A Multi-task Approach to Predict Likability of Books

Suraj Maharjan[1] John Arevalo[2] Fabio A. González[2]
Manuel Montes-y-Gómez[3] Thamar Solorio[1]

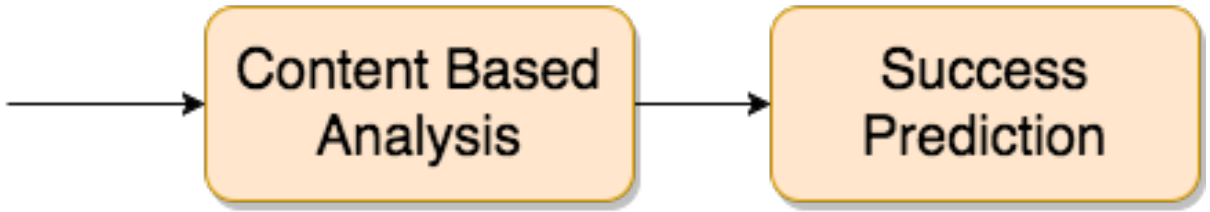
[1]University of Houston
[2]Universidad Nacional de Colombia
[3]Instituto Nacional de Astrofisica Optica y Electronica

EACL 2017
April 7, 2017

Outline

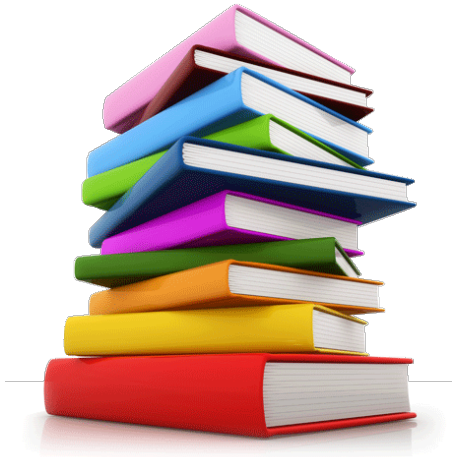
Outline

Task



Motivation

- Publication Industry
- Writers
- Readers
- Book Discovery
- Book Recommendations



Outline

Success

Ashok et al. (2013) Definition

$$Success(book) = \begin{cases} 1, & \text{if } download\_count \geq T_1 \\ 0, & \text{if } download\_count < T_2 \end{cases}$$

Success

Ashok et al. (2013) Definition

$$Success(book) = \begin{cases} 1, & \text{if } download\_count \geq T_1 \\ 0, & \text{if } download\_count < T_2 \end{cases}$$

Our Definition

$$Success(book) = \begin{cases} 1, & \text{if } avg\_rating \geq 3.5, \#\_reviews \geq 10 \\ 0, & \text{if } 0.0 \leq avg\_rating < 3.5, \#\_reviews \geq 10 \end{cases}$$

Outline

Dataset

Genre	Unsuccessful	Successful	Total
Detective Mystery	60	46	106
Drama	29	70	99
Fiction	30	81	111
Historical Fiction	16	65	81
Love Stories	20	60	80
Poetry	23	158	181
Science Fiction	48	39	87
Short Stories	123	135	258
Total	349	654	1,003

Table 1: Goodreads data distribution [1].

[1] Books' content obtained from Project Gutenberg.

Success Definitions Comparison

		Ashok et al. definition	
		Unsuccessful	Successful
Goodreads definition	Unsuccessful	73	32
	Successful	110	184

Table 2: Confusion matrix between two different definitions of success.

Success Definitions Comparison

		Ashok et al. definition	
		Unsuccessful	Successful
Goodreads definition	Unsuccessful	73	32
	Successful	110	184

Table 2: Confusion matrix between two different definitions of success.

- 28 out of 110 books labeled as unsuccessful by Ashok et al. (2013) definition
  - ratings ≥ 3.5
  - reviewed > 100 reviewers

Outline

Setup

- Single Task (ST)
  - Predict Success
- Multi-task (MT)
  - Predict both Success and Genre together
  - Parameter sharing

Hand-crafted Feature

- Lexical
  - Skip-grams, Word and Char n-grams, Typed Char n-grams
  - Term frequency-inverse document frequency (TF-IDF)

Hand-crafted Feature

- Lexical
  - Skip-grams, Word and Char n-grams, Typed Char n-grams
  - Term frequency-inverse document frequency (TF-IDF)
- Syntactic
  - Constituents, Phrasal, Clausal, Production Rules

Hand-crafted Feature

- Lexical
  - Skip-grams, Word and Char n-grams, Typed Char n-grams
  - Term frequency-inverse document frequency (TF-IDF)
- Syntactic
  - Constituents, Phrasal, Clausal, Production Rules
- Sentiment
  - SentiWordNet
  - SenticNet Concepts

Hand-crafted Feature

- Lexical
  - Skip-grams, Word and Char n-grams, Typed Char n-grams
  - Term frequency-inverse document frequency (TF-IDF)
- Syntactic
  - Constituents, Phrasal, Clausal, Production Rules
- Sentiment
  - SentiWordNet
  - SenticNet Concepts
- Readability
  - Gunning Fog Index, Flesch Reading Ease, Flesch Kincaid Grade Level, RIX, LIX, ARI and Smog Index

Hand-crafted Feature

- Lexical
  - Skip-grams, Word and Char n-grams, Typed Char n-grams
  - Term frequency-inverse document frequency (TF-IDF)
- Syntactic
  - Constituents, Phrasal, Clausal, Production Rules
- Sentiment
  - SentiWordNet
  - SenticNet Concepts
- Readability
  - Gunning Fog Index, Flesch Reading Ease, Flesch Kincaid Grade Level, RIX, LIX, ARI and Smog Index
- Writing Density
  - Count of words, characters, uppercase words, exclamations, question marks
  - Average word length, sentence length, words per sentence
  - Lexical diversity

Neural Network Learned Representations

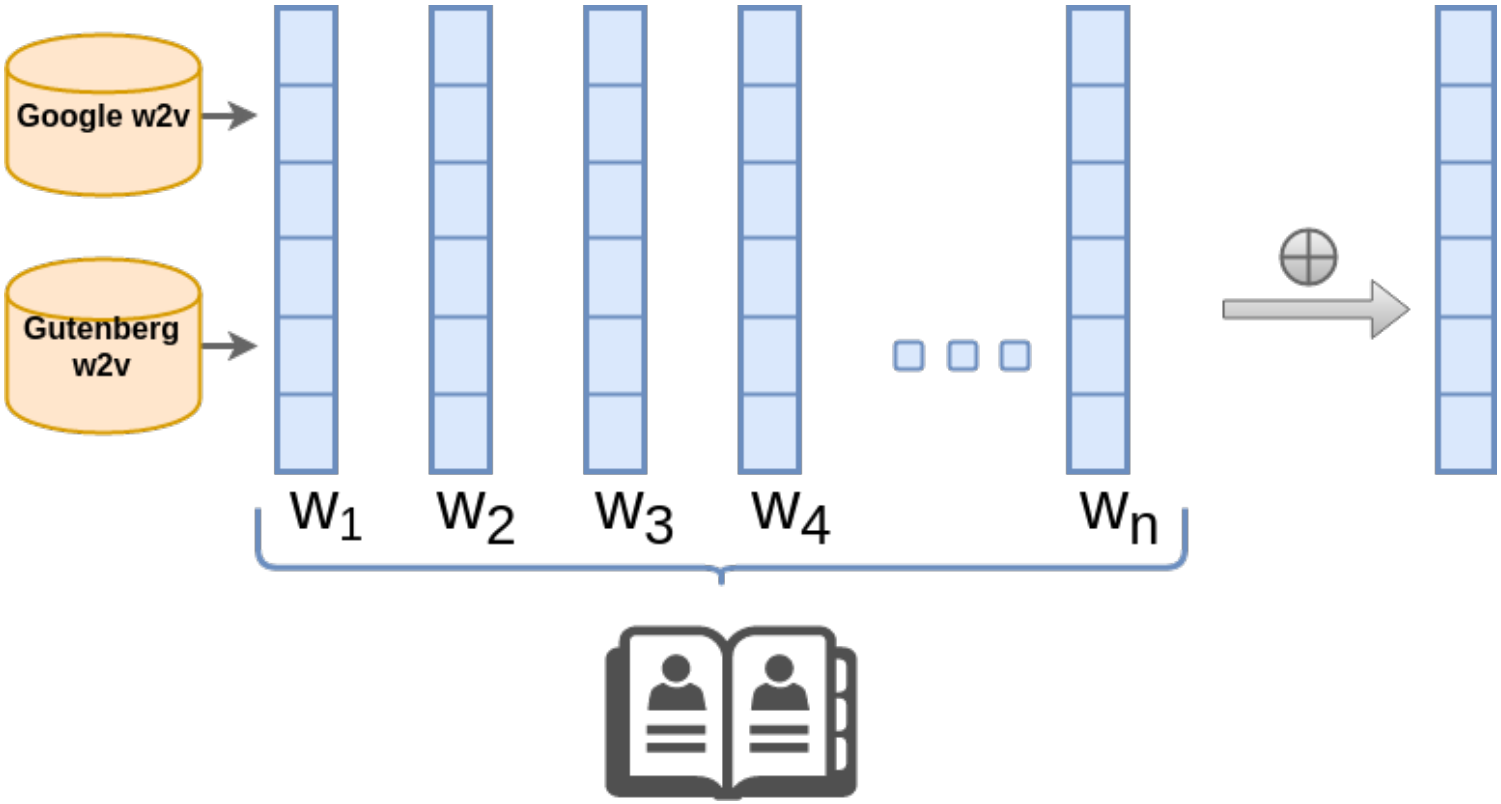


Figure 1: Word Embeddings.

Neural Network Learned Representations

Figure 2: Book Vector using Doc2Vec.

Multi-task RNN

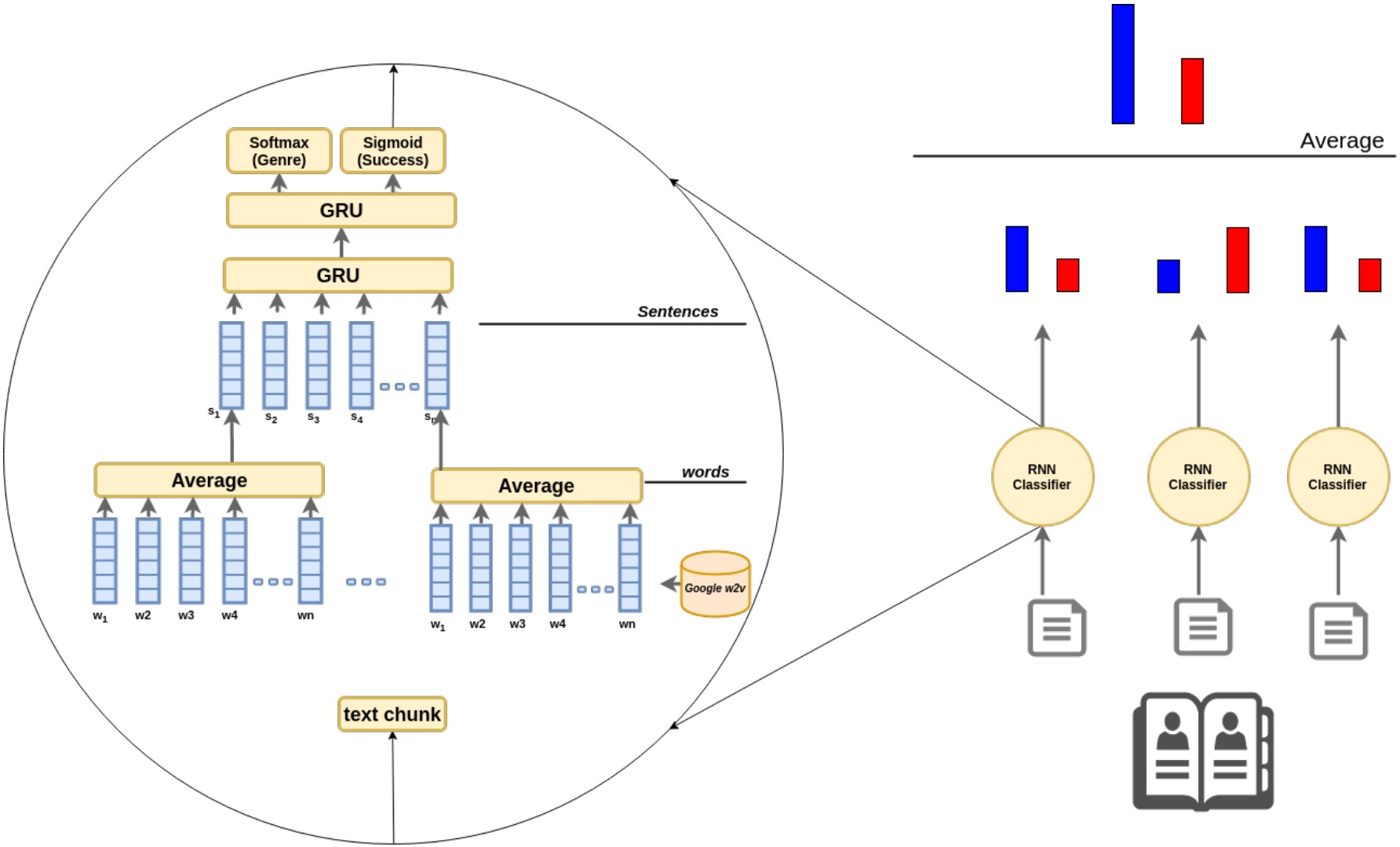


Figure 3: Multi-task RNN Representations.

Experimental Setup

- Stratified random data splitting into 70:30 training/test
- Linear SVM (Binary Classification)
  - C parameter tuned through grid search
- Elastic Net regression algorithm (Regression)
  - $l1\_ratio$ tuned through grid search
- Trained 25 models
  - random hyper-parameter initialization for learning rate, weights initialization ranges and regularization parameters

Outline

Individual Feature Performance

Features	**Single task (F1)**	**Multi-task (F1)**	**MSE**
Word Bigram	0.659	0.685	0.152
Char 3 gram	0.669	0.700	0.155
Typed mid_word 3 gram	0.653	0.687	0.156
Clausal	0.506	0.558	0.156
Writing Density (WR)	0.605	0.640	0.156
Readability (R)	0.506	0.634	0.144
SentiWordNet Sentiments(SWN)	0.582	0.610	0.156
Sentic Concepts and Scores (SCS)	0.657	0.670	0.155
Book2Vec (DBoW)	0.643	0.654	0.130
Book2Vec (DMM)	0.686	**0.731**	0.142
Book2Vec (DBoW+DMC)	0.647	0.677	0.131
Book2Vec (DBoW+DMM)	**0.695**	0.729	0.142
RNN	0.529	0.686	**0.125**

Feature Combination Performance

Features	Single task (F1)	Multi-task (F1)	MSE
Unigram+Bigram	0.660	0.691	0.15
Unigram+Bigram+Trigram	0.660	0.700	0.149
Char 3,4,5 gram	0.682	0.689	0.153
All Typed ngram	0.663	0.691	0.144
SCS+WR+Typed mid word	**0.720**	0.710	0.155
SCS+Book2Vec	0.695	0.731	0.139
R+Book2Vec	0.695	0.729	0.139
WR+Book2Vec	0.693	0.726	0.139
Word Ngram+ RNN	0.691	0.688	0.125
Skip gram + RNN	0.689	0.683	0.125
Typed char ngram+ RNN	0.689	0.702	0.125
Char 3 gram + RNN	0.689	0.688	0.125
Clausal+ RNN	0.689	0.688	0.125
SCS + RNN	0.691	0.688	0.125
WR+Book2Vec+ RNN	0.701	**0.735**	0.129
SCS+WR+RNN	0.675	0.696	**0.123**
All hand-crafted	0.670	0.689	0.148
All hand-crafted+neural	0.667	0.712	0.129

Results on the Ashok et al. Dataset

Features	**Avg Accuracy(%)**
Word Bigram	71.25
Char 3 grams	71.00
Typed mid_word 3-gram	70.25
Writing Density (WR)	68.38
Readability	61.38
Sentic concepts & scores(SCS)	72.38
GoogleNews Word2Vec	69.88
Gutenberg Word2Vec	64.25
Book2Vec	72.38
RNN	55.80
Unigram+Bigram+Trigram	72.75
Book2Vec+SCS	64.75
Book2Vec+WR	66.38
SCS+WR+Typed char ngrams	**73.00**

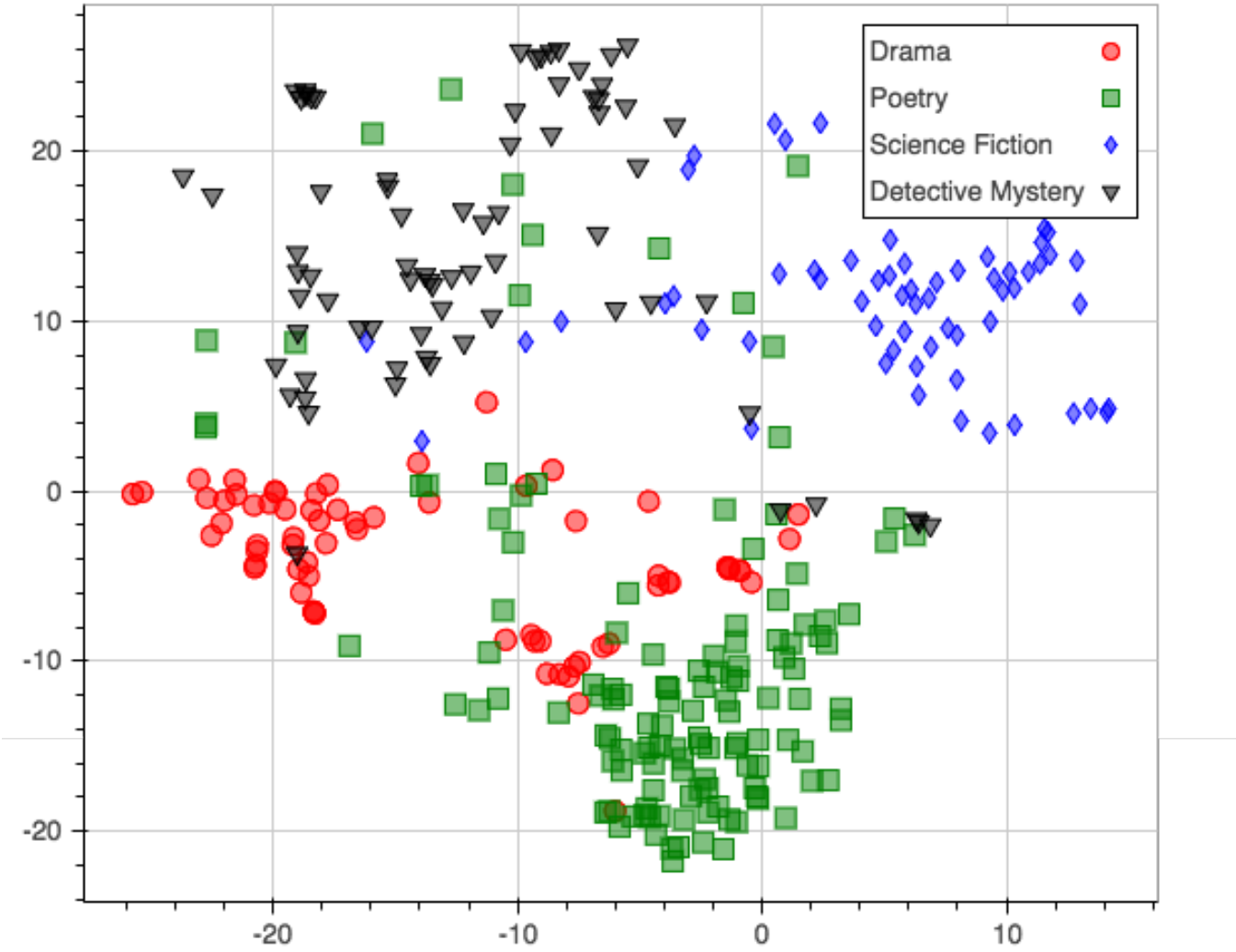
Outline

Discriminative Features

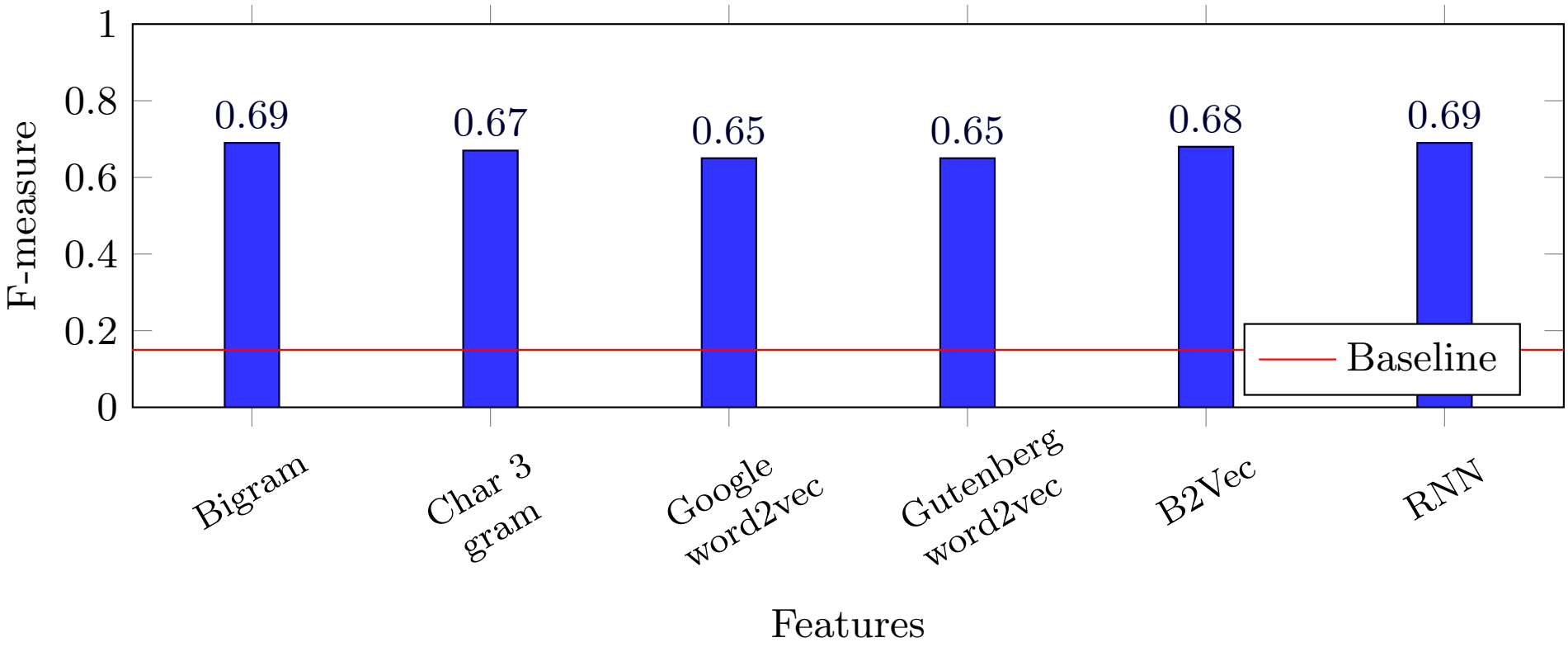
Type	Features
ngrams	. “, . ”, said :, young man, very young man, the young man, boys, . i, father, his father, mother, he said, she said, said NE, princess, lord, colonel, captain, doctor, tour, mr, miss
Sentic concepts	conceive, grieve, zealous, emptiness, bitterness, corpse, hypothesis, irony, theory_of_the, wagon,deep_blue, scarred, screaming, grudging, vigil, vein, beautiful_place, rural, marriage, friendship, cats, 911 avg aptitude, polarity, pleasantness, attention scores
Character and typed character ngrams	mr., mrs., john, thou, amor, pen, his, and,the, ing, n’s,ed, gg’, pt’, d’a, t”, i-t, _, ,”i ,” ”, ” say,” s,” she

- Quotation marks indicate dialogues as an important feature
- Pronouns and titles related to male gender
- First person narration
- Readability and writing density

t-SNE Projection of Book2Vecs

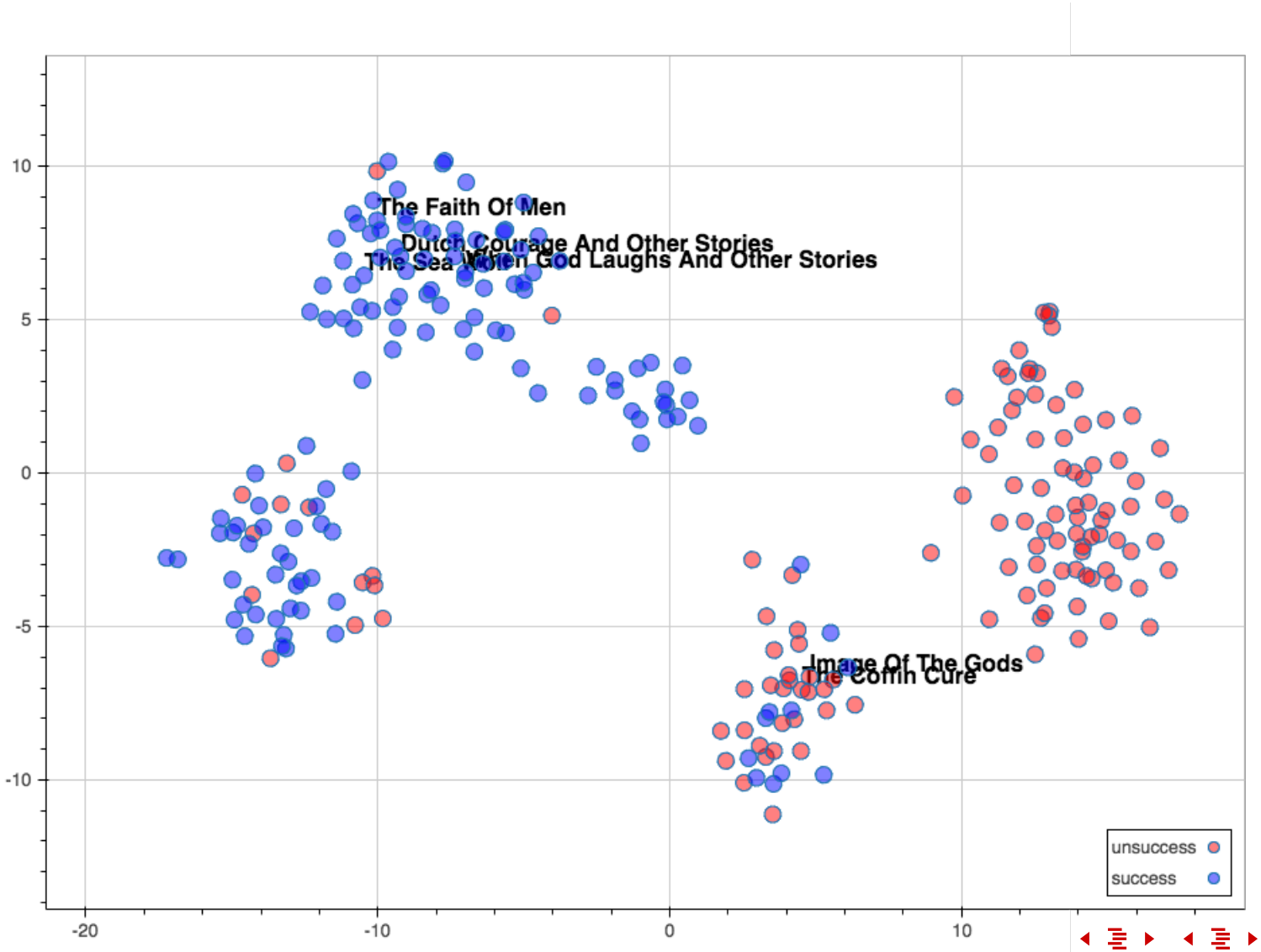


8 way Genre Classification

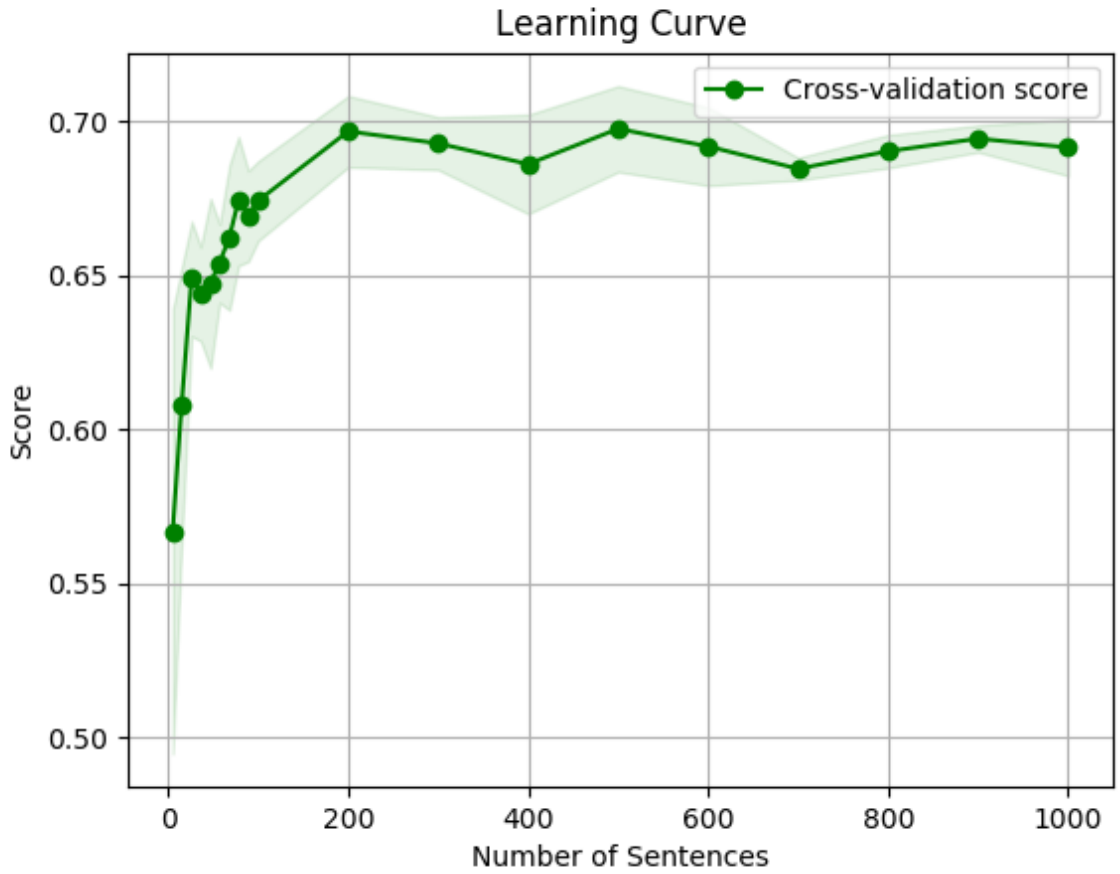


- Poetry and Drama were the most accurately classified genres
- Fiction was the most difficult to classify

t-SNE Projection of Successful and Unsuccessful RNN vectors



How much content is needed for success prediction?



Outline

Conclusion and Future Works

- Multitask approach is better than single task approach
- Hand-crafted features and deep vectors capture complementary information
- Deep vectors capture genre related concepts
- Potential use as editing tool
- Plot and social network analysis

Thank you!

http://ritual.uh.edu/