Aalto University

Malo, P., Sinha, A., Takala, P., Ahlgren, O., and Lappalainen, I.:

Learning the Roles of Directional Expressions and Domain Concepts in Financial News Analysis

---

Pyry Takala ICDM Dallas Dec 2013

Financial sentiment could be used to predict stock and economic metrics

- Predict stock performance: trading volume, volatility, return
- Predict changes in economic metrics, e.g. consumer confidence
- Assess risk of long-term investments, portfolios etc.

Standard sentiment extraction is not well suited for financial news

Financial news

- Well structured and clear language
- Relatively small vocabulary
- Numbers and expressions on changes in them ubiquitous
- Sentiment not always same as in common language

Text of other domains (Politics, sports, short stories…)

User-generated texts (Twitter, movie reviews by users, …)

Nearly 5,000 sentences were annotated for sentiment (1/2)

Examples of dataset

Net sales revenue per passenger is expected to increase.
Cash flow from operations totalled EUR 2.71 mn, compared to a negative EUR 0.83 mn in the corresponding period in 2008.
Cardona slowed her vehicle, turned around and returned to the intersection, where she called 911.
In addition to verification of an identity and digital signatures, new state-approved Mobile ID enables to cast votes in elections as well.
However, the growth margin slowed down due to the financial crisis.
According to Finnish petrol station chain St1's managing director Kim Wiio, the company was forced to make purchases with rising prices in the first half of 2008, and now consumer prices are going down almost daily due to competition.

Nearly 5,000 sentences were annotated for sentiment (2/2)

Examples of dataset

Net sales revenue per passenger is expected to increase.
Cash flow from operations totalled EUR 2.71 mn, compared to a negative EUR 0.83 mn in the corresponding period in 2008.
Cardona slowed her vehicle, turned around and returned to the intersection, where she called 911.
In addition to verification of an identity and digital signatures, new state-approved Mobile ID enables to cast votes in elections as well.
However, the growth margin slowed down due to the financial crisis.
According to Finnish petrol station chain St1's managing director Kim Wiio, the company was forced to make purchases with rising prices in the first half of 2008, and now consumer prices are going down almost daily due to competition.

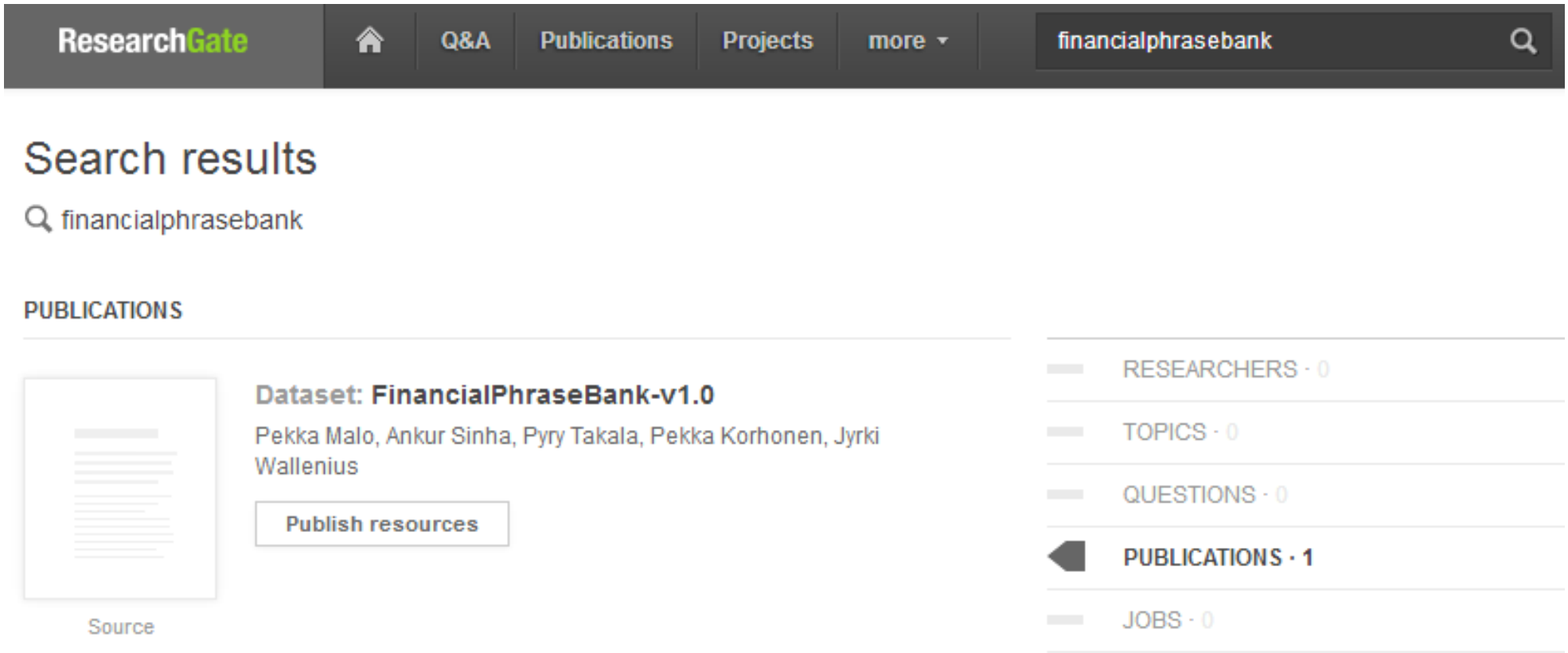
- ~5,000 sentences annotated for sentiment
- Positive, neutral, negative
- Sentences selected randomly from financial news and press releases of stock-exchange listed companies

Annotation quality was strictly monitored to ensure quality output

Annotators knowledgeable on financial news	✓
16[1] annotators with various cultural backgrounds	✓
Annotators unbiased	✓
All annotators instructed in same way	✓
FinancialPhraseBank includes a dataset with sentences where all annotators agree	✓

1) Currently available dataset based on 16 annotators, 3 annotators at time of writing

FinancialPhraseBank is available for your use on ResearchGate



A set of lexicons were used to assist classification

Lexicon		Example
General polarity-bearing expressions	■ Positive, negative, neutral ■ MPQA / Loughran & McDonald's dictionary	■ MPQA: *He started to think about his car debt.* ■ L&M: *The company's debt at end of the period was 100EURm.*

A set of lexicons were used to assist classification

Lexicon		Example
General polarity-bearing expressions	■ Positive, negative, neutral ■ MPQA / Loughran & McDonald's dictionary	■ MPQA: *He started to think about his car debt.* ■ L&M: *The company's debt at end of the period was 100EURm.*
Financial entities	■ Prior polarity and directional dependence (e.g. "positive-if-up")	■ *The company halved its operating loss due to…*

Directional dependence: negative-if-up

Prior polarity: neutral

A set of lexicons were used to assist classification

Lexicon		Example
General polarity-bearing expressions	■ Positive, negative, neutral ■ MPQA / Loughran & McDonald's dictionary	■ MPQA: *He started to think about his car debt.* ■ L&M: *The company's debt at end of the period was 100EURm.*
Financial entities	■ Prior polarity and directional dependence (e.g. "positive-if-up")	■ *The company halved its operating loss due to…*
Direction of events	■ Up & down categories	

Directional dependence: negative-if-up

Prior polarity: neutral

Direction: down

A set of lexicons were used to assist classification

Lexicon		Example
General polarity-bearing expressions	■ Positive, negative, neutral ■ MPQA / Loughran & McDonald's dictionary	■ MPQA: *He started to think about his car debt.* ■ L&M: *The company's debt at end of the period was 100EURm.*
Financial entities	■ Prior polarity and directional dependence (e.g. "positive-if-up")	■ *The company halved its operating loss due to…* Directional dependence: negative-if-up Prior polarity: neutral
Direction of events	■ Up & down categories	Direction: down
Negators, uncertainty and modals	■ Negators, boosters and diminishers, modal operators, uncertainty words	■ *The company's results were not bad* Negator

Detailed next

3 types of classification methods were tested

Voting-rule classifier

- Baseline to compare to
- Vote for the category that majority of polarity-bearing expressions falls under

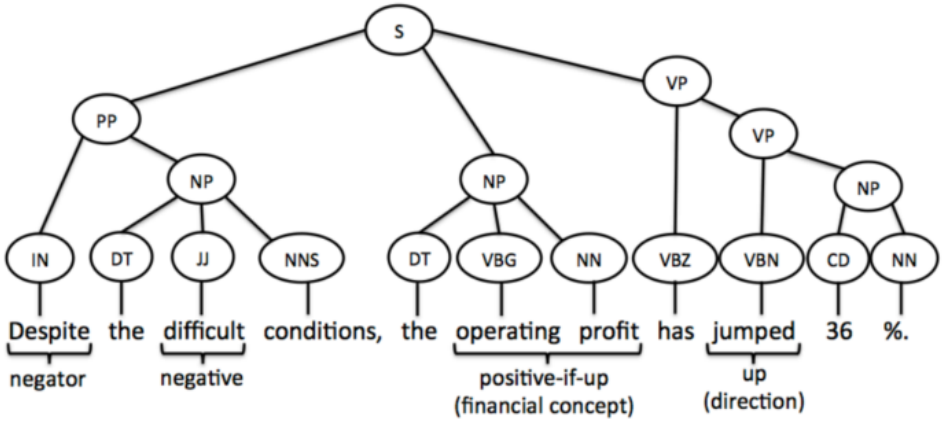
Polarity-interactions

- Tree-structure of sentences: how do concepts relate to each other
- Constituent syntactic trees
- Typed-dependency trees

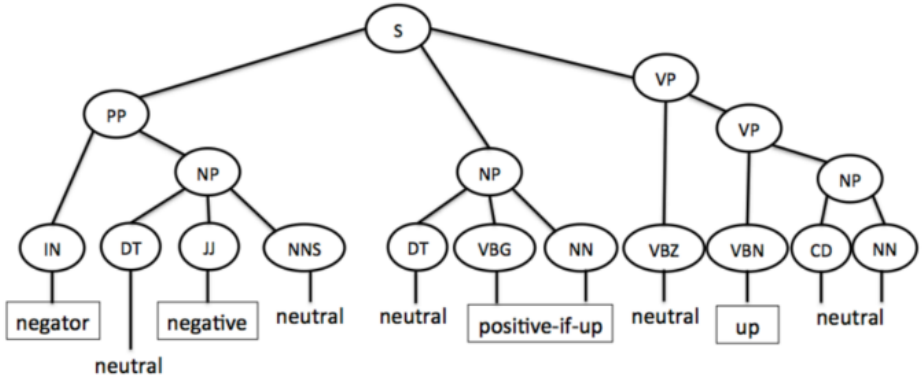
Polarity-orders

- Simple lexical sequence trees

Constituent syntactic trees and Typed-dependency trees were formed, and augmented with polarity-information (1/2)



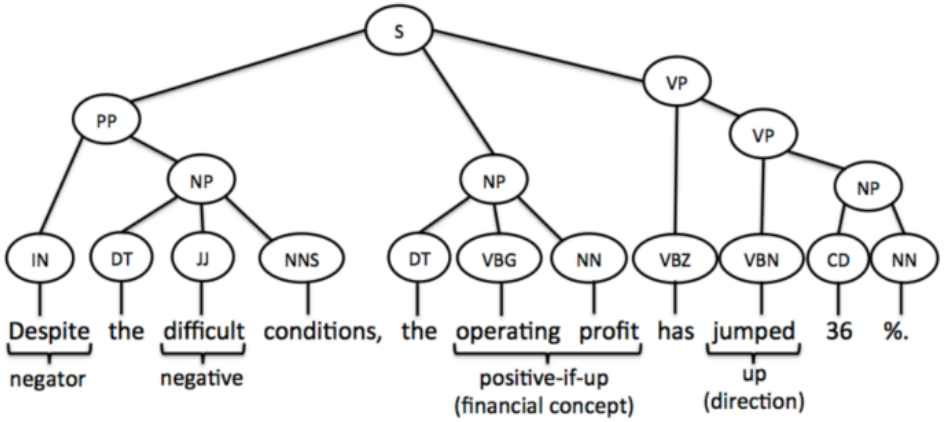
(a) Original constituent syntactic tree



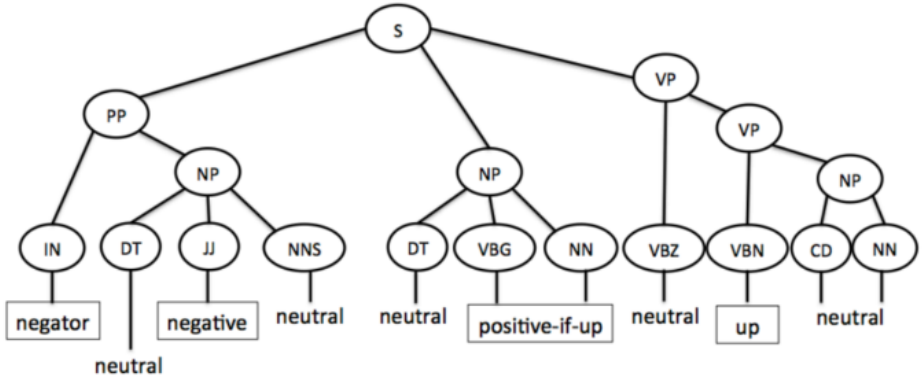
(b) Polarized constituent syntactic tree

Fig. 2. Polarization of constituent syntactic trees.

Constituent syntactic trees and Typed-dependency trees were formed, and augmented with polarity-information (2/2)

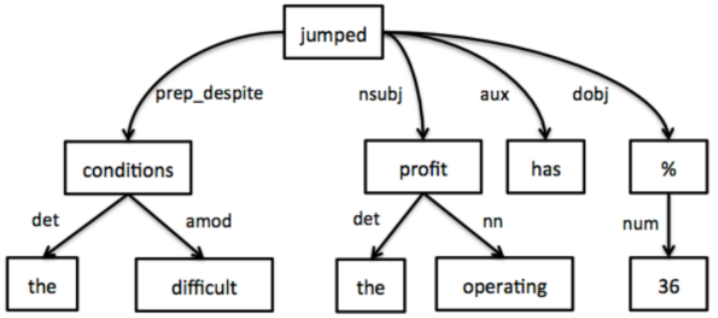


(a) Original constituent syntactic tree

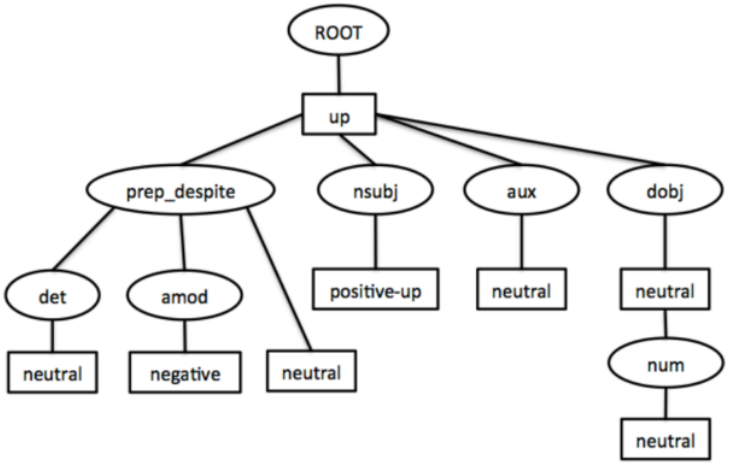


(b) Polarized constituent syntactic tree

Fig. 2. Polarization of constituent syntactic trees.



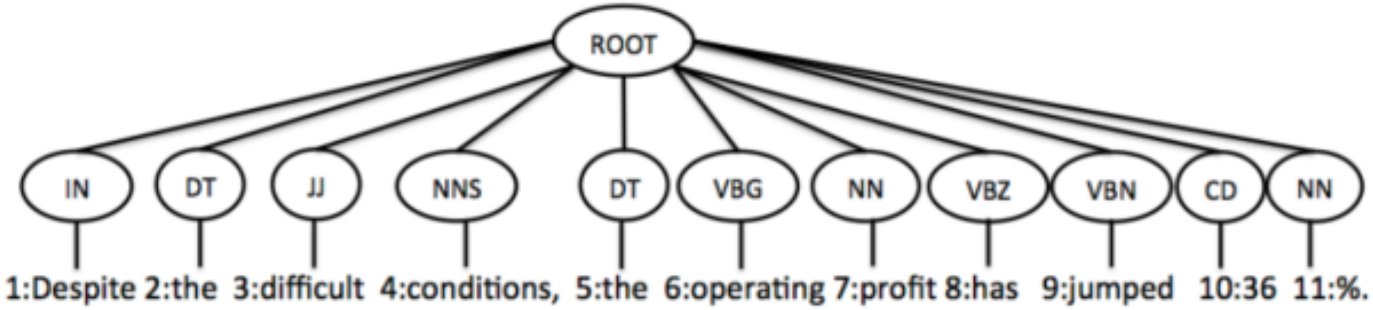
(a) Original typed-dependency graph



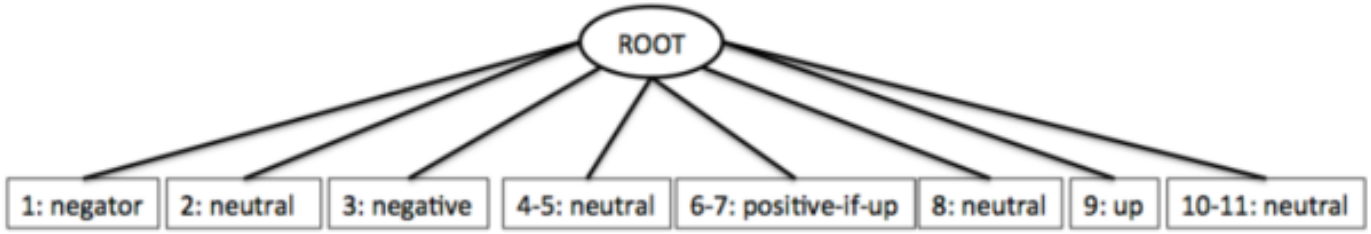
(b) Polarized typed-dependency tree

Fig. 3. Polarization of typed-dependency trees.

Lexical sequence trees were also polarized



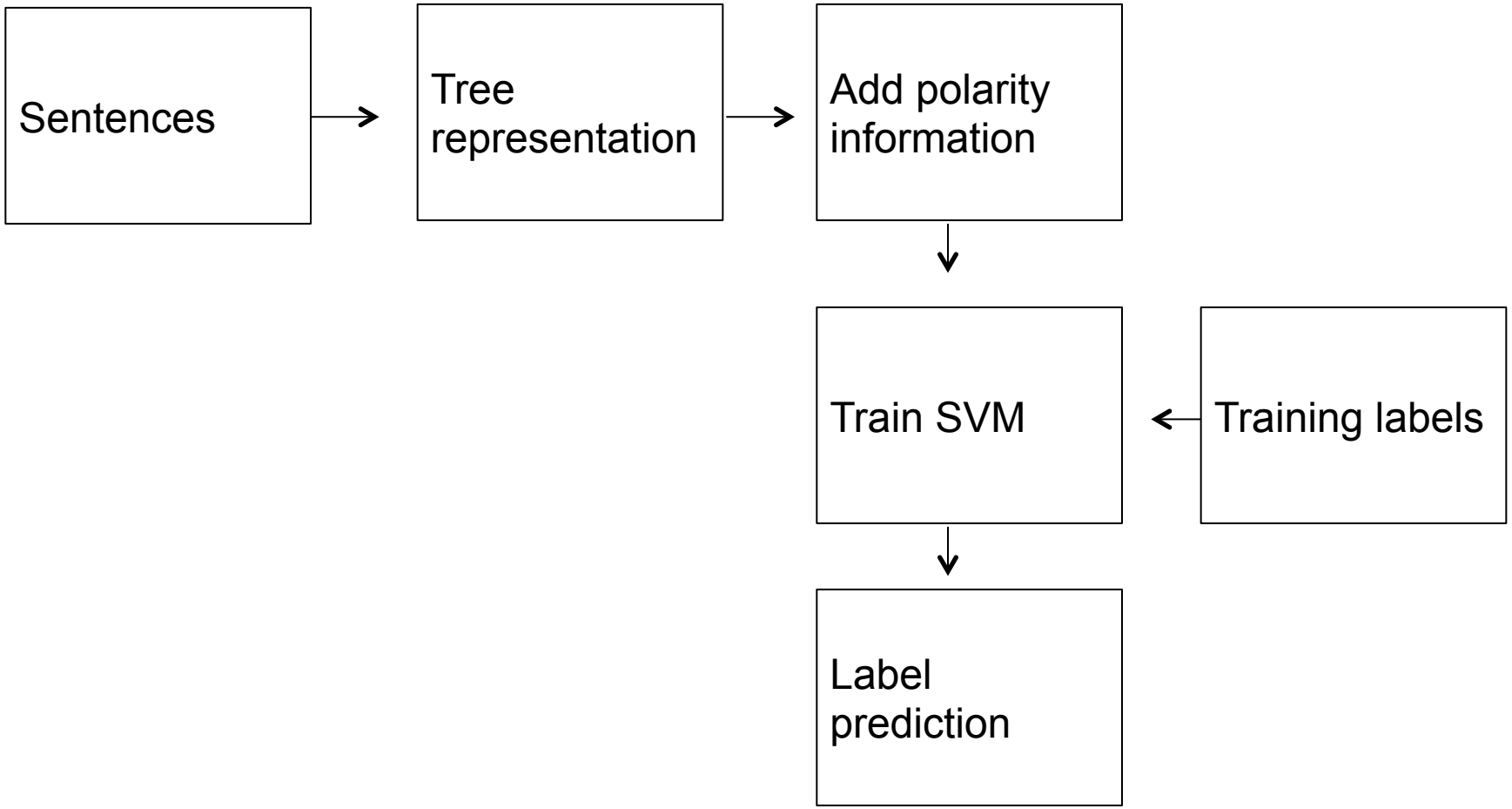
(a) Lexical sequence tree



(b) Polarized lexical sequence tree

Fig. 4. Polarization of lexical sequence trees.

Next, one-against-one SVM was trained with the training data and the tree-structure information



Combined tree models lead to best results (1/2)

	Positive [1)]			
	A	R	P	F
VR-L	0.69	0.19	0.70	0.29
VR-MPQA	0.66	0.60	0.51	0.55
VR-S	0.49	0.61	0.36	0.45
LP-MPQA	0.66	0.53	0.50	0.52
LP-F	0.79	0.70	0.69	0.69
DP-F	0.70	0.35	0.62	0.45
CP-F	0.78	0.58	**0.73**	0.64
DP-LP-F	0.81	0.83	0.69	0.75
CP-LP-F	**0.82**	**0.85**	0.69	**0.77**

Algorithms

VP = Voting

LP = Sequence tree

CP = Constituent syntactic tree

DP = Typed-dependency tree

Lexicons

L = Loughran & McDonald

MPQA = MPQA (by Wilson)

S = SentiWordNet

F = Financial database

Metrics

A = Accuracy

R = Recall

P = Precision

F = F1-score

1) Neutral and negative results show similar patterns

Combined tree models lead to best results (2/2)

	Positive [1]			
	A	R	P	F
VR-L	0.69	0.19	0.70	0.29
VR-MPQA	0.66	0.60	0.51	0.55
VR-S	0.49	0.61	0.36	0.45
LP-MPQA	0.66	0.53	0.50	0.52
LP-F	0.79	0.70	0.69	0.69
DP-F	0.70	0.35	0.62	0.45
CP-F	0.78	0.58	**0.73**	0.64
DP-LP-F	0.81	0.83	0.69	0.75
CP-LP-F	**0.82**	**0.85**	0.69	**0.77**

Domain knowledge is useful

Interaction kernels enhance performance

Sequence is more important than interaction

Sequence and interactions together yield best results

Model also generalizes well outside financial domain

Algorithms

VP = Voting

LP = Sequence tree

CP = Constituent syntactic tree

DP = Typed-dependency tree

Lexicons

L = Loughran & McDonald

MPQA = MPQA (by Wilson)

S = SentiWordNet

F = Financial database

Metrics

A = Accuracy

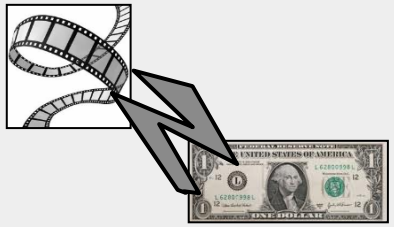
R = Recall

P = Precision

F = F1-score

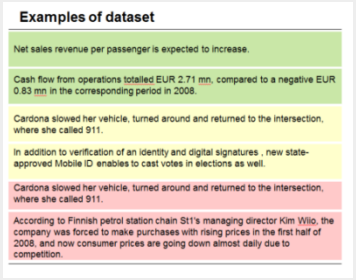
1) Neutral and negative results show similar patterns

To summarize…



- Sentiment in finance is not equal to sentiment in movie reviews

- Tree-kernels, augmented with polarity information, can used to classify text for sentiment



- You are welcome to test your methods with our data (available on ResearchGate: “FinancialPhraseBank”)

Presenter:

pyry.takala@aalto.fi

Lead author:

pekka.malo@aalto.fi