Fake News Challenge

Elliot Greenlee and Patricia Eckhart

CS571 Pattern Recognition

May 4, 2017

THE UNIVERSITY OF TENNESSEE KNOXVILLE

BIG ORANGE. BIG IDEAS.®

About Us

Elliot

- MS in Computer Science
- PhD in the future
- Rock climbing, cooking, travel
- Security research at ORNL
- Machine learning and computer vision GRA with Dr. Hairong Qi

Patricia

- MS in Computer Engineering
- Bodenheimer Fellow
- GRA in Neuromorphic computing group
- Mother of an 8 month old

Motivation

- Fake news discussion is increasing
- Causes confusion amongst adults
- Google, Facebook looking for solutions
- Education initiatives underway
- First step is human in the loop systems

Challenge - FNC 1

- Consists of volunteers and teams from industry and academia
- Stage 1, stance detection
- Classify headlines and body texts based on their relatedness

Challenge

EXAMPLE HEADLINE

"Robert Plant Ripped up $800M Led Zeppelin Reunion Contract"

EXAMPLE SNIPPETS FROM BODY TEXTS AND CORRECT CLASSIFICATIONS

"... Led Zeppelin's Robert Plant turned down £500 MILLION to reform supergroup. ..."

CORRECT CLASSIFICATION: AGREE

"... No, Robert Plant did not rip up an $800 million deal to get Led Zeppelin back together. ..."

CORRECT CLASSIFICATION: DISAGREE

"... Robert Plant reportedly tore up an $800 million Led Zeppelin reunion deal. ..."

CORRECT CLASSIFICATION: DISCUSSES

"... Richard Branson's Virgin Galactic is set to launch SpaceShipTwo today. ..."

CORRECT CLASSIFICATION: UNRELATED

Data

- 40,350 sample training pairs: headlines and body text snippets
- Hand classified
- 10-fold cross validation
- Preprocessing features
- 79.53% using gradient boosting

Methods

- MPP (3) - Python
- KNN - Python
- BPNN - Matlab
- Decision Tree - Matlab
- SVM - Matlab
- Clustering (3) - Python
- Classifier fusion (3) - Python/Matlab
- Dimensionality Reduction (1) - Python

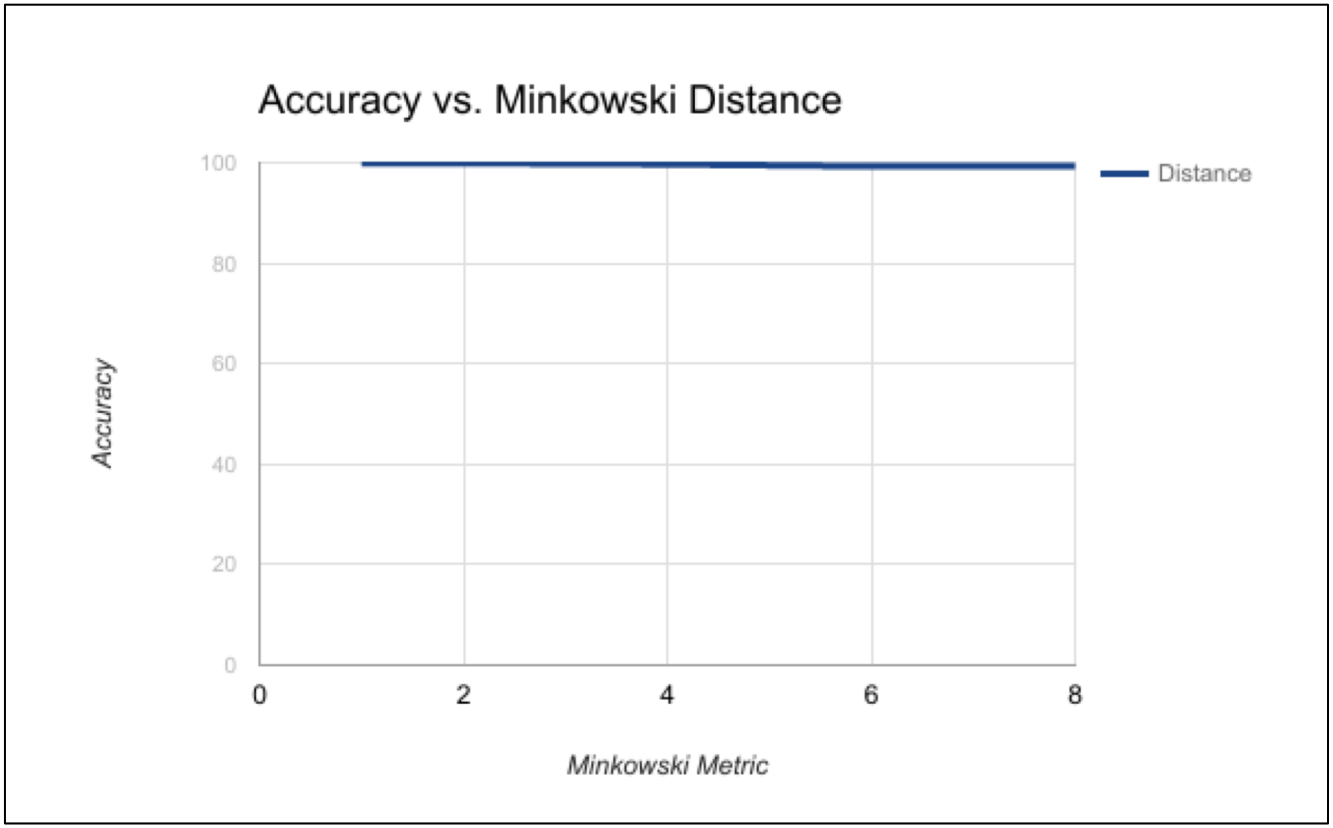
Effects of Prior Probability

Changing Class

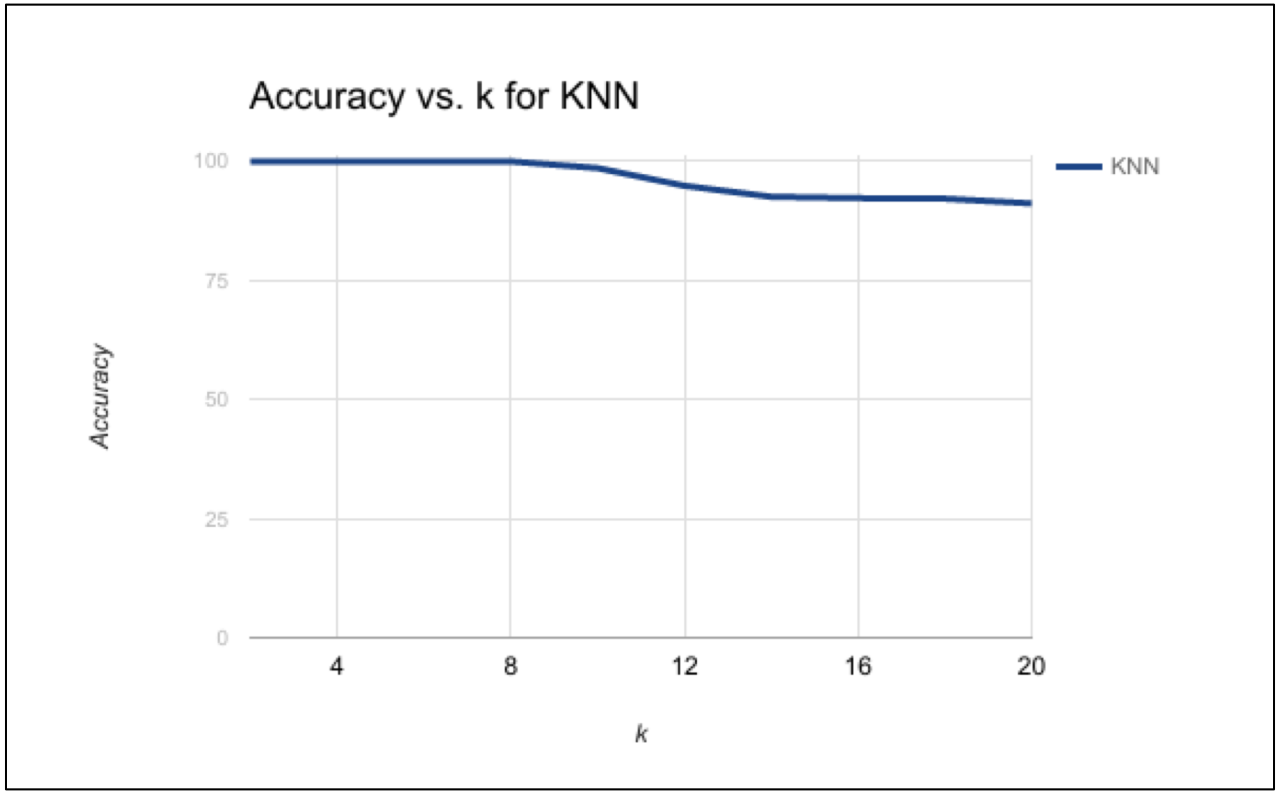
Changing Intensity

Class 1	Class 2	Class 3	Class 4
[0.07, 0.02, 0.18, 0.73]	*	*	*
78.5%	78.5%	78.5%	78.5%
[0.1, 0.3, 0.3, 0.3]	[0.3, 0.1, 0.3, 0.3]	[0.3, 0.3, 0.1, 0.3]	[0.3, 0.3, 0.3, 0.1]
78.3%	79.6%	75.9%	77.3%
[0.25, 0.25, 0.25, 0.25]	*	*	*
78.6%	78.6%	78.6%	78.6%
[0.4, 0.2, 0.2, 0.2]	[0.2, 0.4, 0.2, 0.2]	[0.2, 0.2, 0.4, 0.2]	[0.2, 0.2, 0.2, 0.4]
77.4%	77.0%	79.4%	78.6%
[0.7, 0.1, 0.1, 0.1]	[0.1, 0.7, 0.1, 0.1]	[0.1, 0.1, 0.7, 0.1]	[0.1, 0.1, 0.1, 0.7]
74.6%	71.9%	77.9%	77.4%

Effects of Minkowski Distance



KNN Effects of neighbors



Dimensionality Reduction

	Normalized	PCA 4 (10%)	PCA 6 (5%)	PCA 9 (1%)
MPP Case 1	76.7%	80.7%	80%	78.7%
MPP Case 2	83.7%	83.2%	83.4%	83.6%
MPP Case 3	79.2%	83.4%	83.6%	83.6%
Decision Tree	84.2%	81.2%	81.2%	81.4%
SVM	86.9%	85.6%	85.6%	86.4

Best Results

MPP Case 1	80.7%
MPP Case 2	86.8%
MPP Case 3	83.6%
KNN	99.9%
BPNN	85.1%
Decision Tree	84.2%
SVM	86.9%
K-means	56.7%
Winner Takes All	50.7%
Kohonen	56.0%
Fusion Voting	86.9%
Fusion Naive Bayes	7%
Fusion BKS	88.0%

Future Work

- Testing data comes out June 1st
- Make our framework cohesive
- Improve features
  - Point of view, sentence voice, tense
- Try other methods
  - LSTM Networks
- Tailor methods to the problem

Fake News Challenge

Elliot Greenlee and Patricia Eckhart

CS571 Pattern Recognition

May 4, 2017

THE UNIVERSITY OF TENNESSEE KNOXVILLE

BIG ORANGE. BIG IDEAS.®