Sentiment Analysis on Financial Data Using Neural Networks
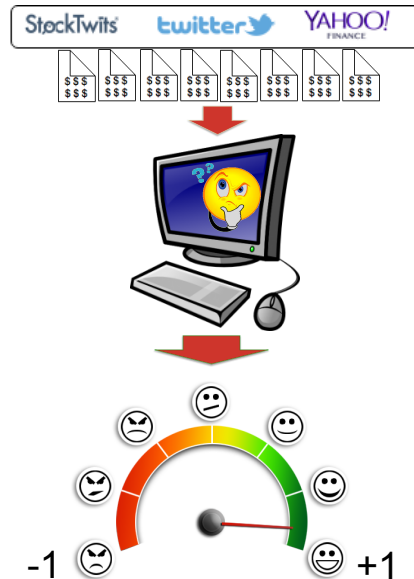
Sudipta Kar and Suraj Maharjan and Thamar Solorio
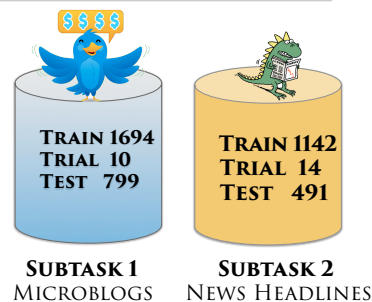
Department of Computer Science
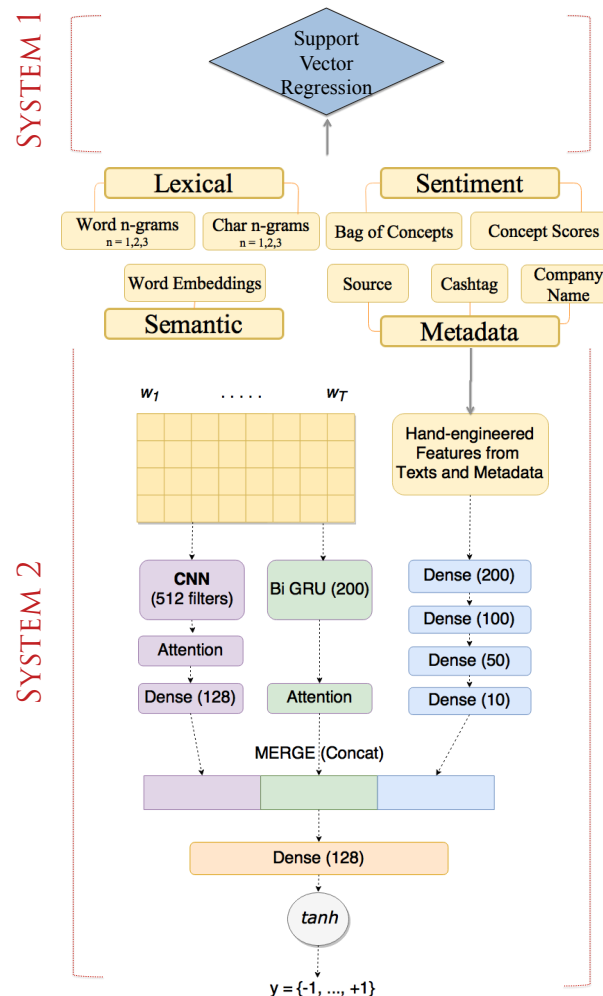UNIVERSITY OF HOUSTON
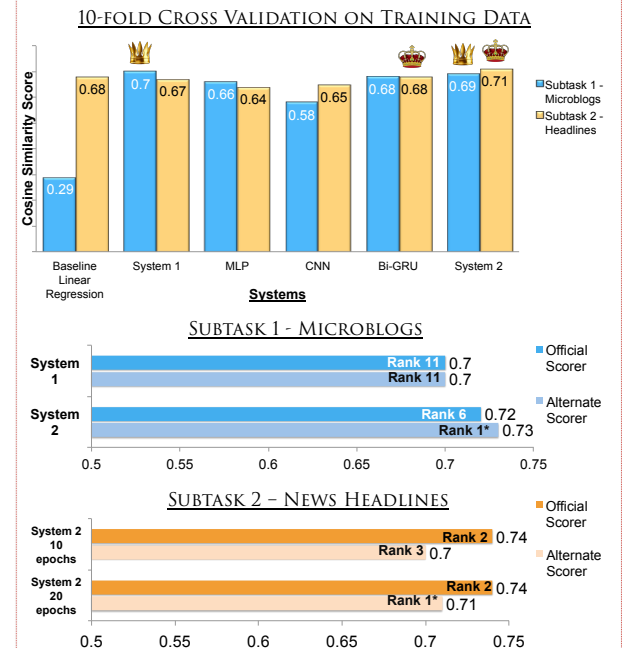{skar3, smaharjan2, tsolorio}@uh.edu

The Problem

The Dataset

Train 1694
Trial 10
Test 799

Train 1142
Trial 14
Test 491

Subtask 1
Microblogs

Subtask 2
News Headlines

Proposed Models

SYSTEM 1

Support Vector Regression

Lexical

Word n-grams
n = 1,2,3

Char n-grams
n = 1,2,3

Word Embeddings

Semantic

Sentiment

Bag of Concepts | Concept Scores

Source | Cashtag | Company Name

Metadata

SYSTEM 2

w_1 . . . . . w_T

Hand-engineered Features from Texts and Metadata

CNN
(512 filters)

Bi GRU (200)

Dense (200)

Attention

Attention

Dense (100)

Dense (128)

Attention

Dense (50)

Dense (10)

MERGE (Concat)

Dense (128)

tanh

$y = \{-1, ..., +1\}$

Results

10-fold Cross Validation on Training Data

■ Subtask 1 - Microblogs
■ Subtask 2 - Headlines

Subtask 1 - Microblogs

System 1 — Rank 11 0.7 / Rank 11 0.7
System 2 — Rank 6 0.72 / Rank 1* 0.73

■ Official Scorer
■ Alternate Scorer

Subtask 2 – News Headlines

System 2 10 epochs — Rank 2 0.74 / Rank 3 0.7
System 2 20 epochs — Rank 2 0.74 / Rank 1* 0.71

■ Official Scorer
■ Alternate Scorer

Conclusion

o Simple machine learning algorithm like Linear Regression worked well for cleaner texts like news headlines.
o Analyzing sentiments for a company from microblog data can be improved by analyzing the related company's statistics provided in the posts.

Example
Worst performers today: $RIG -13% $EK -10% $MGM $IO -6% $CAR -5,5% / best stock: $WTS +15%
Cashtag : $WTS
Annotated Sentiment : 0.857
Predicted Sentiment : -0.365

Find More Details here