=4 > -1 and is accepted.
- [5] has only the keyword **Big John** from *Character* and Rel_{factor_i}=0+0-1=-1 and is rejected.
- [6] has the keyword **audience** from *MovieFeature* and **Big John** from *Character* and its Rel_{factor_i}=0+1-1=0 > -1 and is accepted.
- [7] has the keywords **temper**, **friend** from *Plot* and Rel_{factor_i}=0+0-2=-2 and is rejected.
- [8] has the keywords **sex**, **charm** from *Plot* and Rel_{factor_i}=0+0-2=-2 and is rejected.

*Best remembered for his understated performance as Dr. Hannibal Lecter in Michael Mann's forensics thriller, Manhunter, Scottish character actor **Brian Cox** brings something special to every movie he works on.*

Algorithm Demonstration

- In Sentence [1], **Brian Cox** is the only keyword present and it belongs to the *Cast* list. Rel_{factor_i}=2*1+1*0-0*0=2 > -1 and the sentence is accepted.
- In [2], there is no keyword from the lists and it is rejected.
- [3] has the keyword **acting** from MovieFeature and is accepted.
- [4] has the keywords **Cox**, **L.I.E** from *Cast*, *MovieTitle*, **John Harrigan** from *Character* list and **distributor** from *MovieFeature* list. Rel_{factor_i}=2*2+1-1=4 > -1 and is accepted.
- [5] has only the keyword **Big John** from *Character* and Rel_{factor_i}=0+0-1=-1 and is rejected.
- [6] has the keyword **audience** from *MovieFeature* and **Big John** from *Character* and its Rel_{factor_i}=0+1-1=0 > -1 and is accepted.
- [7] has the keywords **temper**, **friend** from *Plot* and Rel_{factor_i}=0+0-2=-2 and is rejected.
- [8] has the keywords **sex**, **charm** from *Plot* and Rel_{factor_i}=0+0-2=-2 and is rejected.

Usually playing a bit role in some studio schlock (he dies halfway through The Long Kiss Goodnight), he's only occasionally given something meaty and substantial to do.

Algorithm Demonstration

- In Sentence [1], **Brian Cox** is the only keyword present and it belongs to the *Cast* list. Rel_{factor_i}=2*1+1*0-0*0=2 > -1 and the sentence is accepted.
- In [2], there is no keyword from the lists and it is rejected.
- [3] has the keyword **acting** from MovieFeature and is accepted.
- [4] has the keywords **Cox**, **L.I.E** from *Cast*, *MovieTitle*, **John Harrigan** from *Character* list and **distributor** from *MovieFeature* list. Rel_{factor_i}=2*2+1-1=4 > -1 and is accepted.
- [5] has only the keyword **Big John** from *Character* and Rel_{factor_i}=0+0-1=-1 and is rejected.
- [6] has the keyword **audience** from *MovieFeature* and **Big John** from *Character* and its Rel_{factor_i}=0+1-1=0 > -1 and is accepted.
- [7] has the keywords **temper**, **friend** from *Plot* and Rel_{factor_i}=0+0-2=-2 and is rejected.
- [8] has the keywords **sex**, **charm** from *Plot* and Rel_{factor_i}=0+0-2=-2 and is rejected.

*If you want to see some brilliant **acting**, check out his work as a dogged police inspector opposite Frances McDormand in Ken Loach's Hidden Agenda.*

Algorithm Demonstration

- In Sentence [1], **Brian Cox** is the only keyword present and it belongs to the *Cast* list. Rel_{factor_i}=2*1+1*0-0*0=2 > -1 and the sentence is accepted.
- In [2], there is no keyword from the lists and it is rejected.
- [3] has the keyword **acting** from MovieFeature and is accepted.
- [4] has the keywords **Cox**, **L.I.E** from *Cast*, *MovieTitle*, **John Harrigan** from *Character* list and **distributor** from *MovieFeature* list. Rel_{factor_i}=2*2+1-1=4 > -1 and is accepted.
- [5] has only the keyword **Big John** from *Character* and Rel_{factor_i}=0+0-1=-1 and is rejected.
- [6] has the keyword **audience** from *MovieFeature* and **Big John** from *Character* and its Rel_{factor_i}=0+1-1=0 > -1 and is accepted.
- [7] has the keywords **temper**, **friend** from *Plot* and Rel_{factor_i}=0+0-2=-2 and is rejected.
- [8] has the keywords **sex**, **charm** from *Plot* and Rel_{factor_i}=0+0-2=-2 and is rejected.

***Cox** plays the role of Big **John Harrigan** in the disturbing new indie flick **L.I.E.**, which Lot 47 picked up at Sundance when other **distributors** were scared to budge.*

Algorithm Demonstration

- In Sentence [1], **Brian Cox** is the only keyword present and it belongs to the *Cast* list. Rel_{factor_i}=2*1+1*0-0*0=2 > -1 and the sentence is accepted.
- In [2], there is no keyword from the lists and it is rejected.
- [3] has the keyword **acting** from MovieFeature and is accepted.
- [4] has the keywords **Cox**, **L.I.E** from *Cast*, *MovieTitle*, **John Harrigan** from *Character* list and **distributor** from *MovieFeature* list. Rel_{factor_i}=2*2+1-1=4 > -1 and is accepted.
- [5] has only the keyword **Big John** from *Character* and Rel_{factor_i}=0+0-1=-1 and is rejected.
- [6] has the keyword **audience** from *MovieFeature* and **Big John** from *Character* and its Rel_{factor_i}=0+1-1=0 > -1 and is accepted.
- [7] has the keywords **temper**, **friend** from *Plot* and Rel_{factor_i}=0+0-2=-2 and is rejected.
- [8] has the keywords **sex**, **charm** from *Plot* and Rel_{factor_i}=0+0-2=-2 and is rejected.

Big John *feels the love that dares not speak its name, but he expresses it through seeking out adolescents and bringing them back to his pad.*

Algorithm Demonstration

- In Sentence [1], **Brian Cox** is the only keyword present and it belongs to the *Cast* list. Rel_{factor_i}=2*1+1*0-0*0=2 > -1 and the sentence is accepted.
- In [2], there is no keyword from the lists and it is rejected.
- [3] has the keyword **acting** from MovieFeature and is accepted.
- [4] has the keywords **Cox**, **L.I.E** from *Cast*, *MovieTitle*, **John Harrigan** from *Character* list and **distributor** from *MovieFeature* list. Rel_{factor_i}=2*2+1-1=4 > -1 and is accepted.
- [5] has only the keyword **Big John** from *Character* and Rel_{factor_i}=0+0-1=-1 and is rejected.
- [6] has the keyword **audience** from *MovieFeature* and **Big John** from *Character* and its Rel_{factor_i}=0+1-1=0 > -1 and is accepted.
- [7] has the keywords **temper**, **friend** from *Plot* and Rel_{factor_i}=0+0-2=-2 and is rejected.
- [8] has the keywords **sex**, **charm** from *Plot* and Rel_{factor_i}=0+0-2=-2 and is rejected.

*What bothered some **audience** members was the presentation of **Big John** in an oddly empathetic light.*

Algorithm Demonstration

- In Sentence [1], **Brian Cox** is the only keyword present and it belongs to the *Cast* list. Rel_{factor_i}=2*1+1*0-0*0=2 > -1 and the sentence is accepted.
- In [2], there is no keyword from the lists and it is rejected.
- [3] has the keyword **acting** from MovieFeature and is accepted.
- [4] has the keywords **Cox**, **L.I.E** from *Cast*, *MovieTitle*, **John Harrigan** from *Character* list and **distributor** from *MovieFeature* list. Rel_{factor_i}=2*2+1-1=4 > -1 and is accepted.
- [5] has only the keyword **Big John** from *Character* and Rel_{factor_i}=0+0-1=-1 and is rejected.
- [6] has the keyword **audience** from *MovieFeature* and **Big John** from *Character* and its Rel_{factor_i}=0+1-1=0 > -1 and is accepted.
- [7] has the keywords **temper**, **friend** from *Plot* and Rel_{factor_i}=0+0-2=-2 and is rejected.
- [8] has the keywords **sex**, **charm** from *Plot* and Rel_{factor_i}=0+0-2=-2 and is rejected.

*He's an even-**tempered**, funny, robust old man who actually listens to the kids' problems (as opposed to their parents and **friends**, both caught up in the high-wire act of their own confused lives.).*

Algorithm Demonstration

- In Sentence [1], **Brian Cox** is the only keyword present and it belongs to the *Cast* list. Rel_{factor_i}=2*1+1*0-0*0=2 > -1 and the sentence is accepted.
- In [2], there is no keyword from the lists and it is rejected.
- [3] has the keyword **acting** from MovieFeature and is accepted.
- [4] has the keywords **Cox**, **L.I.E** from *Cast*, *MovieTitle*, **John Harrigan** from *Character* list and **distributor** from *MovieFeature* list. Rel_{factor_i}=2*2+1-1=4 > -1 and is accepted.
- [5] has only the keyword **Big John** from *Character* and Rel_{factor_i}=0+0-1=-1 and is rejected.
- [6] has the keyword **audience** from *MovieFeature* and **Big John** from *Character* and its Rel_{factor_i}=0+1-1=0 > -1 and is accepted.
- [7] has the keywords **temper**, **friend** from *Plot* and Rel_{factor_i}=0+0-2=-2 and is rejected.
- [8] has the keywords **sex**, **charm** from *Plot* and Rel_{factor_i}=0+0-2=-2 and is rejected.

*He'll have **sex**-for-pay with them only after an elaborate courtship, **charming** them with temptations from the grown-up world"*

Lexicons and Datasets

Lexicons and Datasets

- Lexicons
  - SentiWordNet (Esuli *et al.*, 2006)
  - Subjectivity Lexicon (Wilson *et al.*, 2005)
  - General Inquirer (Stone *et al.*, 1966)

Lexicons and Datasets

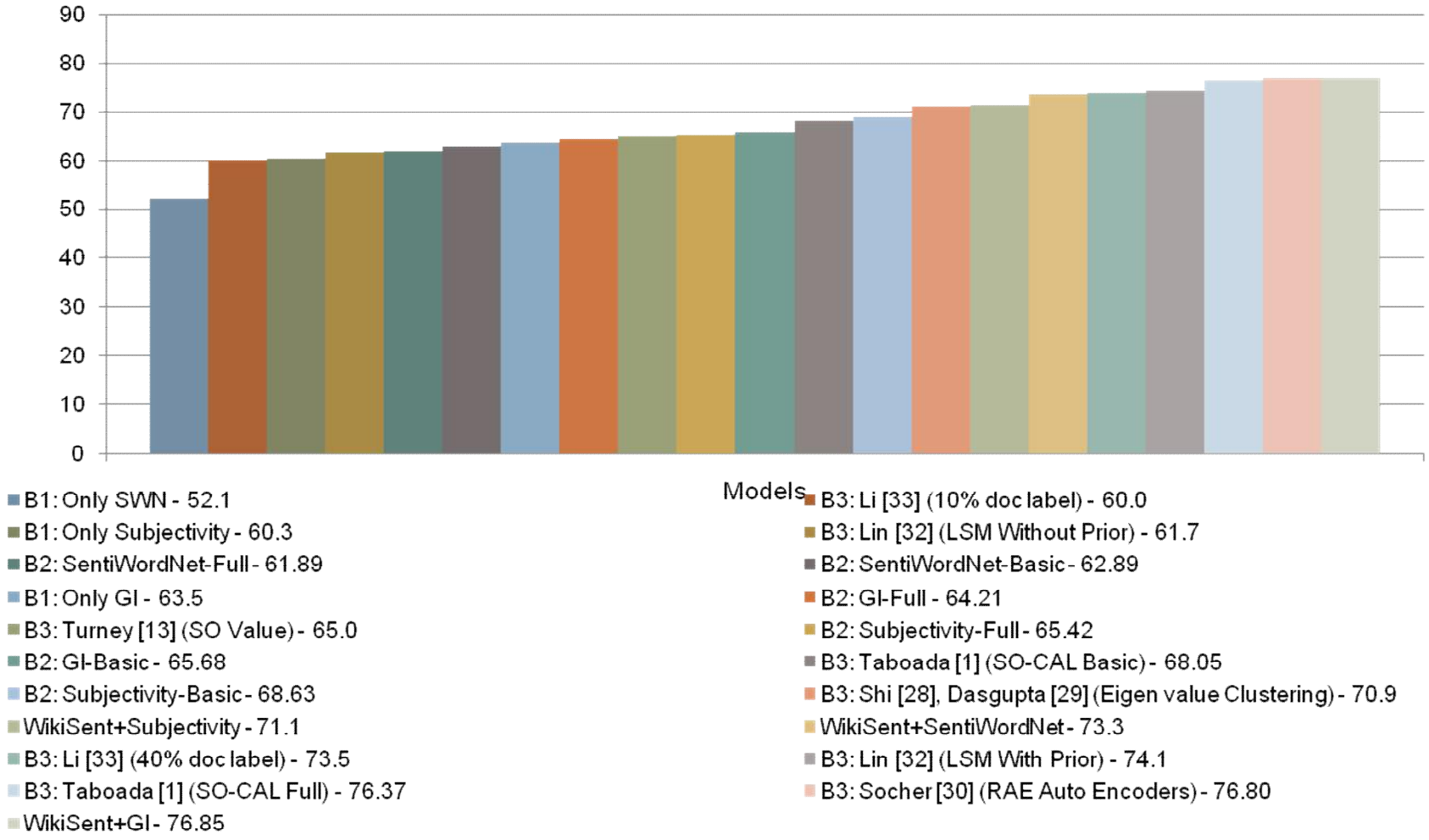
- Lexicons
  - SentiWordNet (Esuli *et al.*, 2006)
  - Subjectivity Lexicon (Wilson *et al.*, 2005)
  - General Inquirer (Stone *et al.*, 1966)

- Baseline 1
  - Bag-of-Words based on the 3 lexicons
- Baseline 2
  - SO-CAL (Taboada *et al.*, 2011)
- Baseline 3
  - All the semi-supervised and unsupervised systems in the domain

Lexicons and Datasets

- Lexicons
  - SentiWordNet (Esuli *et al.*, 2006)
  - Subjectivity Lexicon (Wilson *et al.*, 2005)
  - General Inquirer (Stone *et al.*, 1966)
- Baseline 1
  - Bag-of-Words based on the 3 lexicons
- Baseline 2
  - SO-CAL (Taboada *et al.*, 2011)
- Baseline 3
  - All the semi-supervised and unsupervised systems in the domain
- Movie Review Dataset (Pang *et al.*, 2002)
  - 1000 positive and 1000 negative reviews (labeled at document level)
    27,000 untagged reviews

Accuracy Comparison with All the Semi-supervised and Unsupervised Systems in the Same Dataset



Accuracy Comparison with Different Baselines

Accuracy Comparison with Different Baselines

Baseline 1 : Simple Bag-of-Words	
Only SentiWordNet	52.1
Only Subjectivity	60.3
Only GI	63.5
Baseline 2 : Worse, Median and Best Performing Lexicons with SO-CAL	
SentiWordNet-Full	61.89
SentiWordNet-Basic	62.89
GI-Full	64.21
GI-Basic	65.68
Subjectivity-Full	65.42
Subjectivity-Basic	68.63
WikiSent with Different Lexicons	
WikiSent+Subjectivity	71.1
WikiSent+SentiWordNet	73.3
WikiSent+GI	**76.85**

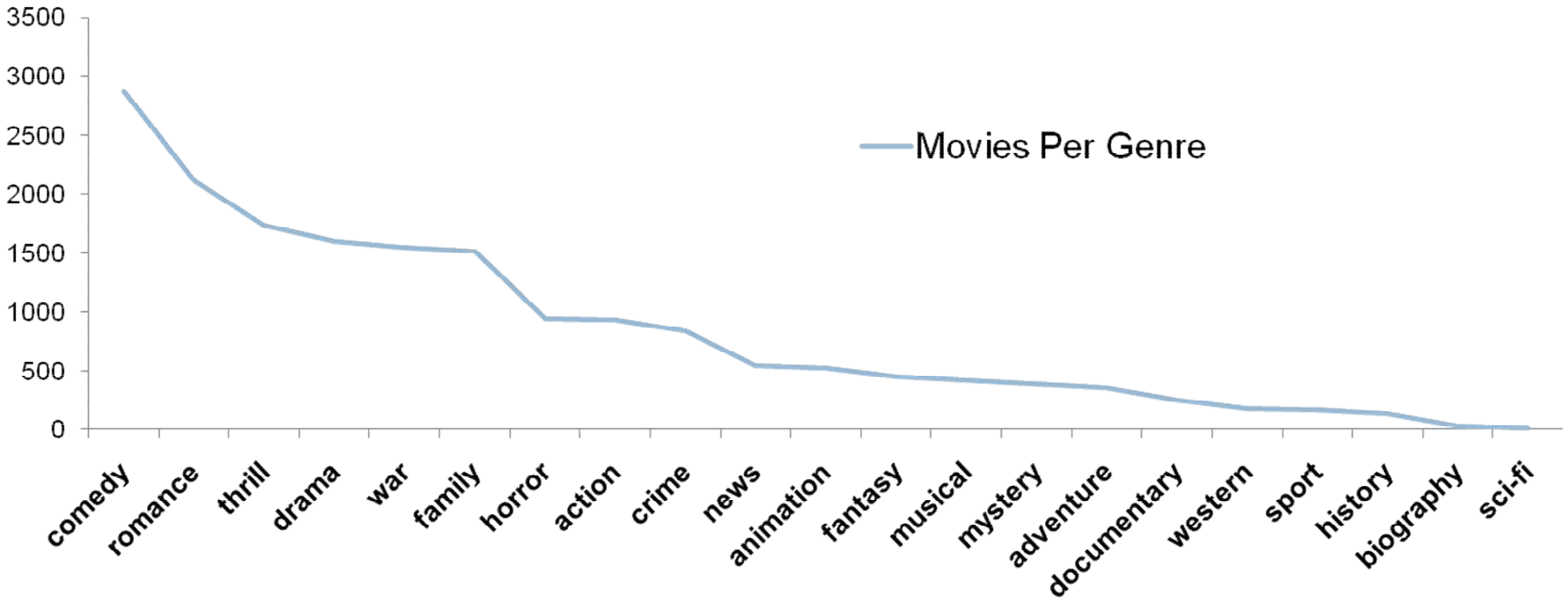
Accuracy Comparison with Different Baselines

Systems	Classification Method	Accuracy
Turney SO value	Unsupervised (PMI)	65
Taboada SO-CAL Basic [1]	Lexicon Generation	68.05
Taboada SO-CAL Full [1]	Lexicon Generation	76.37
Shi [28], Dasgupta [29]	Unsupervised Eigen Vector Clustering	70.9
Socher [30] RAE	Semi Supervised Auto Encoders	76.8
Lin [32] LSM	Unsupervised without prior info	61.7
Lin [32] LSM	Weakly Supervised with prior info	74.1
Li [33]	Semi Supervised 10% doc. Label	60
Li [33]	Semi Supervised 40% doc. Label	60
WikiSent	**Wikipedia+GI Lexicon**	**76.85**

Trend Analysis

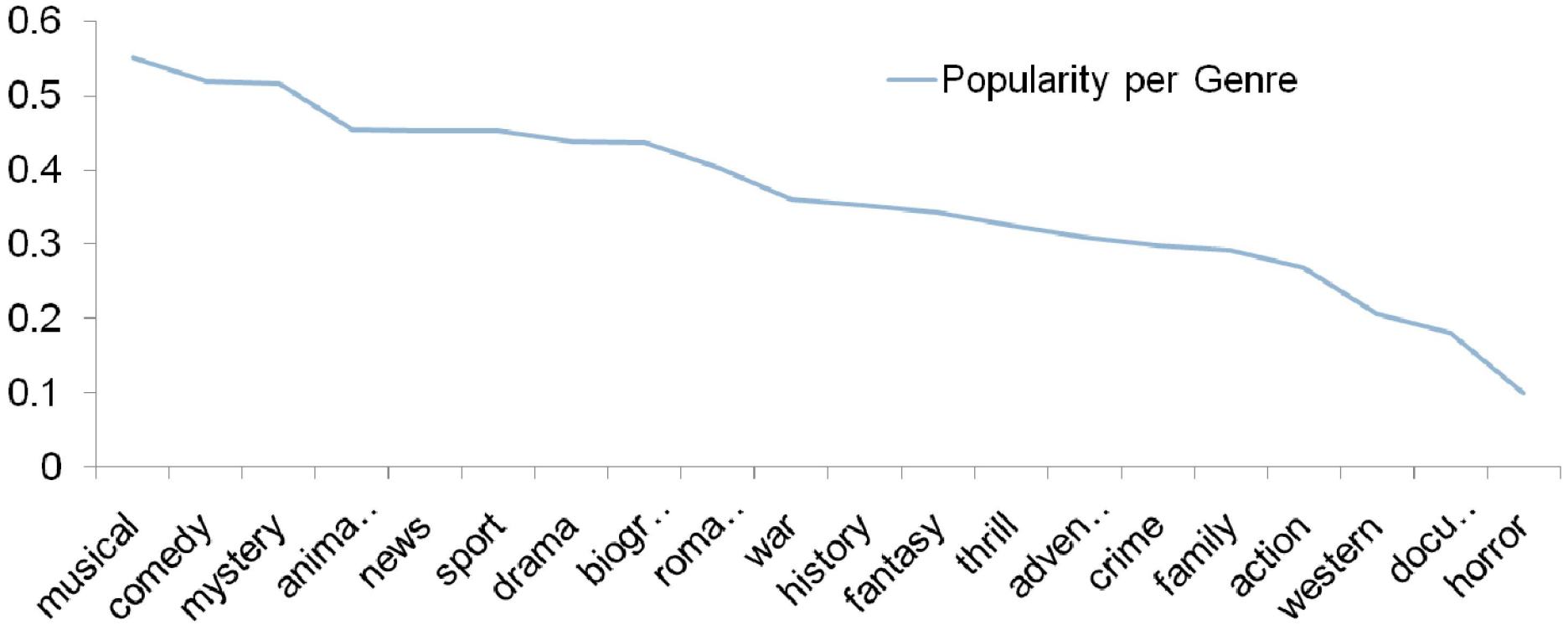
$$Genre\ Popularity = \frac{Positive\ Movie\ Reviews\ per\ Genre}{Total\ Movie\ Reviews\ per\ Genre}$$

Trend Analysis



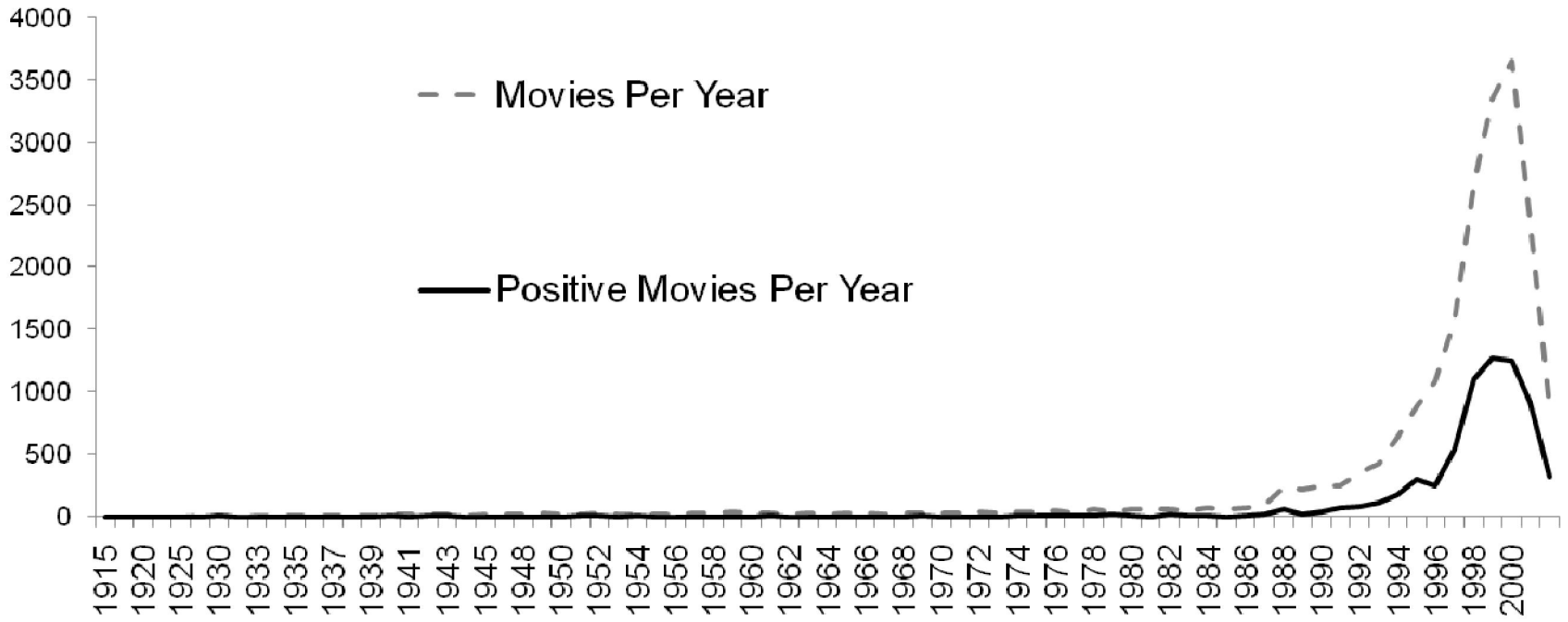
$$Genre\ Popularity = \frac{Positive\ Movie\ Reviews\ per\ Genre}{Total\ Movie\ Reviews\ per\ Genre}$$

Trend Analysis



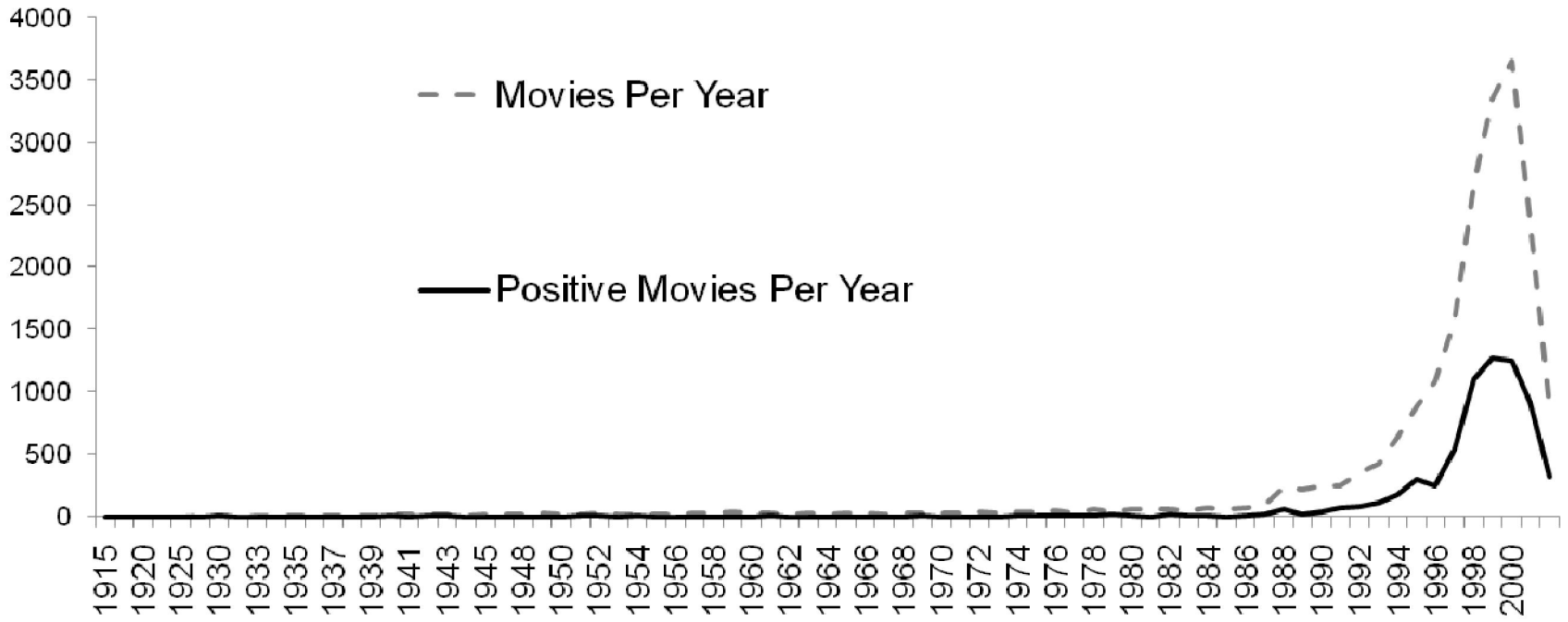
$$Genre\ Popularity = \frac{Positive\ Movie\ Reviews\ per\ Genre}{Total\ Movie\ Reviews\ per\ Genre}$$

Trend Analysis



$$Genre\ Popularity = \frac{Positive\ Movie\ Reviews\ per\ Genre}{Total\ Movie\ Reviews\ per\ Genre}$$

Trend Analysis



	WikiSent	Bag-of-Words Baseline 1
Positive Reviews (%)	48.95	81.2
Negative Reviews (%)	51.05	18.79

Drawbacks

- Absence of a co-reference resolution module
  - *false negative*
- Synonymous concepts not handled
  - Does not matter much as genre-specific concepts like *acting*, *direction*, *story-writer* occur in same lexical form
- Reviewer opinion bias can affect the system
  - *false positive*
- Absence of WSD module affects lexicon-based classification