

Mining Domain-Specific Dictionaries

**Pantelis
Agathangelou**

*Open University
of Cyprus*



Ioannis Katakis

*University of
Athens*



Fotios Kokkoras

*Technological
Educational Institute
of Thessaly*



**Konstantinos
Ntonas**

*International
Hellenic University*



Summary

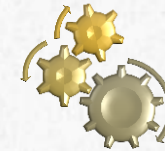
The Opinion Mining Problem



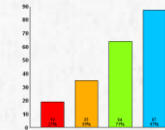
Introduction



Proposed Method



Experimental Evaluation

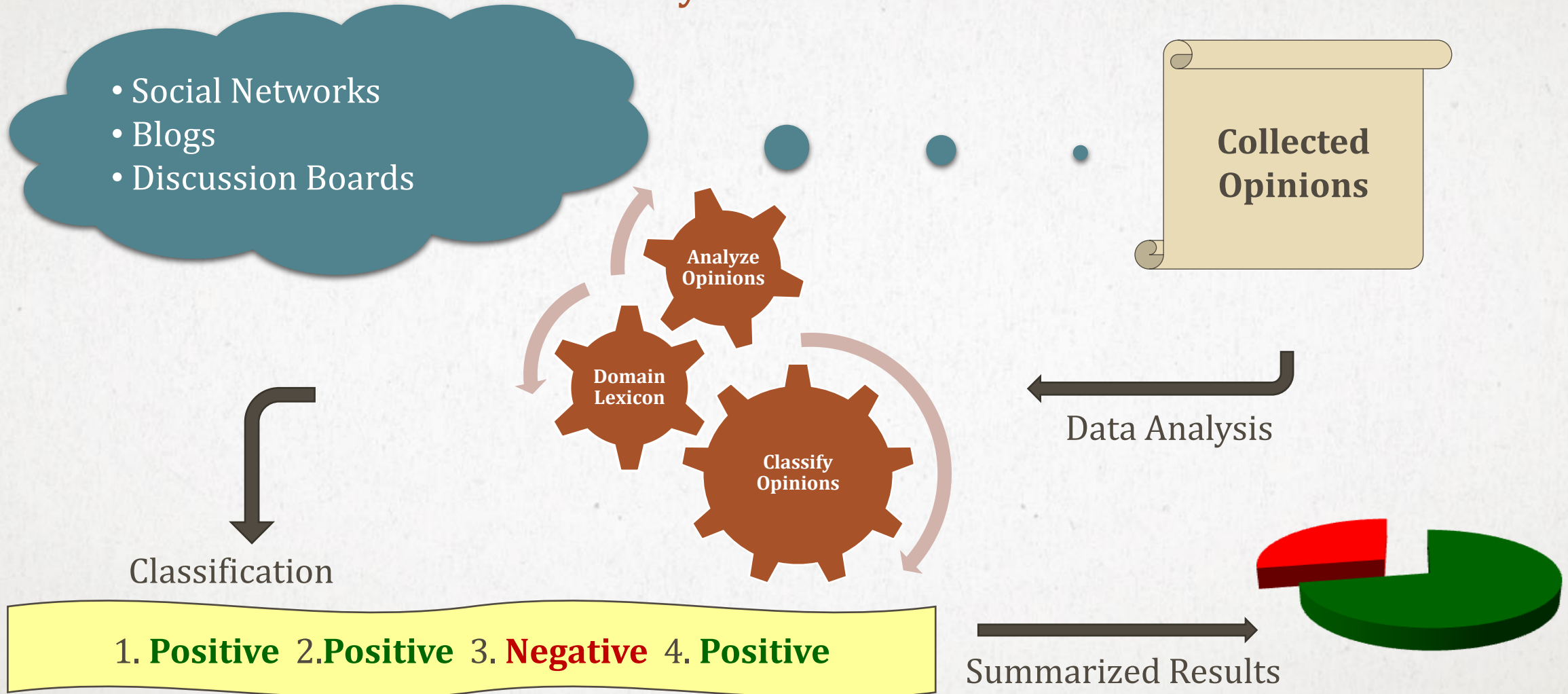


Interface



The Opinion Mining Problem

In Dictionary Based Solutions



DOMAIN SPECIFIC OPINION LEXICON

LEXICON ATTRIBUTES

- List of terms of known polarity (Positive or Negative)
- Strength or Sentiment Tension

SAMPLE

Positive	Sentiment Tension
Beautiful	10
Astonishing	4
Cool	-4
Negative	Tension
Slow	-6
Ugly	-3
Low	+3

Differences
In Comparison
To Generic
Lexicons

INTRODUCTION

❑ What we do in this paper ?

- ✓ We mine a domain specific dictionary
- ✓ We implement a multiple stage approach
- ✓ We utilize language patterns for the extraction process

❑ What is the innovation ?

- The designed algorithm can operate with a small initial seed list
- The method is unsupervised
- It can operate in multiple languages, provided the appropriate patterns
- Produces fast and accurate results.

Proposed Method

Opinion Preprocessing



Auxiliary List Preparation (Modules)



Seed Import and Filtered Seed Extraction



Conjunction Based Extraction



Double Propagation & Opinion Word Validation



OPINION PREPROCESSING



- Receives user opinions in raw form.

A nice little tablet!!,
I had mine @ for almost half a year.
The battery # lasts a few days for me.

- Implement some form of preprocessing

A nice little tablet!!,
I had mine @ for almost half a year.
The battery # lasts a few days for me.

- Sentence splitting – delimitation

- Additionally – Stemmer Engine

A nice little tablet,
I had mine for almost half a year.
The battery lasts a few days for me.

Sentence Splitters

AUXILIARY LIST PREPARATION - MODULES



Articles	Basic Verbs	Comparatives	Decreasers
the	be	cheaper	little
a	bend	finer	clearer
an	chose	newer	slower
one	throw	stronger	poorer

Future Words	Increasesers	Negations	Pronouns
will	better	none	my
to	hotter	no	they
let	harder	any	everybody
if	darker	anyone	her

Sum: 380 word constants

SEED IMPORT AND FILTERED SEED EXTRACTION



FILTER SEED EXTRACTION PATTERNS

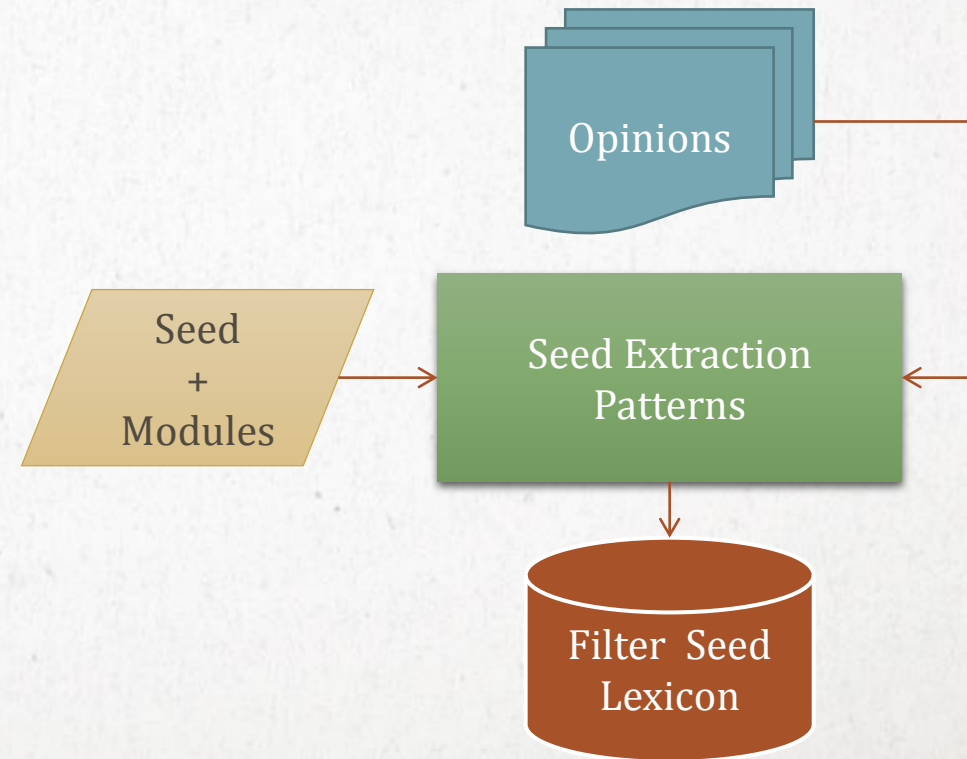
Positive Seed Patterns

	{decr}	{comp}	{pos}	-1	
e.g.	little	more	useful		
	{neg}	{pos}		-1	
	{pos}			+1	
	{art}	{neg}	{verb}	{decr}	+1
e.g.	the	noise	has	depleted	
	{neg}	{verb}	{neg}		+1
	{pos}	{neg}			+1
	{neg}				-1

Negative Seed Patterns

21 positive, 12 negative polarity patterns in total

EXTRACTION PROCESS



CONJUNCTION-BASED EXTRACTION



CONJUNCTION BASED EXTRACTION PATTERNS

Positive
Conj Based
Patterns

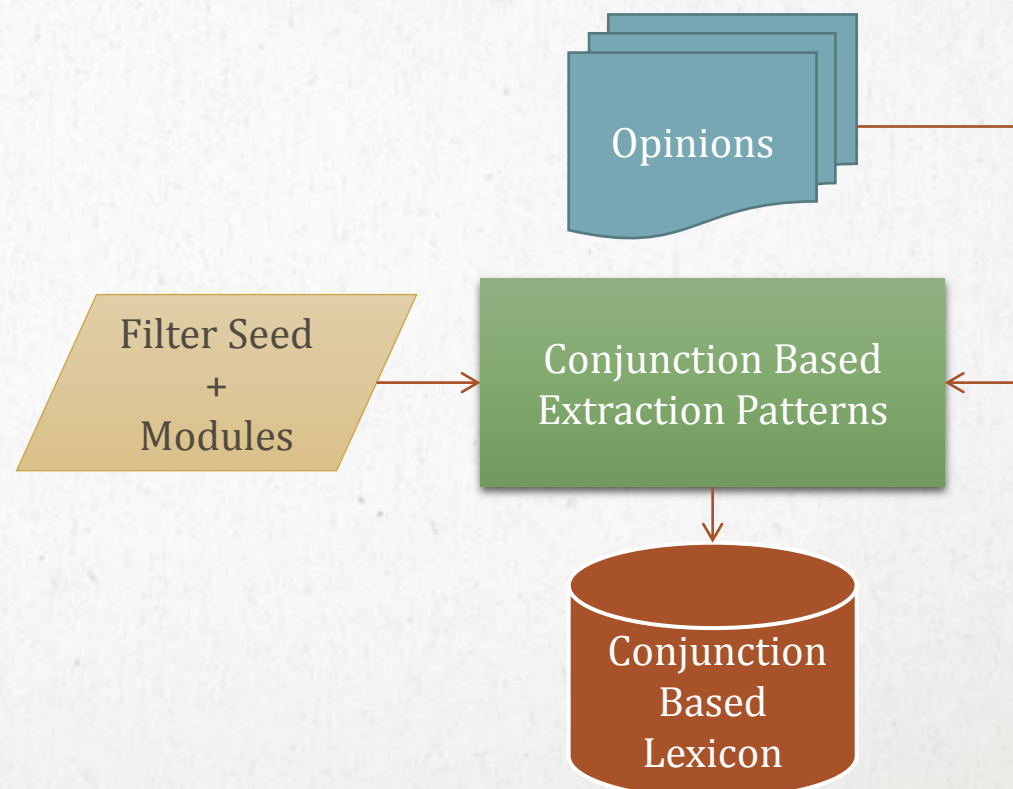
	{pos}	{conj}	{comp}	{art}	{cpos}
	{pos}	{conj}	{neg}	{cpos}	
e.g.	thin	and	not	sticky	
	{pos}	{conj}	{cpos}		

	{neg}	{conj}	{fut}	{verb}	{cneg}
	{neg}	{conj}	{incr}	{cneg}	
e.g.	expensive	and	too	small	
	{neg}	{conj}	{cneg}		

Negative
Conj Based
Patterns

6 positive, 4 negative extraction patterns in total

EXTRACTION PROCESS



DOUBLE PROPAGATION EXTRACTION METHOD



DOUBLE PROPAGATION PATTERNS

{art}	{copw}	{pron}	{verb}	{copw}
{copw}	{conj}	{art}	{copw}	
e.g. amazing	and	the	cost	
{copw}	{copw}			

6 extraction patterns in total

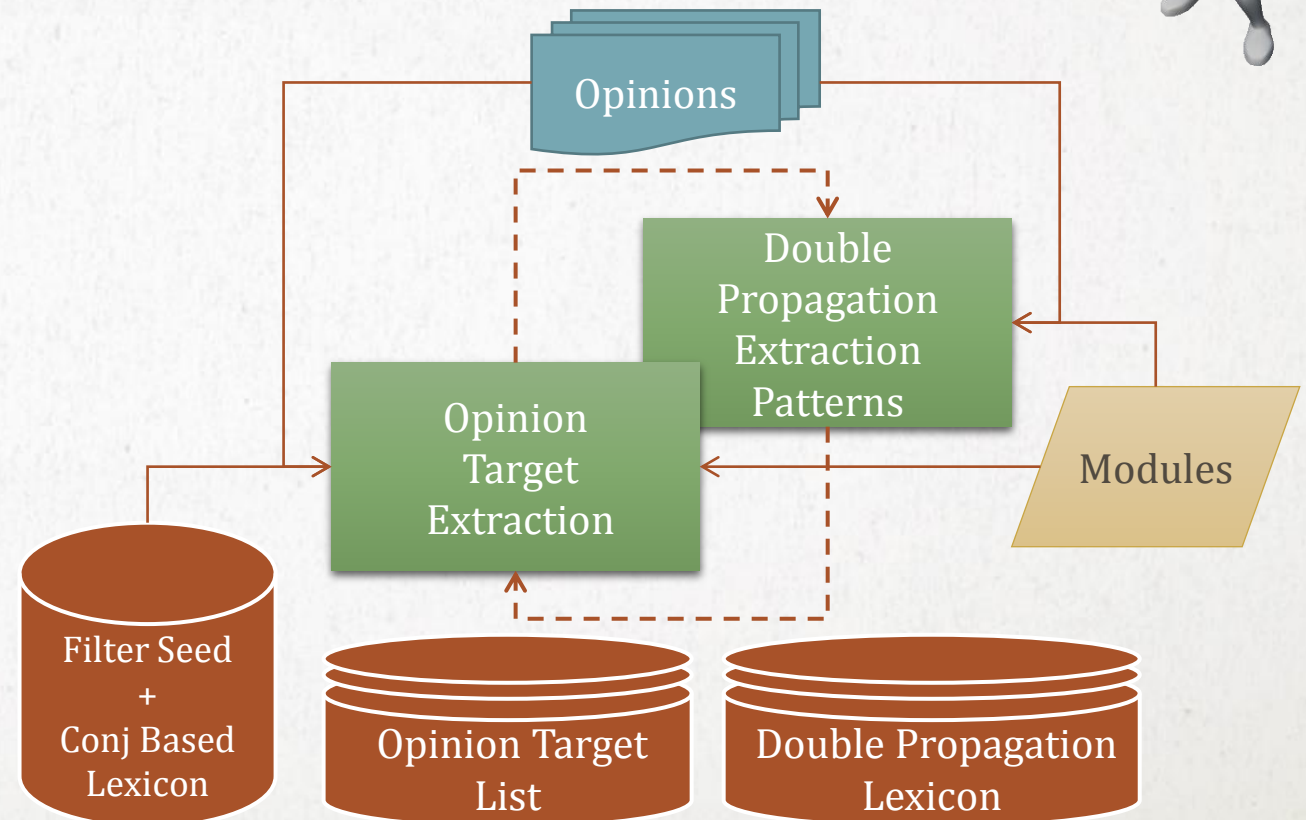
OPINION TARGET EXTRACTION

e.g. nice phone, amazing screen

Opinion Targets

e.g. wide and tall

EXTRACTION PROCESS



DOUBLE PROPAGATION SENTIMENT EXTRACTION



STEP 1

- Intra –Sentential Sentiment Consistency

INTRA SENTENTIAL EXAMPLE

(magnificent cool screen)[+1]

Opinion word **cool** is extracted from opinion target **screen**, inherits sentence polarity [+1]

STEP 2

- Inter –Sentential Sentiment Consistency (max depth = 3)

INTER SENTENTIAL EXAMPLE

(very good that mobile)[+1]

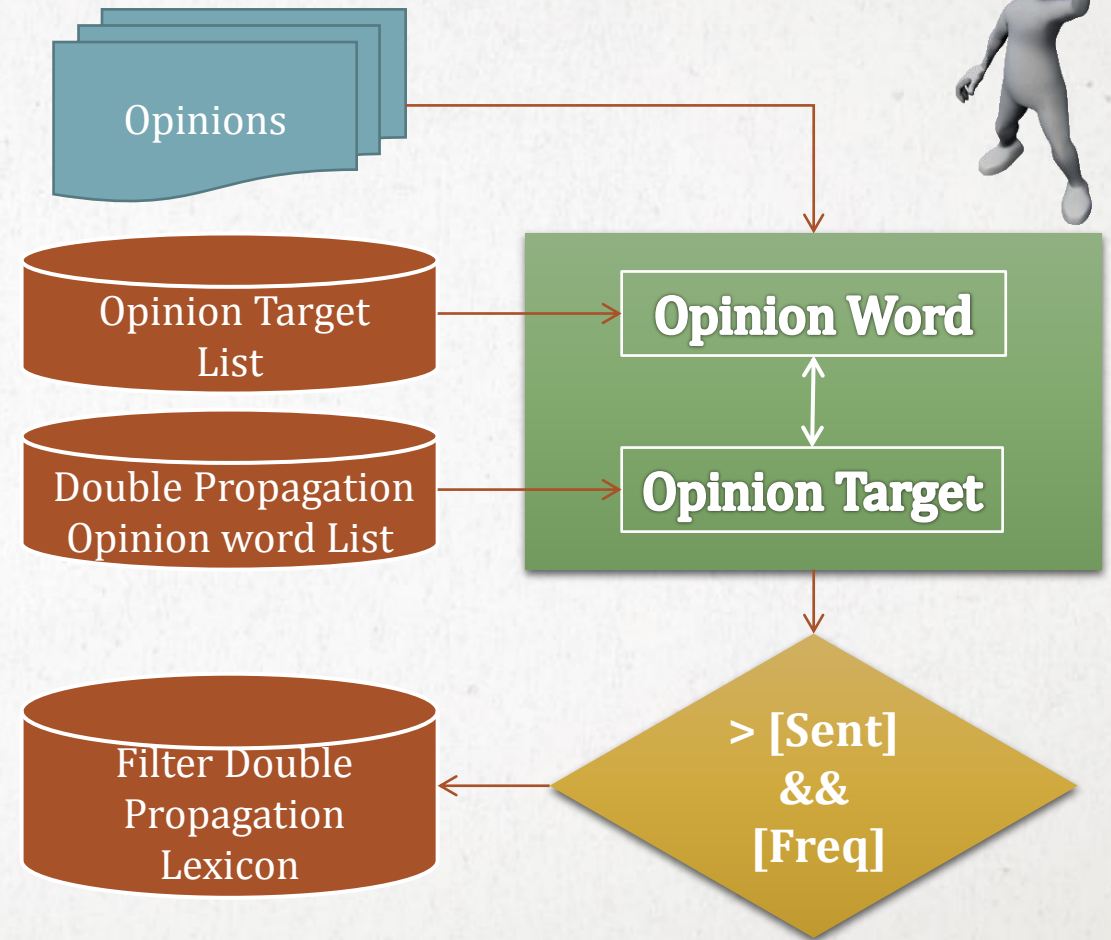
(awesome screen)[0]

(easy browsing fabulous graphics)[+2]

awesome extracted from **screen**, inherits sentence polarity [+2] at depth 1

OPINION WORD VALIDATION

- We use the extracted double propagation opinion word set and opinion target word set
- Sentiment Threshold **[Sent]**: Minimum accepted polarity
- Frequency Threshold **[Freq]**: Minimum accepted frequency co-existence of opinion word – opinion target



EXPERIMENTAL RESULTS

ALGORITHM FEATURES

- When Conjunction Based Extraction Fails to discover seed words, double propagation fills in the extraction gap.
- When Conjunction Based Extraction value is balanced so is double propagation.

AVERAGE PRECISION – RECALL METRICS

Source Category	Opinions	Average Values	
		Conj. Precision	Double Prop. Recall
Televisions	322	0%	49%
Air Conditioners	139	28%	15%
Washing Machines	83	3%	5%
Cameras	166	0%	46%
Refrigerators	103	35%	10%
Mobiles	3626	19%	38%
Tablets	448	54%	35%

Conclusion: The above results justify the unsupervised manner of the proposed method.

EXPERIMENTAL RESULTS

COMPARISON BETWEEN EXTRACTION STEPS

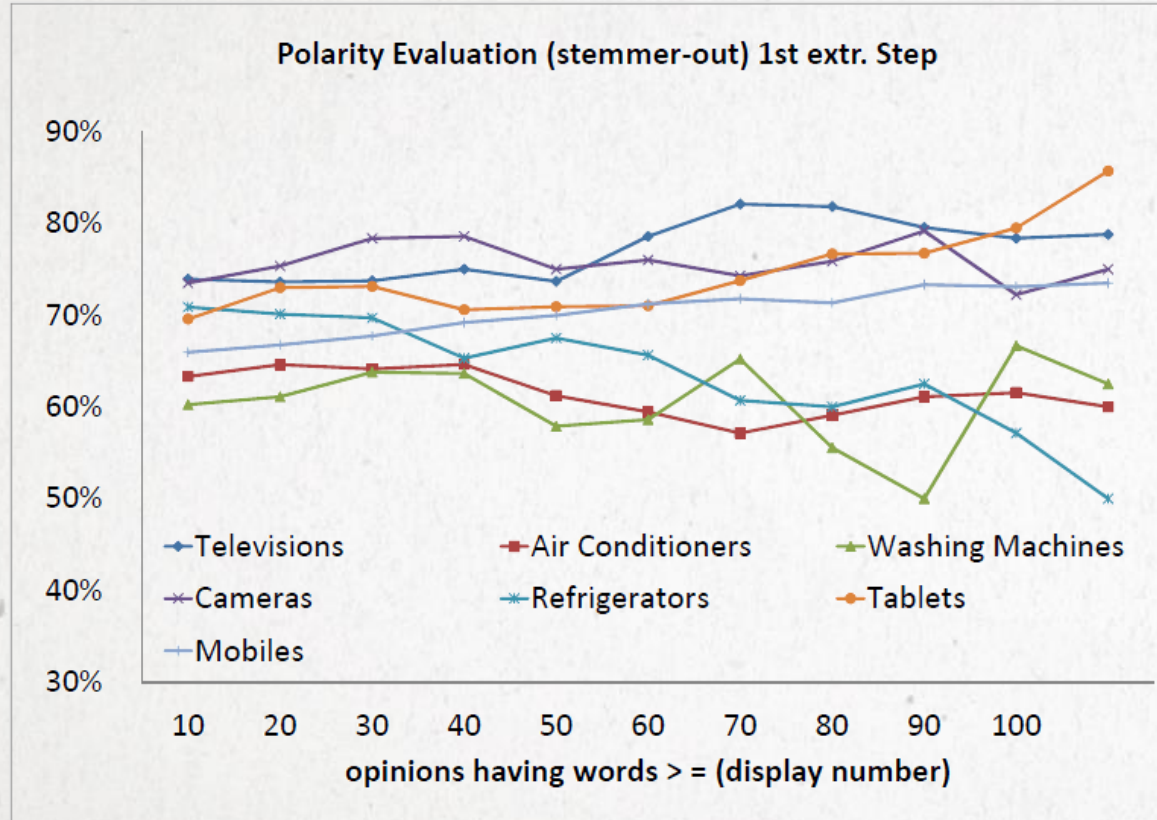
- Filter Seed builds the base of classification, but double propagation extents it.
- Conjunction based has low impact at overall classification

QUALITY OF THE EXTRACTED LEXICON BY EVALUATION CLASSIFICATION

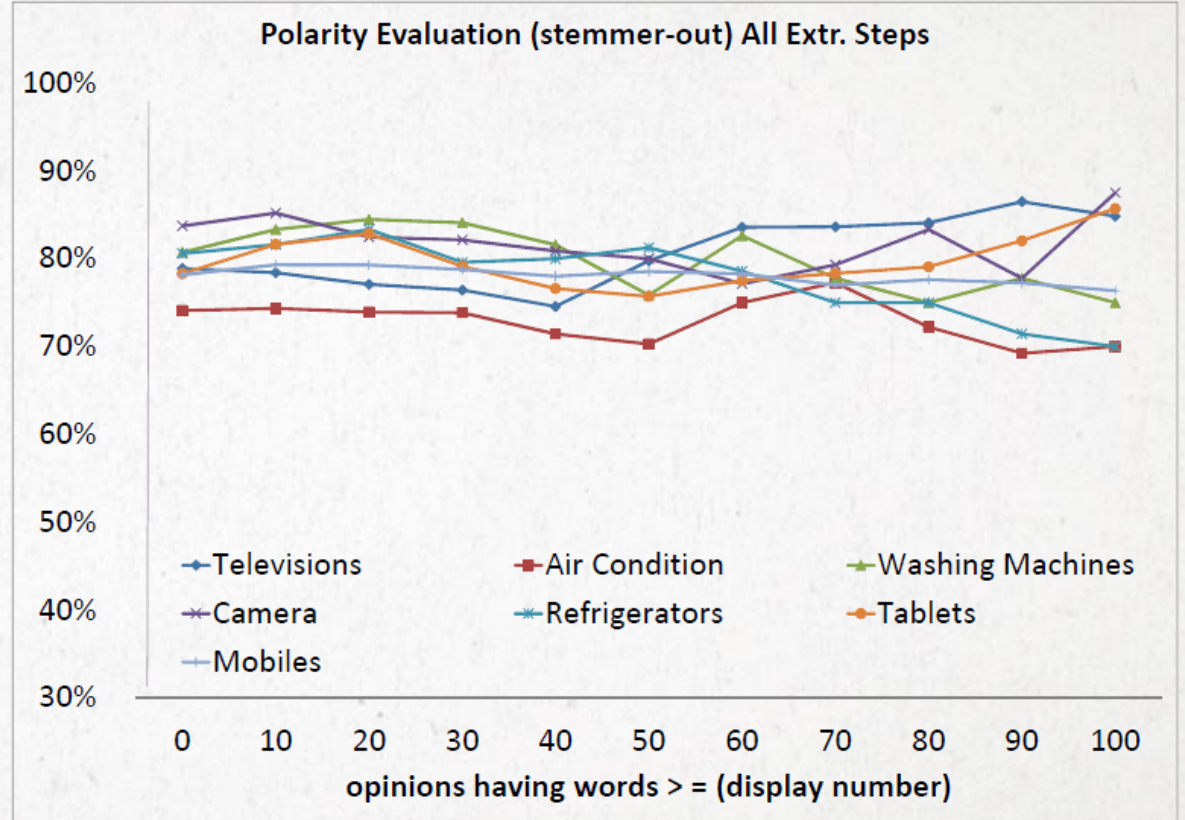
			Average polarity evaluation (stemmer out)		
Source Category	Opinions	Sentences Processed	Filtered Seed	Conj. extr.	Double prop.
Televisions	322	1630	77,19%	77,19%	80,71%
Air Conditioners	139	847	61,47%	65,14%	72,87%
Washing Machines	83	515	60,48%	60,48%	79,84%
Cameras	166	872	75,75%	75,75%	81,77%
Refrigerators	103	539	63,59%	63,59%	77,85%
Mobiles	3626	20284	70,35%	70,39%	78,05%
Tablets	448	2142	74,60%	74,60%	79,71%
Total Average			69,06%	69,59%	78,69%
Step Contribution:				1%	12%

EXPERIMENTAL RESULTS

Evaluation Based on Sentiment Classification



Stemmer out, first step

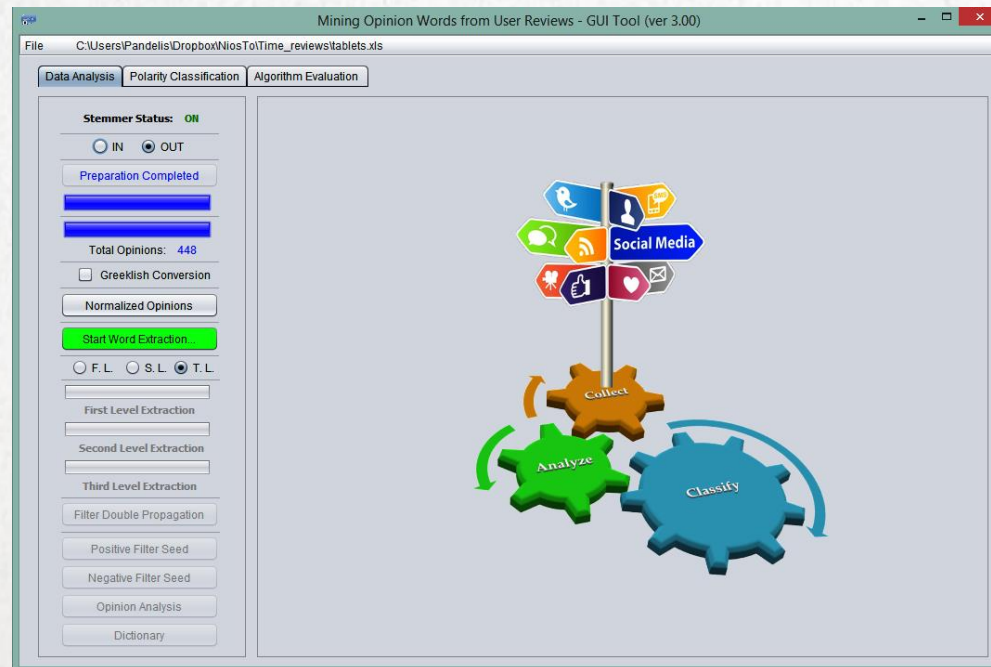


Stemmer out, all steps

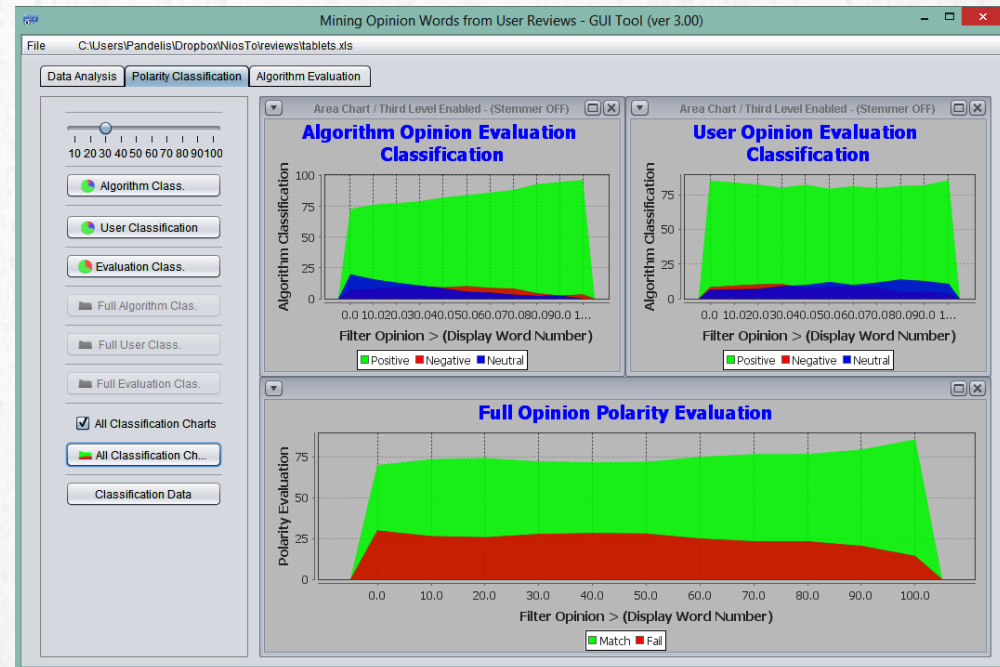
Conclusion: double propagation normalizes the quality of the lexicon upwards

INTERFACE

WELCOME SCREEN



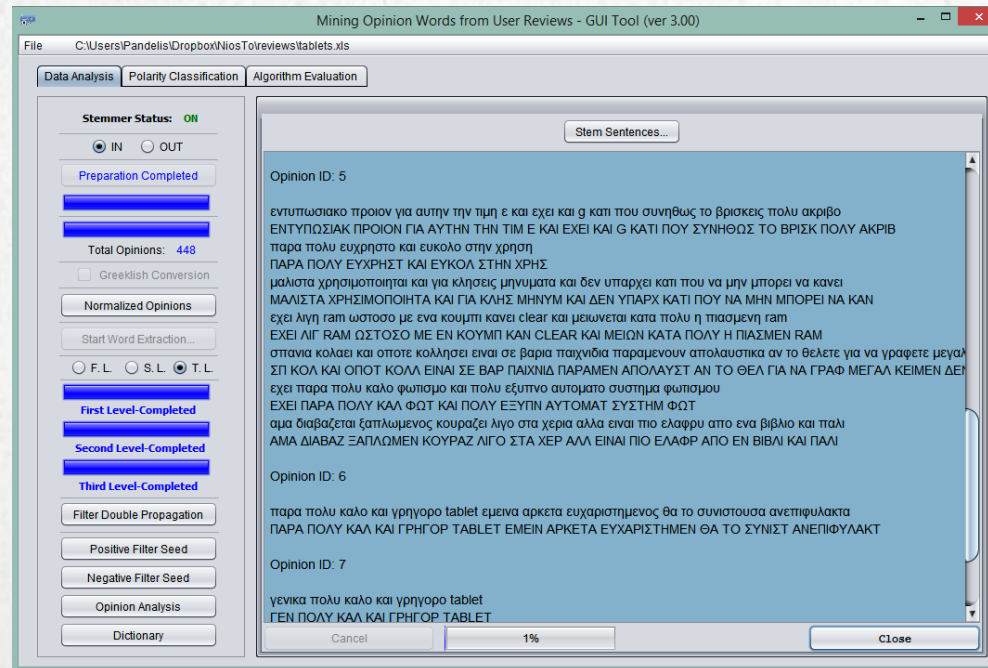
POLARITY CLASSIFICATION OPTIONS



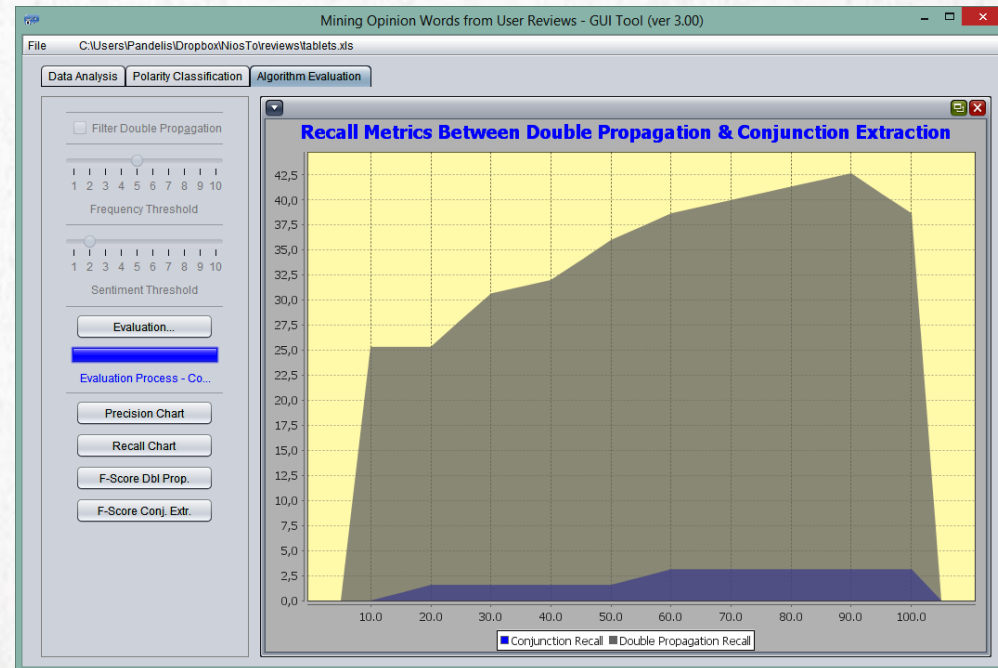
<http://deixto.com/niosto/>

INTERFACE

STEMMER OPINION OPTIONS



ALGORITHM EVALUATION OPTIONS



<http://deixto.com/niosto/>

More...

Mining Domain-Specific Dictionaries

Pantelis Agathangelou, Katakis Ioannis, Fotios Kokkoras, Konstantinos Ntonas

Thank you for
your attention!

DEiXTo - Web Extraction Tool: <http://deixto.com/>



NiosTo – Dictionary Extraction Tool : <http://deixto.com/niosto/>



pandelisagathangelou@gmail.com