Key point analysis and explanations for quantitative text analysis

Master Lab Course – Explainable AI for Machine Learning

Daniel Schroter
Yuyin Lang
Hannes Schroter
08.02.2021

Agenda

Argument Keypoint Matching

Prior Research

Our Models

Explainability

Key Point Matching

Topics

"Homeschooling should be banned"

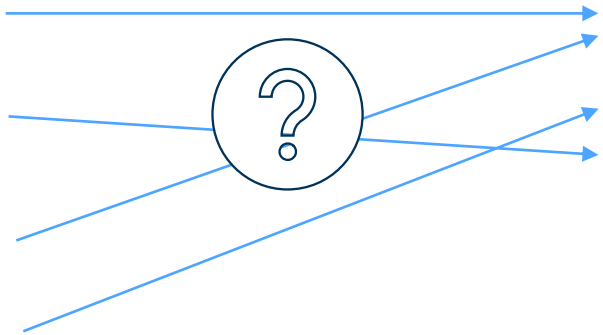
Arguments

Children can not learn to interact with their peers when taught at home.

It is impossible to ensure that homeschooled children are being taught properly.

Homeschooling a child denies them valuable lifeskills, particularly interaction with their own age group and all experiences stemming from this.

Parents are usually not qualified to provide a suitable curriculum for their children. additionally, children are not exposed to the real world.



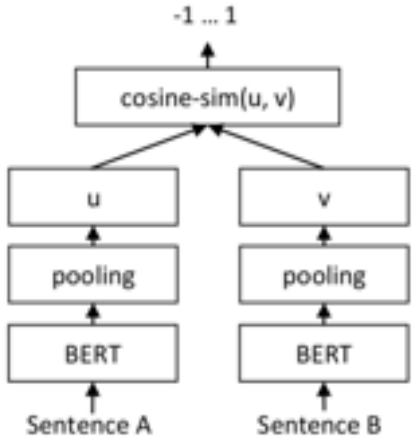
Key Points

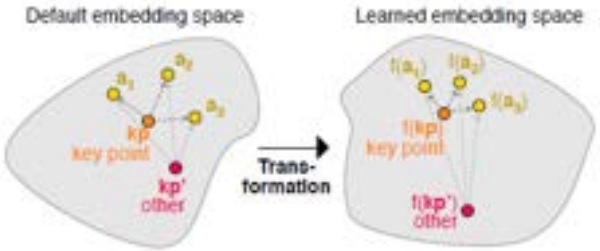
Mainstream schools are essential to develop social skills.

Parents are not qualified as teachers.

Homeschools cannot be regulated/standardized.

Prior Research: SMatchToPageRank





$$yd^2 + (1-y)\max(margin - d, 0)^2$$

Model

Siamese Neural Network

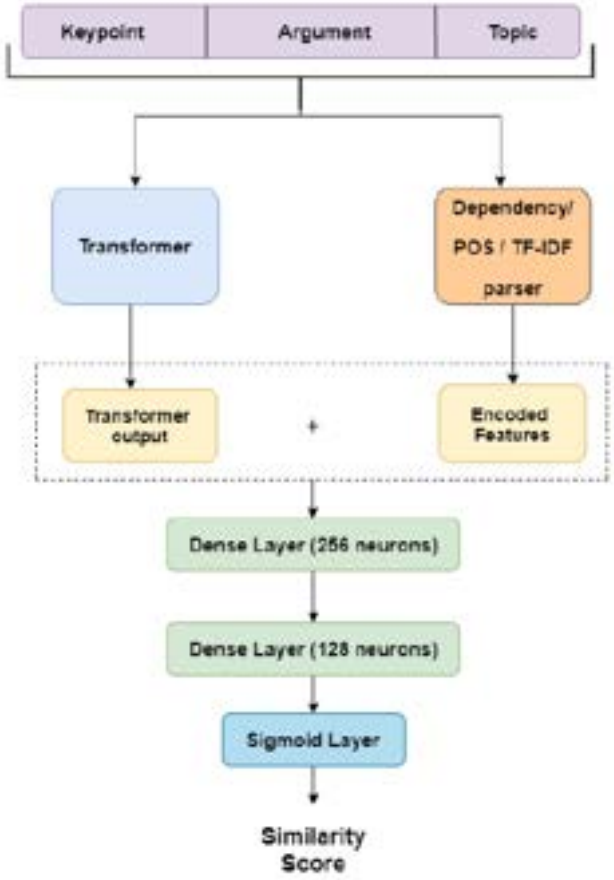
Idea

Learn embeddings where similar pairs of inputs are close to each other and dissimilar inputs are far away. Contrastive Loss.

Strict	Relaxed
0.864	0.950

Result: Mean Average Precision

Prior Research: ENIGMA



Model

Classification

Idea

Use Additional features like PoS, TF-IDF, Dependency to capture similarity, Keypoints, Arguments and Topics are a concatenated input

Strict	Relaxed
0.844	0.931

Result: Mean Average Precision

Models: Overview

Data

Dataset

KPM data
Arguments, Topics, Keypoints

Dataset

STS, Arg30k

Dataset

Web Crawler collecting articles for the topics in the KPA dataset

Pretraining

Unsupervised Learning

Transformer Based Denoising AutoEncoder

Unsupervised Learning

Masked Language Modeling

Unsupervised Learning

Simple Contrastive Learning of Sentence Embeddings

Unsupervised Learning

Semantic Re-Tuning with Contrastive Tension

Model

Supervised Learning

Siamese Neural Network with Contrastive Loss

Supervised Learning

Siamese Neural Network with Contrastive Loss and Additional Features

Supervised Learning

Enigma Model with similarity score and finetuned Embedding

Models: Overview

Data

Dataset

KPM data
Arguments, Topics, Keypoints

Dataset

STS, Arg30k

Dataset

Web Crawler collecting articles for the topics in the KPA dataset

Pretraining

Unsupervised Learning

Transformer Based Denoising AutoEncoder

Unsupervised Learning

Masked Language Modeling

Unsupervised Learning

Simple Contrastive Learning of Sentence Embeddings

Unsupervised Learning

Semantic Re-Tuning with Contrastive Tension

Model

Supervised Learning

Siamese Neural Network with Contrastive Loss

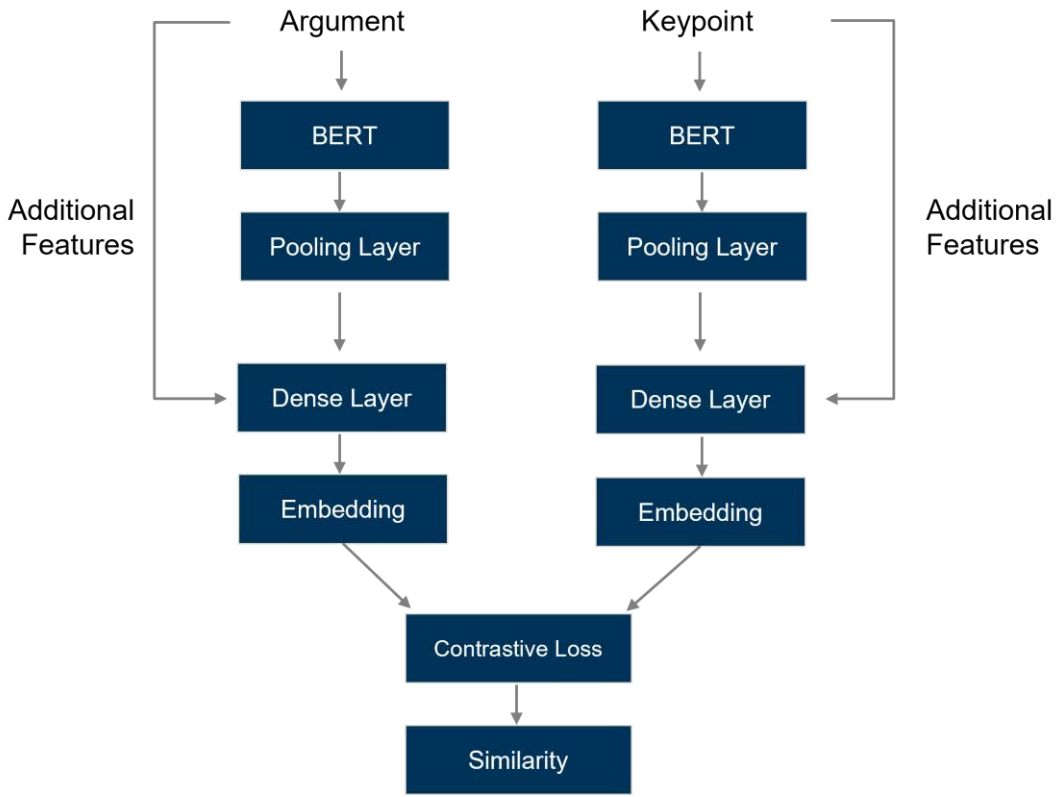
Supervised Learning

Siamese Neural Network with Contrastive Loss and Additional Features

Supervised Learning

Enigma Model with similarity score and and finetuned Embedding

Model 1: PoS SNN



Idea 1

Siamese Neural Networks can approriately learn similarity score between a pair of Inputs

Idea 2

Additional Features like PoS can capture the structural information of a sentence beyond it‘s implicit meaning

PoS SNN
0.912

Enigma
0.844

SNN
0.864

Models: Overview

Data

Dataset

KPM data
Arguments, Topics, Keypoints

Dataset

STS, Arg30k

Dataset

Web Crawler collecting articles for the topics in the KPA dataset

Pretraining

Unsupervised Learning

Transformer Based Denoising AutoEncoder

Unsupervised Learning

Masked Language Modeling

Unsupervised Learning

Simple Contrastive Learning of Sentence Embeddings

Unsupervised Learning

Semantic Re-Tuning with Contrastive Tension

Model

Supervised Learning

Siamese Neural Network with Contrastive Loss

Supervised Learning

Siamese Neural Network with Contrastive Loss and Additional Features

Supervised Learning

Enigma Model with similarity score and and finetuned Embedding

Model 2: TSDAE SNN

TSDAE

Transformer Based Denoising Autoencoder

MLM

Masked Language Modeling

SimCSE

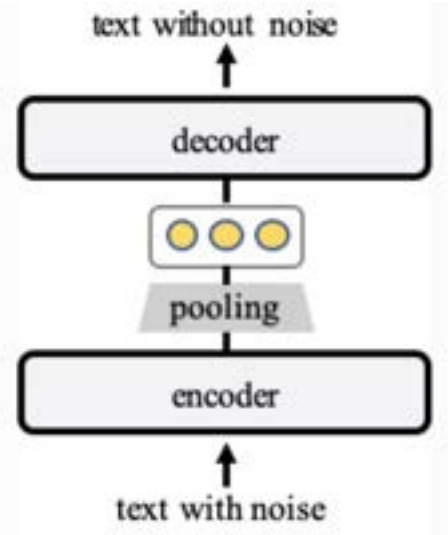
Simple Contrastive Learning of Sentence Embeddings

CT

Semantic Re-Tuning with Contrastive Tension

Transformer Based Denoising Autoencoder

Learn sentence embeddings by reconstructing the original sentence from it‘s damaged version (noise).



Model 2: TSDAE SNN

Dataset

KPM data:
Arguments, Topics, Keypoints

Unsupervised Learning

Transformer Based Denoising AutoEncoder

Supervised Learning

Siamese Neural Network with Contrastive Loss

Explainability

Task

Model predicts the similarity between arguments and keypoints

Explainability Problem of Siamese Neural Networks

No global prediction – No Regression – No Classification

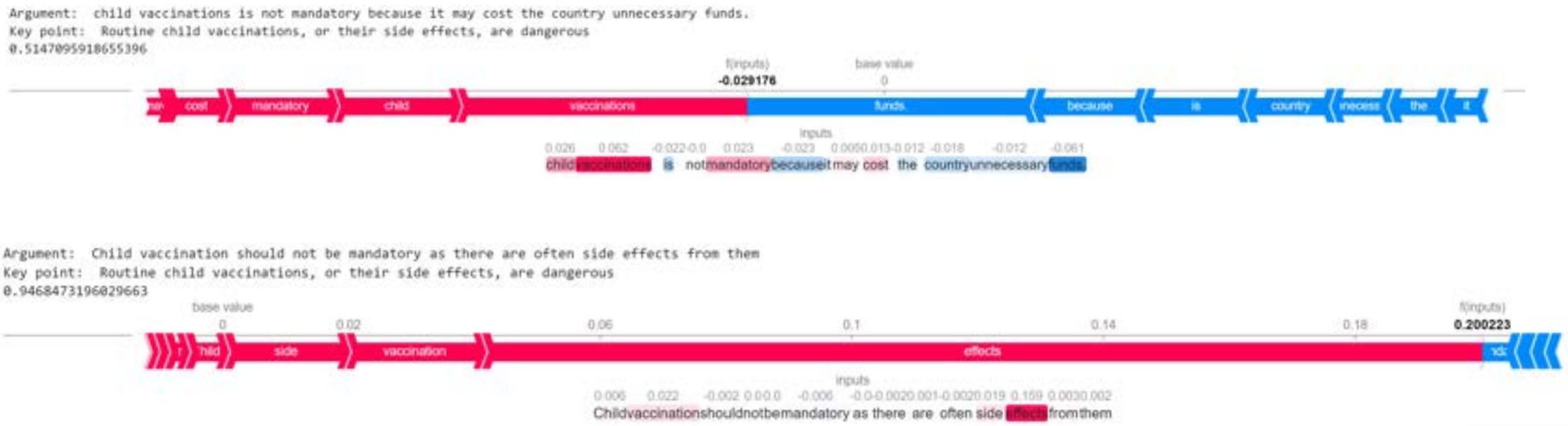
Local Explanation

Why is an argument similar to a keypoint or why are they dissimilar?

LeaveOneOut Explanations

Idea

Marginal contribution of each word to the similarity score



Local interpretable model-agnostic explanations (LIME)

Idea

Approximate a prediction locally by training a simple, interpretable model

Problem

Siamese Neural Networks don‘t perform a regression/classification task since the outcome depends on a pair of inputs

Solution

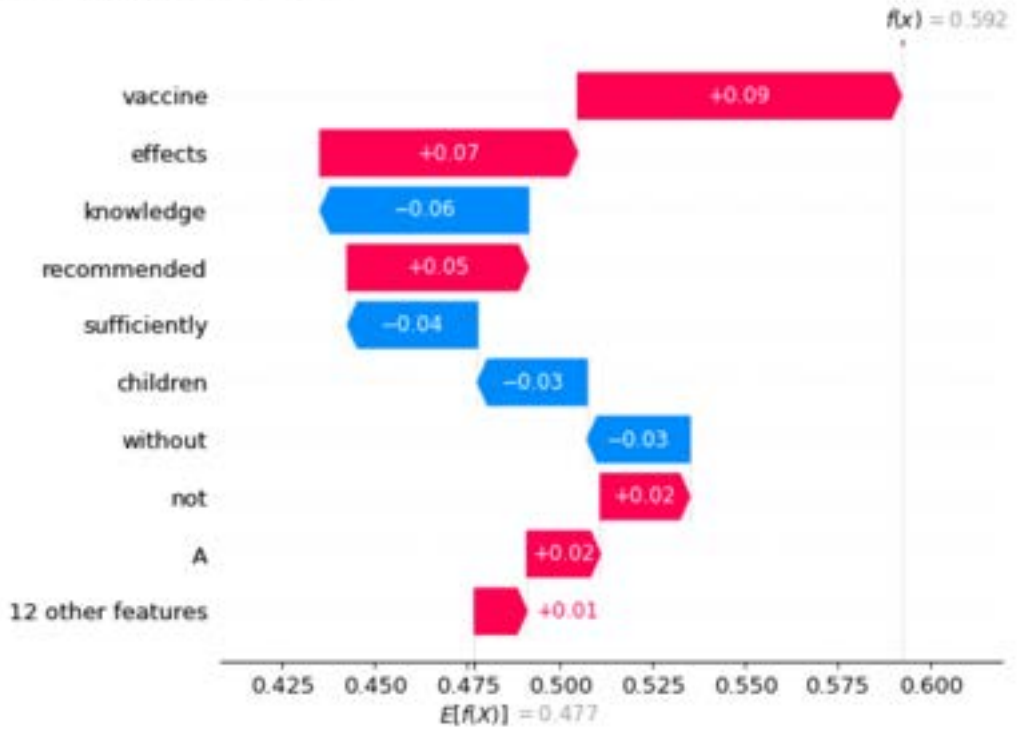
Zoom in to one keypoint and reformulate the prediction as a classification between similar/dissimilar

SHAP – A Local Perspective

Idea

Calculate the feature importance of each token by using Shapley values

```
Topic:  Routine child vaccinations should be mandatory
Argument:  A vaccine that has not been sufficiently tested and without knowledge of side effects is not recommended for children
Key Point:  Routine child vaccinations, or their side effects, are dangerous
Similarity = 0.7575106024742126 True Label:  1
0 = dissimilar, 1 = similar
```



SHAP – A Global Local Perspective

Before

How important are the tokens of an **specific** argument for the similarity to a **specific** keypoint

Now

How important are tokens across **all** arguments for the similarity to a **specific** keypoint

Idea

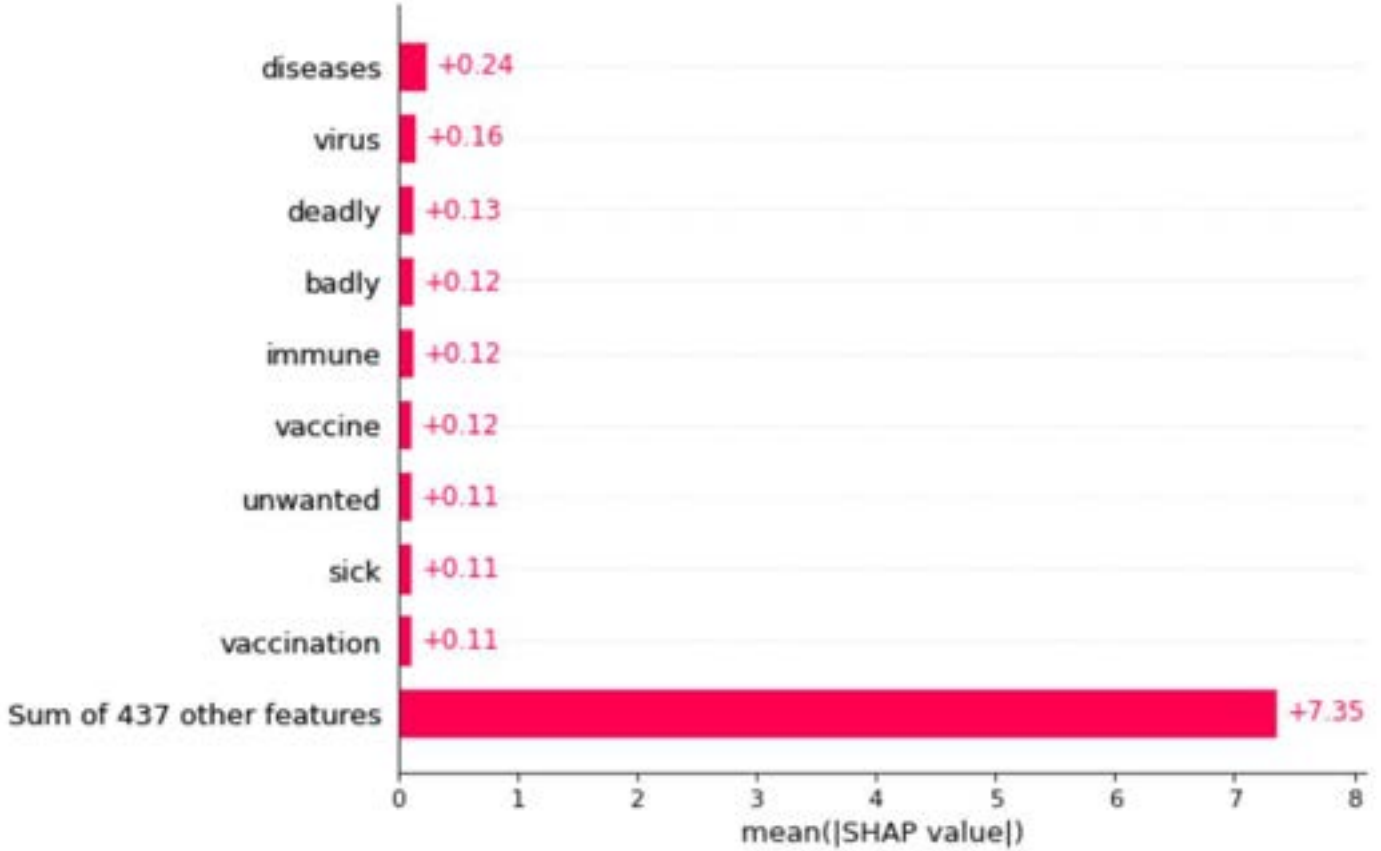
Average the Shapley values across all arguments (global) for a specific keypoint (local).

SHAP – A Global Local Perspective

Idea

Average the Shapley values across all arguments (global) for a specific keypoint.

Topic: Routine child vaccinations should be mandatory
Key Point: Routine child vaccinations, or their side effects, are dangerous



Take Aways

Occams Razor

Simple model has superior performance

Overfitting

Model are likely to overfit

Explainability

Use multiple explainability methods

C ‘people reach their limit when it comes to their quality of life and should be able to end their suffering. this can be done with little or no suffering by assistance and the person is able to say good bye.

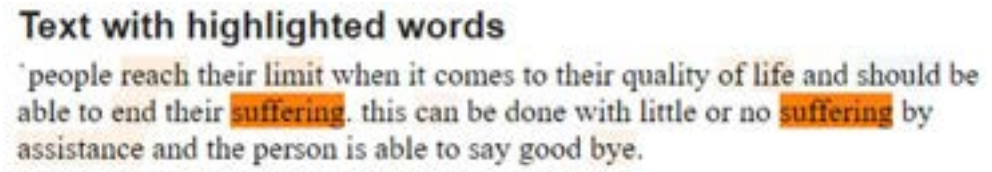
Assisted suicide reduces suffering

LOO

Suffering appears twice

LIME

Strong contribution of suffering



Thank You.

Happy to answer your questions!

Selected Results

Data	Unsupervised Learning	Supervised Learning	Map Strict
KPM	TSDAE	SNN	0.921
KPM		PoS SNN	0.912
KPM	MLM	SNN	0.903
Crawled	TSDAE	SNN	0.900
KPM	SimCSE	SNN	0.899
KPM	TSDAE	Pos SNN	0.895
KPM		ENIGMA + SIM	0.869
KPM		SNN	0.864
KPM		ENIGMA + SNN	0.861
KPM,STS,30K		ENIGMA + SIM	0.858
KPM		ENIGMA	0.844