THE OHIO STATE UNIVERSITY

Data Visualization for Twitter Sentiment Analysis by Denali

OUTLINE:

- Background
- Data source
- Denali Introduction
- Experiments Setup and Results
- Conclusion

BACKGROUND

- **Twitter sentiment analysis on candidate in GOP debates**
- **3 sentiment polarities**
  - Positive
  - Negative
  - Neutral
- **Feature extraction and sentiment score computation**
- **Represent tweet as a sparse feature vector**
- **3-parallel DNN architecture**
- **Visualize high dimensional data**
  - Feature vectors
  - Parameter space

DATA SOURCE

- **Raw Twitter Data**
  - A labeled data set with 1,288 tweets from Kaggle
  - Focus on the most controversial “Donald Trump”

Tweets	Sentiment	Sentiment Confidence
@JGreenDC @realDonaldTrump In all fairness #BillClinton owns that phrase.#GOPDebate	Negative	0.6332
@MsPackyetti: Donald Trump's campaign reveals 1 important thing: Twitter Trolls are real people.	Positive	0.6957

Table 1 Example for Labeled Data

DATA SOURCE (cont.)

- **Representing Feature Vector in Feature Space**
  - Features collection induce a feature space
  - Take the #occurrences into account
  - Embed text sample as feature vector
  - Sparse feature vector with dimension of 3198

Tweets	"RT@TrumpIssues #GOPDebate #2016Debate elect Trump, elect Trump! Only he and @SarahPalinUSA can save America!"										
Feature	…	america	…	elect	…	gopdebate	…	save	…	trump	…
#Occurrence		1	0	2	0	1	0	1	0	2	0

Table 2 Example for Feature Vectors

DATA SOURCE (cont.)

- **Parameter Space induced from a 3-Parallel-DNN Architecture**
  - Different # epochs give different configurations of the architecture

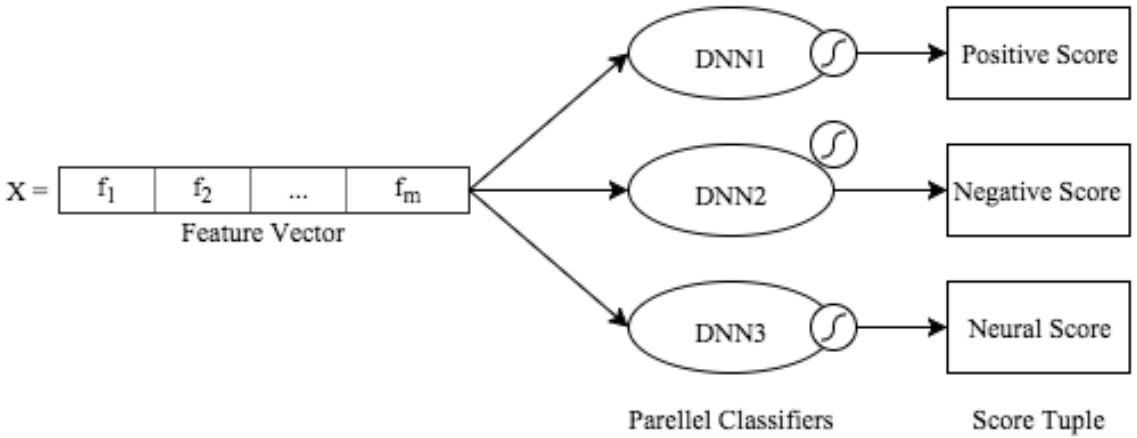


Figure 1-1 3-Parallel-DNN Architecture

DATA SOURCE (cont.)

- **2 hidden layer DNN: (3198,100), (100, 50)**
- $3 \times (3199 * 100 + 101 * 50 + 51) = 3 \times 325001 = 975003$

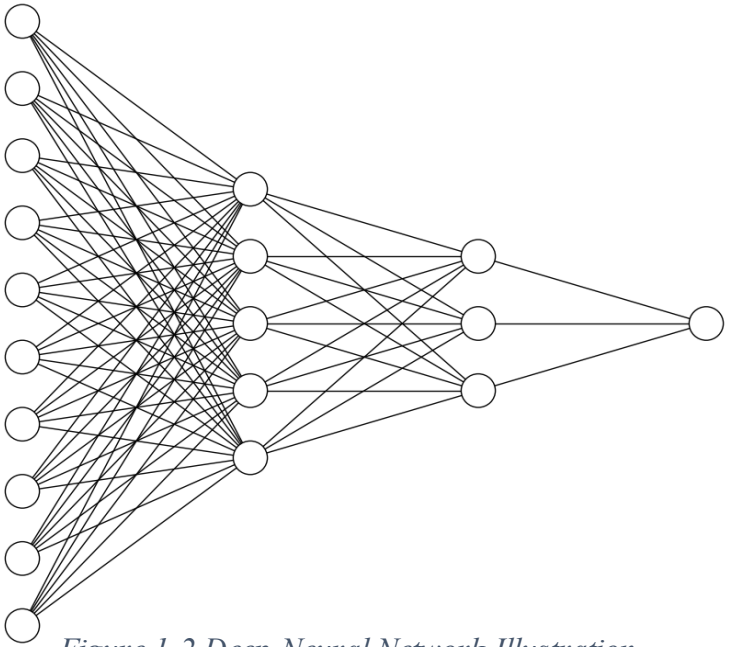


Figure 1-2 Deep Neural Network Illustration

- **A tool for visualizing trees as landscape metaphors**
- **Two requisites for the data**
  - Tree structure extractable, usually hierarchical
  - Scalar function defined on it
- **Two common approaches to extract tree structure**
  - Hierarchical Clustering Tree
  - Contour Tree

EXPERIMENT SETUPS AND RESULTS

- **Visualizing Contour Tree Extracted from Feature Vectors**
  - A topological space X and function $f: X \rightarrow R$
  - X is1,288 feature vectors of dimension 3198
  - f is the variance for the score of 3 sentiment polarities
  - Sentiment score range is $[0,1]$
  - Scalar function range is $[0.0042, 0.2222]$

EXPERIMENT SETUPS AND RESULTS (cont.)

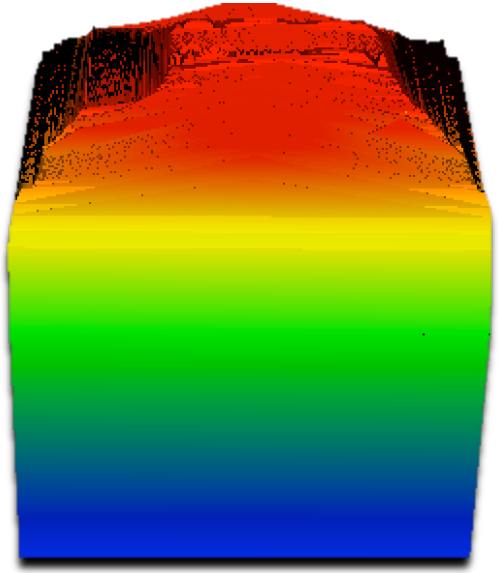


Figure 2 Contour Tree Extracted from Feature Vectors

EXPERIMENT SETUPS AND RESULTS (cont.)

- **Visualizing Hierarchical Clustering Tree Extracted from Feature Vectors**
  - Applying average linkage clustering algorithm
  - A hierarchical clustering tree is extracted base on the dissimilarity between feature vectors
  - Born with scalar function defined (the distance where merges)

EXPERIMENT SETUPS AND RESULTS (cont.)

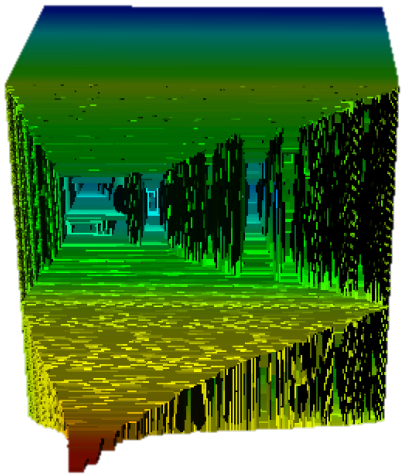


Figure 3 HCT (Euclidean)

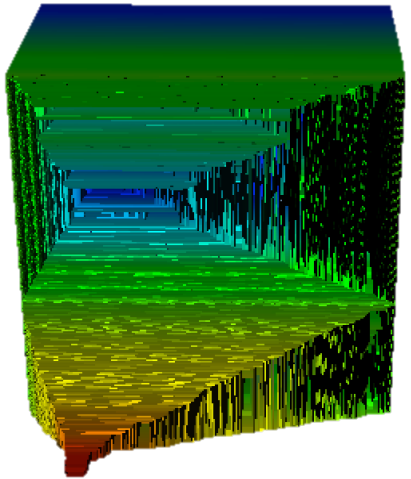


Figure 4 HCT (Manhattan)

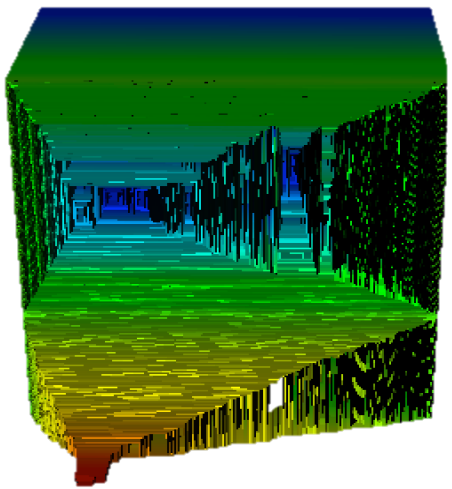


Figure 5 HCT (Hamming)

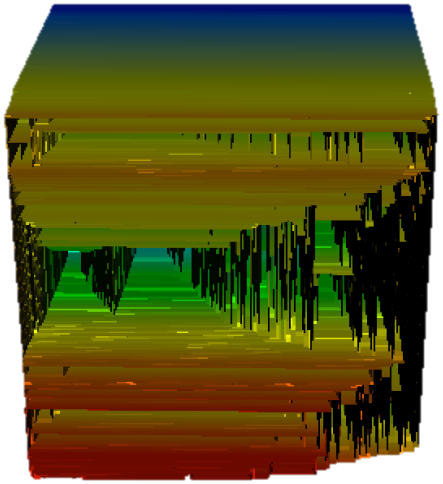


Figure 6 HCT (Cosine)

EXPERIMENT SETUPS AND RESULTS (cont.)

- **Visualizing the parameter space induced by 3-parallel-DNN Architecture**
  - The cost function is defined on the parameter space
  - Fig 7 uses testing error as scalar value
  - Fig 8 uses training error as color map (MAX scalar value is red & MIN scalar value is blue
  - Blue region are suffering from over-fitting & Red region are suffering from under-fitting

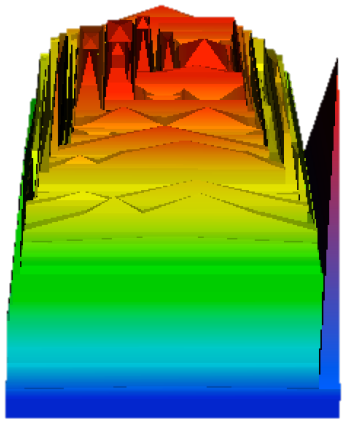


Fig 7 Contour Tree for Testing Error

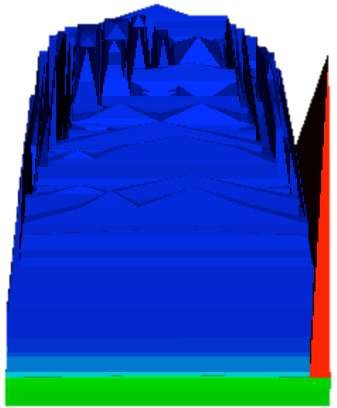
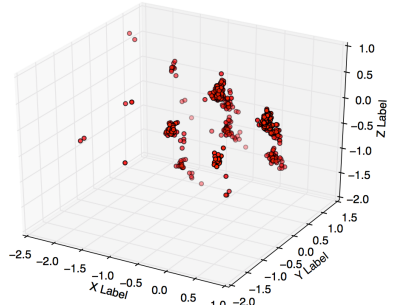


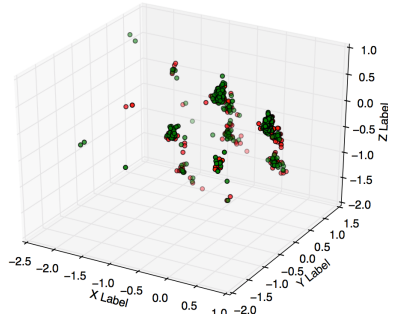
Fig 8 Configured with Color Map for Training Error

EXPERIMENT SETUPS AND RESULTS (cont.)

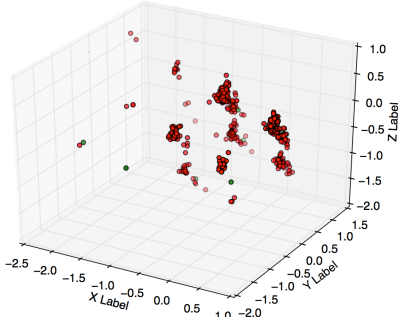
- Apply PCA to the feature vector
- Red means being correctly classified while green means being wrongly classified
- Upper plots shows the performance on training data while lower plots shows the performance on testing data



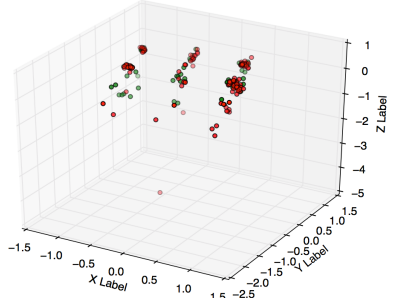
Over-fitting

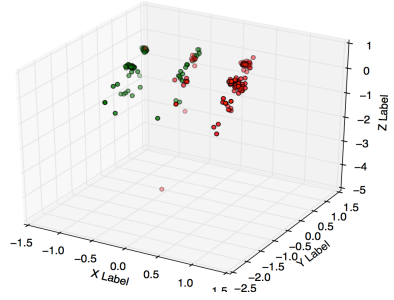


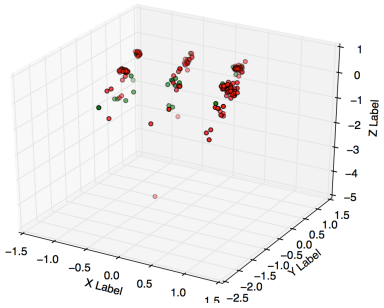
Under-fitting



Optimal







CONCLUSION:

- Denali can be used to visualize high dimensional data
- Visualize by extracting tree structure
  - hierarchical structure tree
  - Contour tree

Q & A