Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

Evaluation of Semantic Textual Similarity Approaches for Automatic Short Answer Grading(ASAG)

Ramesh Kumar

March 26, 2018

Motivation

- As [2] shows, almost **30%** of British teacher's **time** is spent only for **grading**
- Possibility of **errors** and **unfairness** due to **bias**, **fatigue** or **lack of consistency**

Figure: Computerized Examination [1]

Problem Statement

- Evaluate **one** of the ASAG approach already implemented using **Stanford CoreNLP library[3]**
- **Re-implement** using two different open source libraries; **NLTK** and **Spacy**
- **Evaluate** on the **Texas[6]** and **Mathematics and Robotics course(MRC)** dataset
- **Compare performance** of all three libraries on **Pearson correlation coefficient, root mean square error and runtime**

Natural Language Processing(NLP) Basic Concepts(1)

- Lemmatization
  - Example : “going”, “gone”, “goes” shares same root; “go”
- Named Entity Recognition
  - Example : “Bonn is one of the beautiful city of Germany”
  - Result : [(‘Bonn’, ‘LOCATION’), (‘Germany’, ‘LOCATION’)]
- Part of Speech Tag
  - Example : “I book the flight”
  - Result : [(‘I’, ‘PRP’), (‘book’, ‘VBP’), (‘the’, ‘DT’), (‘flight’, ‘NN’)]

Natural Language Processing(NLP) Basic Concepts(2)

- Parse Trees
  - Dependency Tree

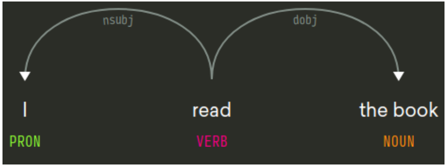

Figure: Dependency tree of sentence = "I read the book"

Approach(1)

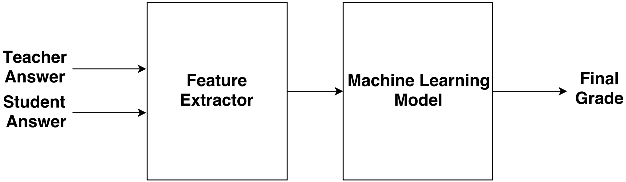

Figure: ASAG pipeline overview

Approach(2)

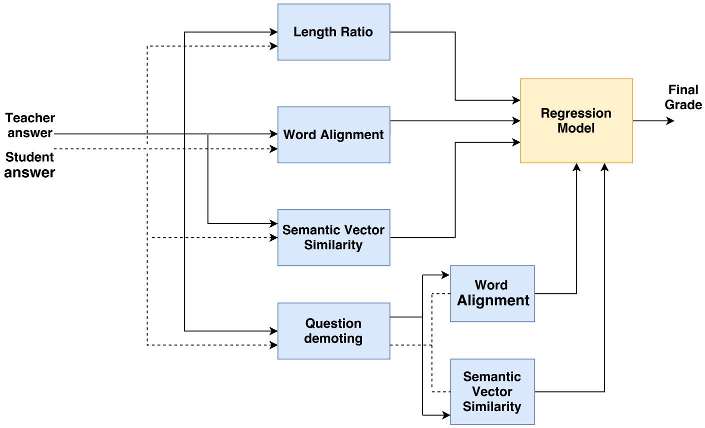

Figure: Detailed view [3]

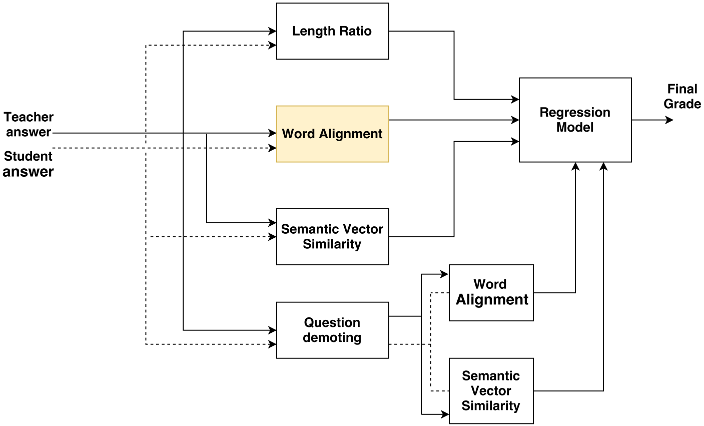

Figure: Word Alignment [3]

Word Alignment

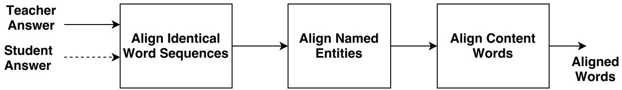

Figure: Word Aligner pipeline [3]

Examples where word aligner works better

Teacher Answer	Student Answer	Aligned Words
A homogeneous transformation matrix is a 4 x 4 matrix that combines a rotation and a translation into a single compact form, thereby representing the transformation between two coordinate frames.	the,translation and rotation that apply to any point	[‘rotation’, ‘rotation’], [‘translation’, ‘translation’]
Rotation matrices are orthogonal, i. e. their inverse is equal to their transpose, and their determinant is equal to 1.	Determinate is 1	[‘determinant’, ‘Determinate’], [‘1’, ‘1’]
Separation of variables, educated guess,(Ansatz) , variation of parameters, numerical,(e.g. Runge - Kutta methods)	1. Solve by separating variables. 2. Euler method. 3. Runge Kutta method.	[‘Runge’,‘Runge’], [‘Kutta’, ‘Kutta’], [‘methods’, ‘method’], [‘Separation’, ‘separating’], [‘variables’, ‘variables’]

Table: Examples taken from MRC data set, where word aligner works better

Examples where word aligner does not works better

Teacher Answer	Student Answer	Aligned Words
A homogeneous transformation matrix is a **4 x 4** matrix that combines a rotation and a translation into a single compact form, thereby representing the transformation between two coordinate frames.	The homogeneous transform matrix is a **4x4** matrix that casts translation and rotation matrices into a single transformation matrix.	['matrix', 'matrix'], ['is', 'is'], ['a', 'a'], ['into', 'into'], ['a', 'a'], ['single', 'single'], ['matrix', 'matrix'], ['that', 'that'], ['transformation', 'transformation'], ['rotation', 'rotation'], ['translation', 'translation'], ['homogeneous', 'homogeneous'], ['transformation', 'transform']
Rotation matrices are orthogonal, i. e. their inverse is equal to their transpose, and their determinant is equal to 1.	$A^T = A^{-1}$	['1', '1']

Table: Examples taken from MRC data set, where word aligner does not works properly

Examples where word aligner does not works better

Teacher Answer	Student Answer	Aligned Words
The characteristic polynomial of a matrix X is given as $\|X - lambdaI\|$ $= 0$ and is used for calculating the eigenvalues of X.	det X - lambda I $=$ 0 is the characteristic polynomial of matrix X. The polynomial is used to determine the eigenvalues and eigenvectors of the matrix.	['X', 'X'], ['-', '-'], ['lambda', 'lambda'], ['I', 'I'], ['0', '0'], ['The', 'the'], ['characteristic', 'characteristic'], ['polynomial', 'polynomial'], ['of', 'of'], ['matrix', 'matrix'], ['X', 'X'], ['is', 'is'], ['used', 'used'], ['the', 'the'], ['eigenvalues', 'eigenvalues'], ['calculating', 'determine']

Table: Example taken from MRC data set, where word aligner does not works properly

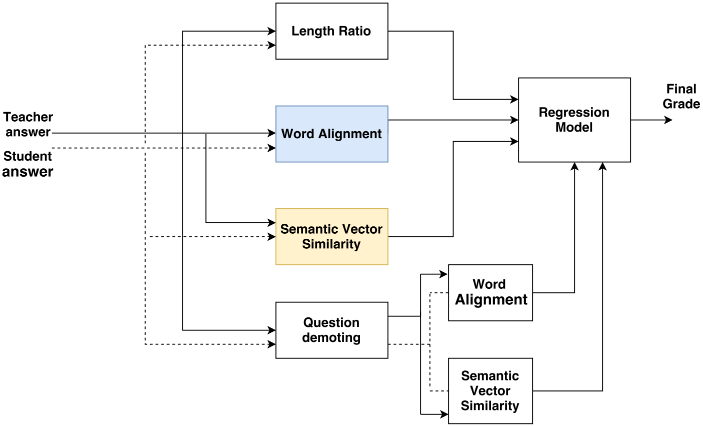

Figure: Semantic Vector Similarity [3]

Word Embeddings

- Enables the machine learning algorithms to **process raw text** and **understand natural language**
- Approach that **represent words** in the **form of vectors** and also **captures their meaning**
- **word2vec trained Model** is used to get the **embeddings** of the words present in student and teacher answer

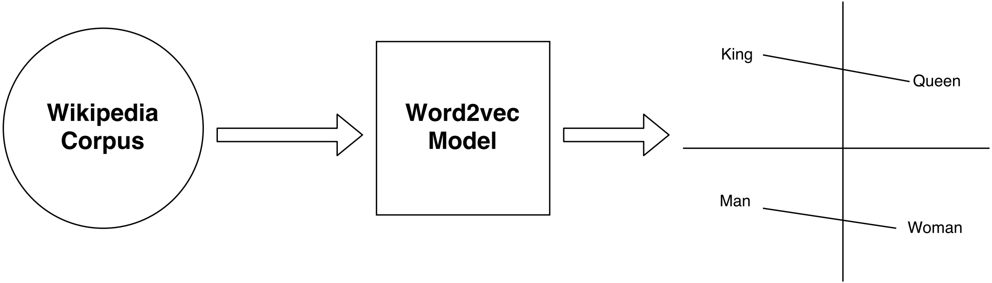

Figure: Word embeddings pipeline [4]

Semantic Vector Similarity(Example 1)

- **Cosine similarity** of two sentences is **0.99**

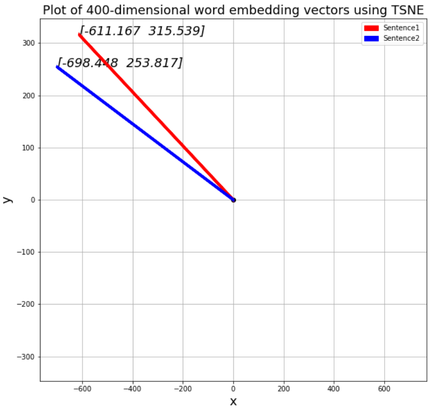

Figure: Embeddings Vector of two sentences; Sentence1 = "Five men are dead from an accident", Sentence2 = "Five people died from a collision".

Example 2

- **Cosine similarity** b/w two answers is **0.86**
- **Grade** assigned by **human grader** is **1**

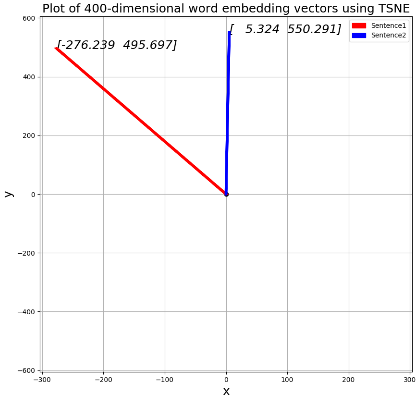

Figure: Embeddings vector of two answers from MRC data set; **Teacher Ans.** ="We can end up in gimbal lock if we use Euler angles, i.e. since rotations are done sequentially, certain axes can get aligned, which reduces the degrees of freedom. Representing rotations with quaternions can be used to overcome this problem." **Student Ans.** = "We can get gimbal lock, which means that we lose one DOF. An alternative would be to use Quaternions to represent rotation"

Example 3

- **Cosine similarity** b/w two answers is **0.6**
- **Grade** assigned by both **human graders** is **0**

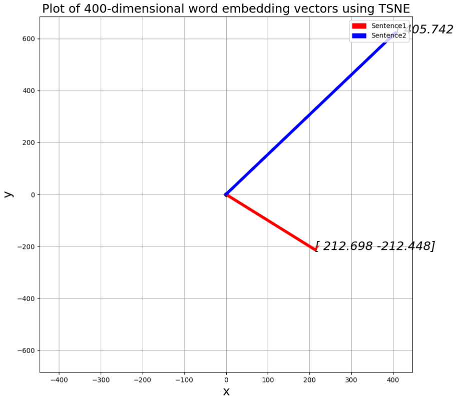

Figure: Embeddings vector of two answers from MRC data set; **Teacher Ans.** =" An orthonormal basis of a vector space is a basis whose vectors are all unit vectors that are orthogonal to each other." **Student Ans.** = "They represent the direction towards the vectors are pointing. "

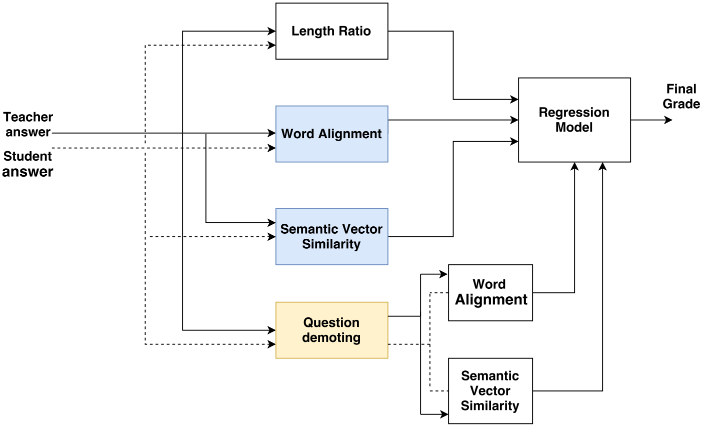

Figure: Question Demoting [3]

Question Demoting

- **Ignoring words** that appear in **questions** from **student** and **teacher answers**
- **Improves** the **performance** of the system

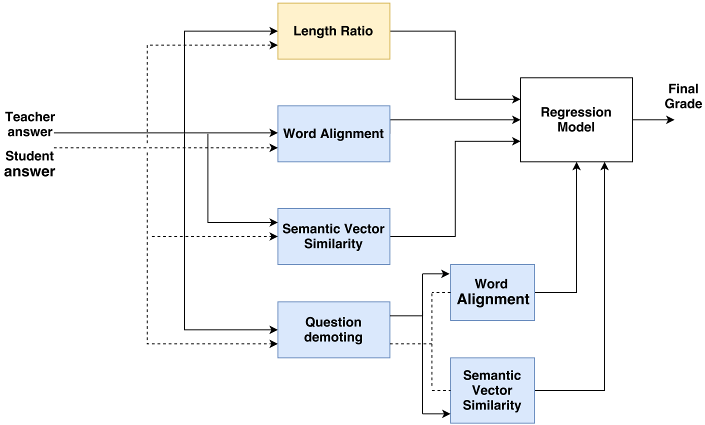

Figure: Length Ratio [3]

Length Ratio

- Idea is to **determine** whether **student answer** contains **enough detail** or not
- Computed as
  - Ratio of number of content words in student answer to the number of content words in teacher answer

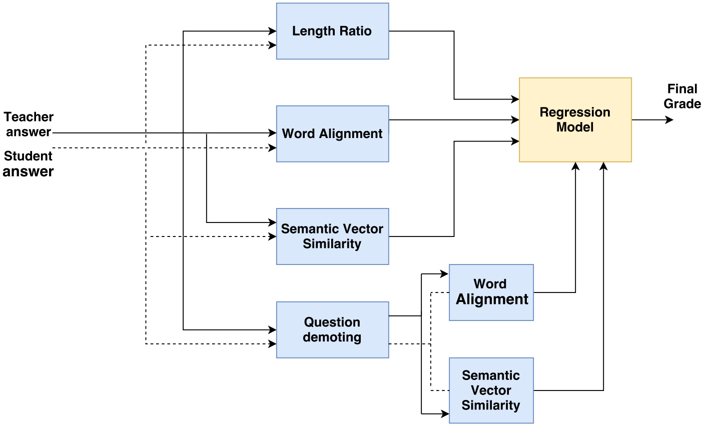

Figure: Regression Model [3]

Datasets

- Texas[1] [6]
  - Created on **data structure course**
  - Consists of **80 questions** and around **2273 answers**
  - Graded by **two human graders**
  - Answers are normalized in the **range** of **0 to 1**
- MRC[2]
  - Created on **mathematics and robotics control course**
  - Consists of **10 questions** and **around 170 answers**
  - Graded by **two human graders**
  - Answers are normalized in the **range** of **0 to 1**

[1] Texas - Open source dataset used for short answer grading

[2] MRC - Mathematics for Robotics and Control

Results(1)

Cases where human grades and system grades are closer or equal

Teacher Answer	Student Answer	Grader 1	Grader 2	Average grade	Stanford	NLTK	Spacy
Rotation matrices are orthogonal, i.e. their inverse is equal to their transpose, and their determinant is equal to 1.	Rotation matrix has determinant of 1. Rotation matrix preserves the norm of the vector. Rotation matrix are orthogonal matrix. Hence the inverse is equal to its transpose. Rotation matrix are linear	1	1	1	1.06	0.99	0.94
Rotation matrices are orthogonal, i.e. their inverse is equal to their transpose, and their determinant is equal to 1	* When rotation matrix and its transpose is multiplied we get an Identity matrix. i.e Rotation matrix is equal to its inverse. * The determinant of the rotation matrix is 1.	1	1	1	0.95	1.07	1.02

Table: Comparison of grades assigned to students answers by human graders and ASAG systems; Stanford coreNLP([3]), NLTK and Spacy. It shows examples where human grades and system grades are closer or equal

Cases where system grades are incorrect

Teacher Answer	Student Answer	Grader 1	Grader 2	Average grade	Stanford	NLTK	Spacy
A homogeneous transformation matrix is a 4 $\times$ 4 matrix that combines a rotation and a translation into a single compact form, thereby representing the transformation between two coordinate frames.	homogeneous transform is [0,0,0,1] which is used in transformation matrix which contains the rotation matrix and translation vector	0	1	0.5	0.79	0.75	0.79
Rotation matrices are orthogonal, i.e. their inverse is equal to their transpose, and their determinant is equal to 1.	No answer	0	0	0	0.05	0.16	0.03

Table: Comparison of grades assigned to students answers by human graders and ASAG systems; Stanford coreNLP([3]), NLTK and Spacy. It shows cases where system grades almost incorrectly

Results(3)

Teacher Answer	Student Answer	Grader 1	Grader 2	Average grade	Stanford	NLTK	Spacy
We can end up in gimbal lock if we use Euler angles, i.e. since rotations are done sequentially, certain axes can get aligned, which reduces the degrees of freedom. Representing rotations with quaternions can be used to overcome this problem.	Gimmberlock arises in euler rotations. Quaternions are used to overcome this.	1	1	1	0.64	0.51	0.61
An eigenvector (of a matrix A) is a vector that doesnt change its direction when multiplied by A; in other words, an eigenvector obeys the relation $Ax = lambdax$, where $lambda$ is a constant called an eigenvalue.	Eigenvalues are the values obtained after solving the characteristic equation of a matrix i.e they are the roota of the characteristic equation. Eigenvectors will contain the maximum variant and important part of the data projected. $\|\|Ax - b\|\| = 0$ Determinant of the characteristic equation is zero	0.25	0.75	0.5	0.7	0.65	0.76

Evaluation of ASAG Systems

Table: Performance on Texas dataset using Stanford CoreNLP[3], NLTK and Spacy

System	Pearson's r	RMSE
Stanford[3]	0.63	0.85
NLTK	0.55	0.20
Spacy	0.55	0.20

Table: Performance on MRC dataset using Stanford CoreNLP[3], NLTK and Spacy

System	Pearson's r	RMSE	Runtime(Average questions graded per min)
Stanford[3]	0.66	0.26	9
NLTK	0.62	0.28	2
Spacy	0.62	0.27	7

What is good?

- Word Aligner **implemented** using **two different libraries**; NLTK and Spacy
- Word Aligner can be used as an **assistant** for **human grader**
- Word Aligner **does not need** any **machine learning**
- Semantic vector similarity **works** in most of cases, but it is **difficult to debug**, when its **not working**

What is not good?

- **Performance reduces**, if answers involve **mathematical equations**
- **Regression model** does **not learn** properly, if **grades** assigned by two human graders has **variation**
- **System assigns grade** based on **current knowledge** rather than **external knowledge**, unlike human grader
- **Length ratio assign grades**, even if answers contain **irrelevant content words**

Word aligner integration with Nbgrader

What is a homogeneous transform?

Teacher Answer

A homogeneous transformation matrix is a 4 x 4 matrix that combines a rotation and a translation into a single compact form, thereby representing the transformation between two coordinate frames

Student Answer

The homogeneous transform matrix is a 4x4 matrix that casts translation and rotation matrices into a single transformation matrix

Figure: Integration of word aligner with Nbgrader in the Jupyter notebook

Future Work(2)

Word embeddings training with different corpus

- Results may **improve**, if word embeddings model is **trained** on **corpus of particular course**

Integration of other state-of-the-art features

- [6] uses **pseudo-relevance feedback technique** to improve the performance of the system, by **integrating** the **correct answers** provided by the students.

References

- http://dailypost.ng/2017/12/30/waec-speaks-2018-exams-use-computer-based-test-platform/
- Mason and Grove-Stephensen. Automated free text marking with paperless school. IN: Proceedings of the 6th CAA Conference, 2002.
- Md Arafat Sultan, Cristobal Salazar and Tamara Sumner, "Fast and Easy Short Answer Grading with High Accuracy", Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016
- https://docs.google.com/presentation/d/1yQWN1CDWLzxGeIAvnGgDsIJr5xmy4dB0VmHFKkLiibo/pub?start=false&loop=false&delayms=3000&slide=id.p
- Sultan, M. A., Bethard, S. & Sumner, T. (2014). Back to Basics for Monolingual Alignment: Exploiting Word Similarity and Contextual Evidence.. TACL, 2, 219-230.
- Michael Mohler, Razvan Bunescu, and Rada Mihalcea. 2011. Learning to grade short answer questions using semantic similarity measures and dependency graph alignments. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1 (HLT '11), Vol. 1. Association for Computational Linguistics, Stroudsburg, PA, USA, 752-762.

Questions?