What’s in a gist? Towards an unsupervised gist representation for few-shot large document classification.

Jaron Mar, Jiamou Liu

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

Motivation

- We live in a world where there are vast amount of texts and resources distributed and accessible online.
- The ***gist*** can be viewed as an abstract concept that represents the quintessential meaning derived from a single or multiple sources of information.
- Identifying the gist contextualises information which facilitates the fast disambiguation and prediction of related concepts.
- Use psychological insights and notions of the gist as inspiration to perform tasks in a more human like way.

Problem Definition: Few-Shot Classification

Typical few-shot problem formulation: Given a set of labelled training examples x_{train} with classes y_{train}. The goal is to create a model that acquires knowledge from the training examples such that the knowledge facilitates the prediction of the test set x_{test} using a few (N) labelled examples of each class in y_{test} and classes in y_{test} are disjoint but related to those in y_{train}.

Our few-shot problem formulation: In a more restrictive but more realistic scenario where given only the test set x_{test} to classify and N examples of each class from y_{test} the goal is to predict the classes of x_{test} i.e. perform N-shot learning in a completely unsupervised manner.

Gist in Psychology: Fuzzy Trace Theory

- ***Fuzzy Trace Theory*** [1] (FTT) is a cognitive theory in psychology which uses a dual-trace model of memory to predict and explain cognitive phenomena, particularly in memory and reasoning.

- A key tenant of FTT posits information can be encoded as gist or verbatim (detailed) representations but humans make decisions based on the gist rather than on exact details.

[1] Brainerd and Reyna, et al, Developmental Review, 1990

Gist Score: Gist Extraction

- A key question in psychology is how can humans manifest or extract the gist of some information. A question we attempt to answer in this work is how can the gist be computationally represented?

Algorithm 1 Gist Extraction Framework, GistEmbedding($T, size$)

Input: Text T, Segmentation size
Output: 1-Dimensional gist embedding of each word

1: $T \leftarrow preprocess(T)$
2: $embeddingArray \leftarrow embed(T)$
3: $segments \leftarrow segment(embeddingSet, size)$
4: $embeddingArray \leftarrow centroid(segments)$
5: $gist \leftarrow reduce1Dim(embeddingArray)$

Gist Score

- In this work we define the ***gist score*** that represents the gist or gists of the input texts.

- Based on the key principles of FTT ***fuzzy-to-verbatim continua*** and ***fuzzy processing preference***, we reduce/map the representation of each word from a word embedding vector to a single real number.

- If you were to judge the similarity between two vectors of or two real numbers, at a glance a human can easily determine the difference between two numbers.

Gist Score: Psychological Motivation

- ***Fuzzy-to-verbatim continua*** states that the that people encode multiple representations at varying levels of precision along the fuzzy-to-verbatim continua.

- ***Fuzzy processing preference*** states that to make decisions they will use the least precise gist representation which allows for faster inferencing.

- Often simple representations and methods are sufficient and lead to surprisingly good results.

Gist Score: Psychological Notions of Gist

- It has been proposed that the gist can be differentiated into two types, the ***global gist*** which captures the meaning of an entire event as a whole and the ***local gist*** which captures the meaning of a more discrete event.

- We therefore define the three types of gist scores, the ***global gist score***, ***local gist score*** and ***combined gist score*** based on the above notion.

Gist Score for N-Shot Learning

Algorithm 2 Gist Score Extraction, GSE($d_1, d_2, ..., d_n, size$)

Input: Documents set $d_{i:n}$, Segmentation size
Output: Gist score for each input document $d_{i:n}$

$text \leftarrow concatenate(d_1, d_2, ..., d_n)$ ▷ join all text together
$gistEmbed \leftarrow gistEmbedding(text, size)$
for i in 1 to n **do**
    $score_{d_i} \leftarrow average(gistEmbed[d_i])$ ▷ segment in embedding related to text d_i
return $score_{d_1}, score_{d_2}, ..., score_{d_n}$

Algorithm 3 Gist Score Similarity N-Shot Instance, GSS($q, S_{C_1}, S_{C_2}, ..., S_{C_n}, size$)

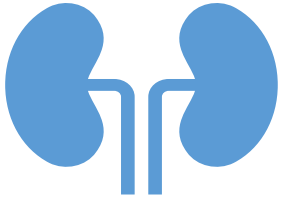
Input: Query document q, Support documents $S_{C_{1:n}}$, Segement size
Output: Probability of q belonging to each class $C_{1:n}$

$global \leftarrow GSE(q, S_{C_1}, S_{C_2}, ..., S_{C_n}, size)$ ▷ consider gist across all support texts
for i in 1 to n **do**
    for $document$ in S_{C_i} **do** ▷ extract pairwise gist similarity
        $local_{q,S_{C_i}} \leftarrow GSE(q, document, size)$
    $gist_{q,S_{C_i}} \leftarrow average(local_{q,S_{C_i}}, global_{q,S_{C_i}})$
return $softmax(-gistq, S_{C_1}, -gistq, S_{C_1}, ..., -gistq, S_{C_n})$ ▷ probability of classes

Gist Score: A Worked Example

Extract: C029 (chapter from a medical textbook)

Most clinicians rate the patient's medical history as having greater diagnostic value than either the physical examination or results of laboratory investigations Rich...



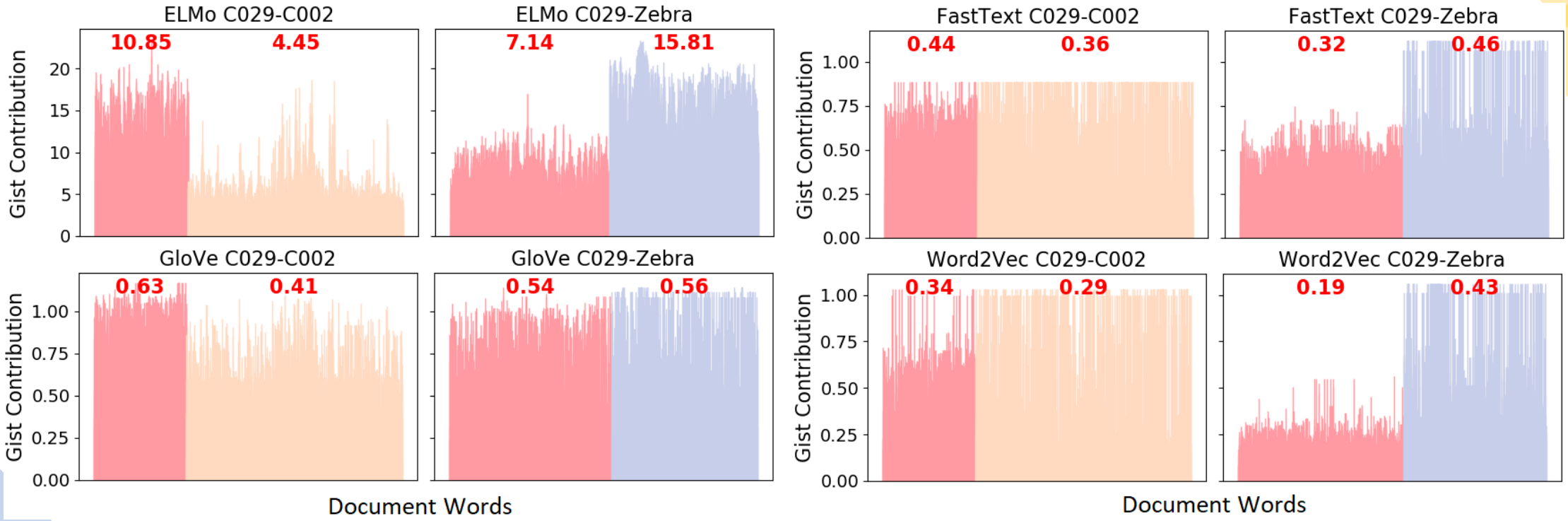
Extract: C002 (chapter from a medical textbook)

Assessment of the peripheral vascular system is done to determine the characteristics of the pulse to ascertain the presence of an arterial bruits and to detect the occurrence of venous inflammation...

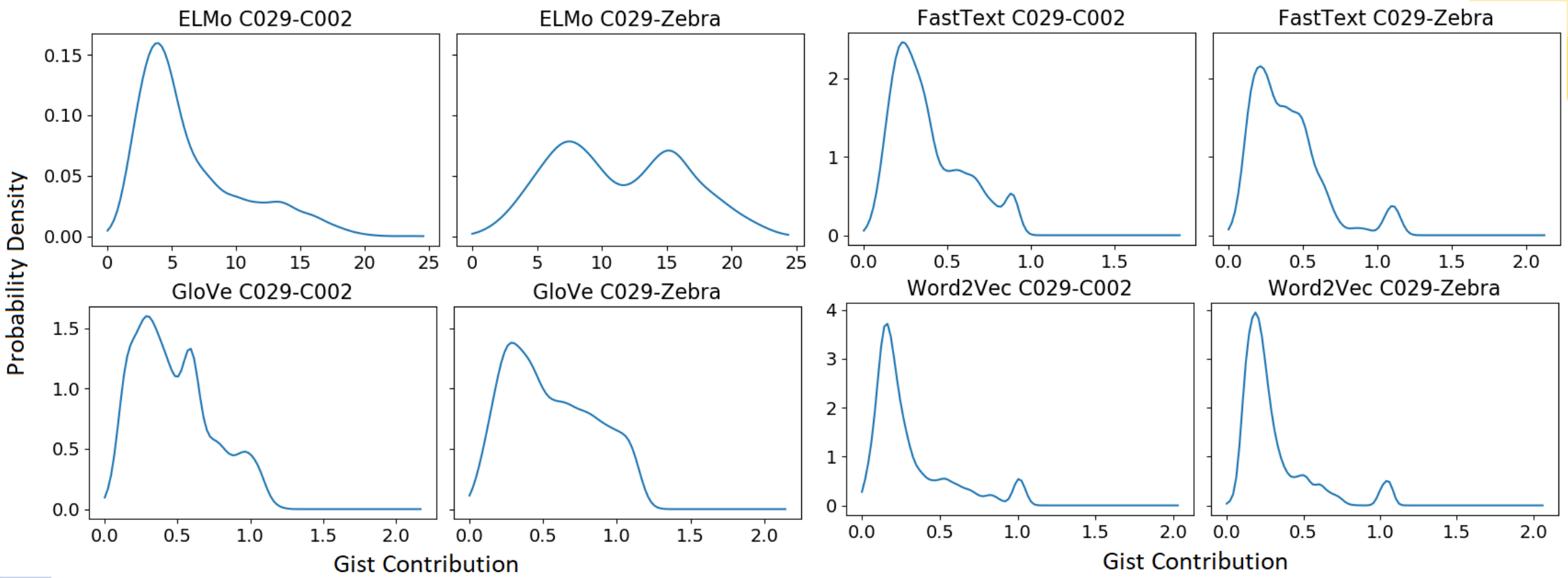
Extract: Zebra (Wikipedia page about Zebras)

Zebras (ZEE-brə, UK also ZEB-rə) are several species of African equids (horse family) united by their distinctive black-and-white striped coats...

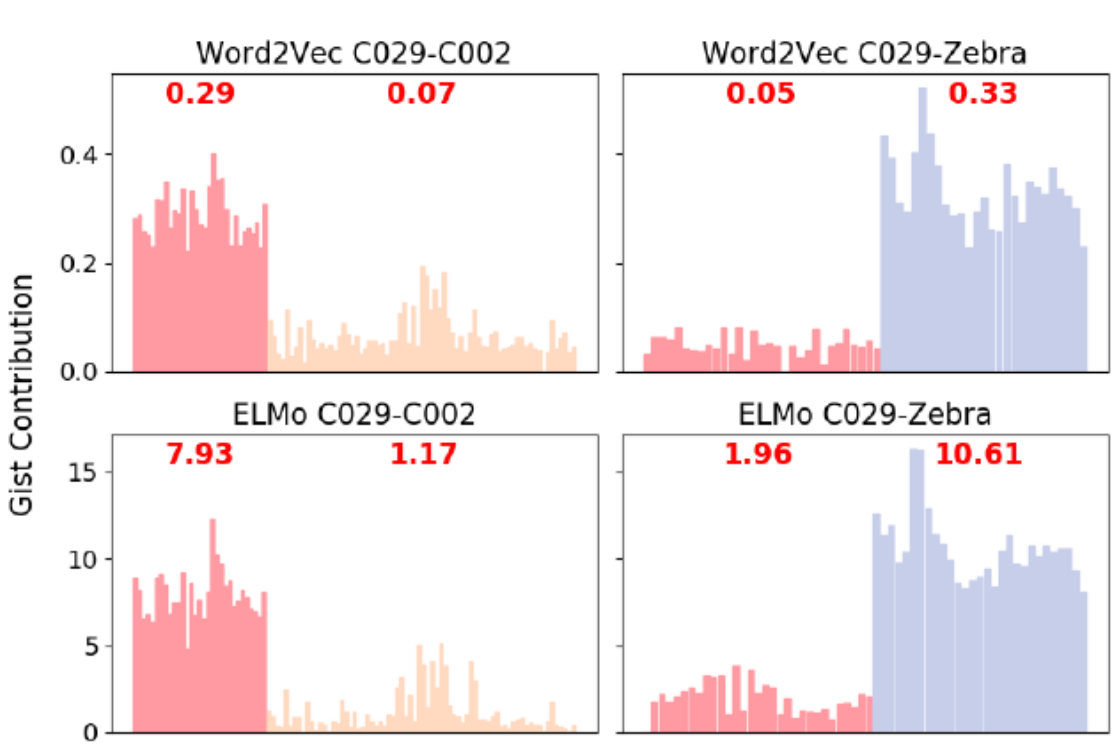
Gist Score: A Visualization

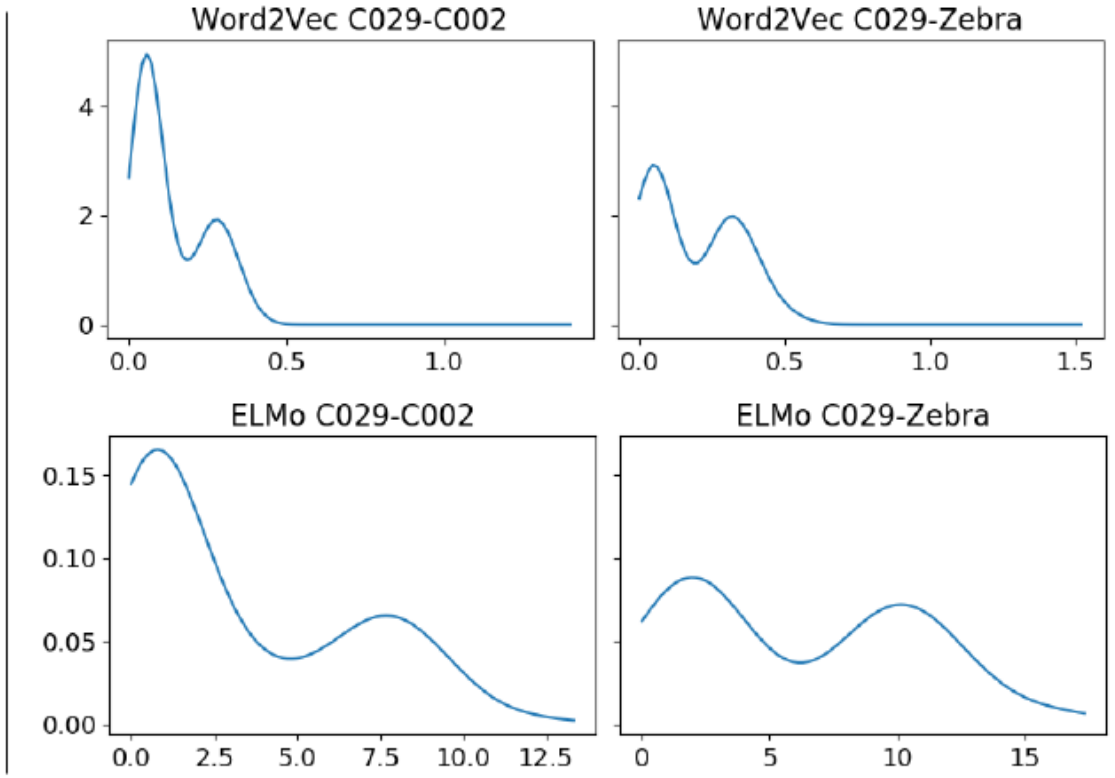


Gist Score: Distributions



Gist Score: Uncovering the Underlying Distribution





Experiments: Datasets

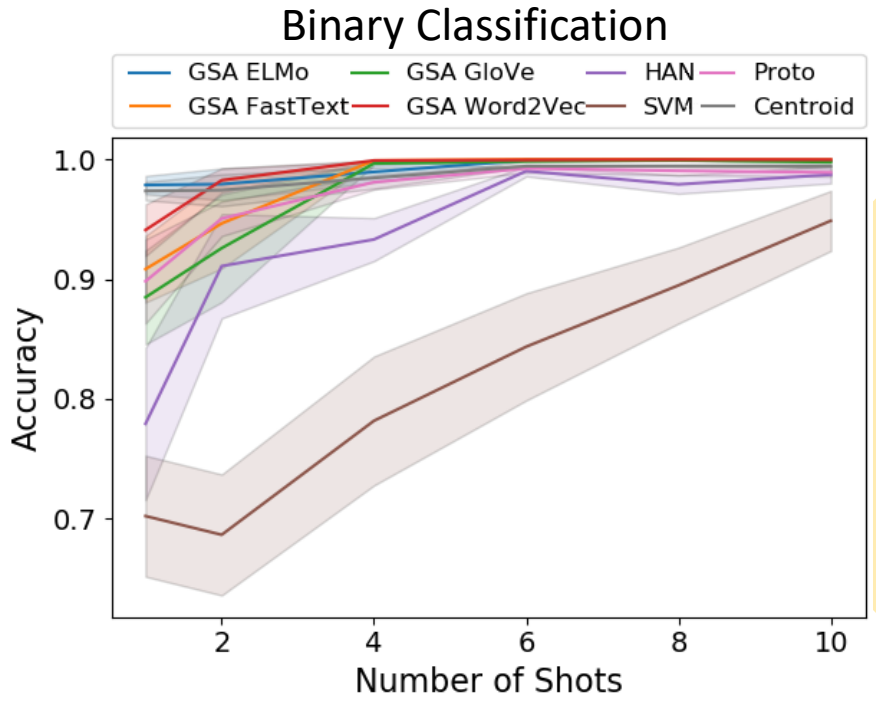
	Documents	Minimum Words	Maximum Words	Average Words	Total Words
Wiki Animal	214	162	15696	3885	831480
Clinical	227	353	14750	2648	601171
Fiction	82	4578	53494	33542	2750472
Physics	33	6325	7956	7155	236127

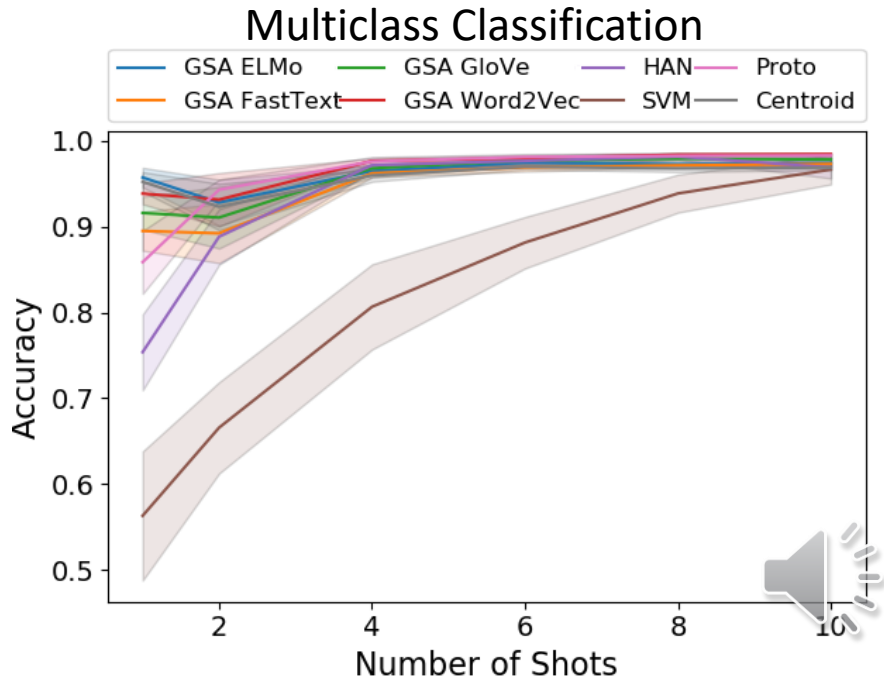
Table 1. Statistics for datasets.

- Compared to existing text classification datasets which generally on average have hundreds of words per document e.g. Yelp, IMDB etc. we apply our experiments to large documents with thousands of words on average.

Experimental Results

- ***Gist Score Association*** (GSA) using ELMo embeddings outperforms all methods in 1-shot learning including:
  - Current N-shot methods (prototypical networks, MAML),
  - Neural network text classifiers (Hierarchical Attention Network)
  - Standard text classifiers (centroids, WCD, SVM using TF-IDF feature vectors).

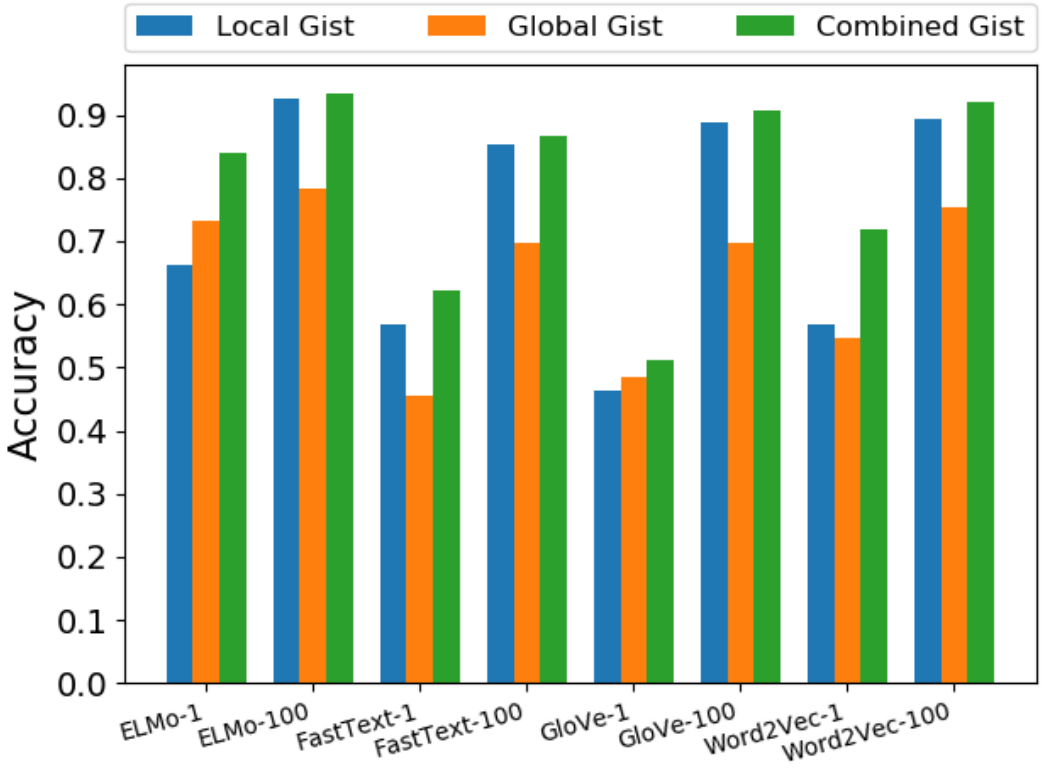




Experimental Results: Gist Scores

- Combination of local and global gist increases the performance in multiclass one-shot learning across all embedding types.

- This suggests that as the number of classes or shots increases evaluating the pairwise similarity, which most current n-shot methods use is not sufficient and global or contextual information about the other documents in the labelled n-shot set can increase accuracy.



Conclusion

- Our gist based method can be used to perform n-shot learning where we don’t have sufficient similar labelled data to apply existing n-shot learning methods and on documents with large numbers of words.

- We take inspiration from psychology to motivate and build a representation of the gist and make predictions can mimic the way humans make decisions.

Thank You For Listening