Methods

NN for AES

Challenges

References

Automated essay scoring: where do you stand and where are we going?



Thomas François



57th ALTE Conference

April 22, 2022

Introduction
0000000

NN for AES

Challenges

References



1 Introduction

- 2 Methods for AES
- 3 Neural Networks for AES
- 4 Opportunities and challenges

000

ntroduction	
000000	

NN for AES

Challenges

References

Plan

Ir

1 Introduction

- 2 Methods for AES
- 3 Neural Networks for AES
- 4 Opportunities and challenges

000

NN for AES

Challenges

References

Automated essay scoring : definition

- First discussion of automated essay scoring [Page, 1966]
- Various names : Automated Essay Grading (AEG); Automated Essay Evaluation (AEE), Automated Writing Evaluation (AWE); or Analytic Writing Assessment (AWA)
- Various goals : assessing content quality or writing skills
 - \longrightarrow we are more concerned by the latter in this talk

Definition

AES is "the process of evaluating and scoring written prose via computer programs" [Shermis and Burstein, 2003]

NN for AES

Challenges

References

Rationales for AES

Save time! [Page, 1966]

 \rightarrow Originally, aimed at reducing teacher's burden, offering good and personalized feedback to L2 learners (very time-consuming for teachers)

Save money!

 \longrightarrow Most visible uses of AES is connected to standardized testing

Computers can do it [Page, 1966]

 \longrightarrow They can sometimes carry out a task better than humans

reproductibility and consistency [Williamson et al., 1999]

 \longrightarrow AES system will always use the same criteria, whatever the context.

22-04-2022

NN for AES

Challenges

References

Rationales for AES (I)

Save time

- Automated assessment can be nearly immediate, 24 hours a day
- Our own current test : we can assess a written production in less than 10 seconds.
- As regards feedback production, we will discuss that later.

NN for AES

Challenges

References

Rationales for AES (II)

Save money

- [McNamara and Lynch, 1997] showed that, for a written task, reliability of assessment increases by 14% when moving from 1 to 2 evaluators and +5% from 2 to 3.
 - \longrightarrow Having 2 or even 3 evaluators is critical, but costly !
- Major test-takers have already adopted AES to cut costs !
- Pearson's Intelligent Essay Assessor scores essays from the Pearson Test of English (PTE) for a decade (no human involved).
- Graduate Record Examination (GRE) or TOEFL also use AES along with a human evaluation.

NN for AES

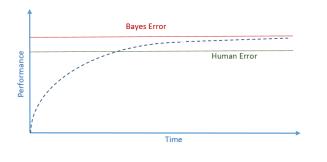
Challenges

References

Rationales for AES (III)

Computers are now as good as human on various tasks

- E.g. AlphaGo is the first IA system to have defeated a human go champion !
- Minimal error on a task can be lower than average human error, especially when humans have trouble to agree.



Sac

NN for AES

Challenges

References

Rationales for AES (IV)

Increase reproductibility and consistency

It is well-known that human evaluators have trouble reaching high agreement

 \longrightarrow severity may vary and systematic bias can occur [Bachman et al., 1995]

 [Williamson et al., 1999] : Human and machine produced holistic scores of candidate performance.

 \longrightarrow The human graders were reconvened to review cases where discrepencies with the machine arose. After that, about half of the score discrepancies were reduced or eliminated.

ction	Methods	NN for AES	Challenges
000	•000000000000	0000000	0000000

Plan

1 Introduction

2 Methods for AES

3 Neural Networks for AES

4 Opportunities and challenges

DAC

Methods ••••••••••

NN for AES

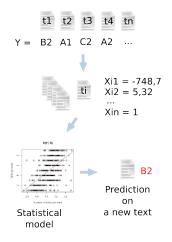
Challenges

References

AES : the Machine Learning Approach

- 1 Gather a corpus of written productions that have been scored in reference with a proficiency scale (e.g. CEFR).
- 2 Define a set of engineered features that are correlated with written proficiency (e.g. lexical sophistication)
- 3 Based on the corpus and these variables, train a statistical model
- 4 Validate the model on unseen data

Supervised approach of AES



22-04-2022 11/48

Methods

NN for AES

Challenges

References

Supervised methods for AES

Linear regression

- First model used [Page, 1966]
- Main advantage : is readily interpretable (useful for high-stake scenarios)
- In some cases, it remains competitive with other ML algorithms [Loukina et al., 2018]
- Not optimal to combine a large set of variable, but can be helped by regularization (L1 = Lasso or L2 = Ridge or L1+L2 = Elastic-net) [Dronen et al., 2015, Somasundaran et al., 2015]

22-04-2022

Methods

NN for AES

Challenges

References

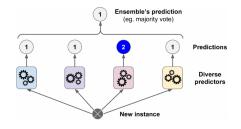
Supervised methods for AES

Ensembles

- Ensembles combine several machine learning algorithms (most commonly trees) to get better performance.
- Examples :

[Larkey, 1998, Chen et al., 2010, Tack et al., 2017,

Vajjala and Rama, 2018]



NN for AES

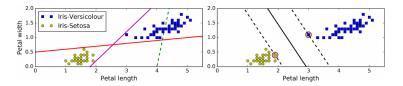
Challenges

References

Supervised methods for AES

Support vector machine

- Considered as the best classification algorithm before the area of neural networks
- Aims at discriminating two classes, while maximizing the margin
- Example from my lab : [Tack et al., 2017]
- Variant : learning to rank with SVM [Yannakoudakis et al., 2011, Chen and He, 2013]



Methods

NN for AES

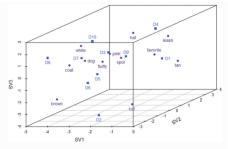
Challenges

References

Unsupervised methods for AES

LSA and other dimensionality reduction techniques

- Essays are projected into a vector space model, whose dimensions are later reduced with SVD
- Essays and target answer or instructional texts are compared based on this semantic space.
- Examples : [Foltz et al., 1999, Lemaire and Dessus, 2001]



Terms and documents in three-dimensional LSA vector space

FIGURE - Source : [Anandarajan et al., 2019]

The features

All these methods generally relies on engineered features.

- surface features : word length, sentence length, number of commas, etc.
- discourse features : essay organization [Burstein et al., 2003], essay development [Attali and Burstein, 2006], coherence [Burstein et al., 2010]
- vocabulary : frequency [Attali and Burstein, 2006], sophistication, collocational usage [Bestgen, 2016]
- grammar errors [Wang et al., 2021], spelling errors [Flor et al., 2019]

Introduction	Methods	NN for AES	Challenges	References
000000	0000000000000	0000000	0000000	00000000

We have conducted a systematic classification of features per languages (unpublished)

Feature	Count	t Language families								
		EN	FR	Ger	Rom	Sin	Jap	Sem	Fin	Sla
Number of words	28	х	x	x	х	x	x	х	x	x
Average word length	18	х	x	x		x			x	
Average sentence length	17	х	x	x		x	x		x	
Number of sentences	15	х	x	x		x	x	х		
Number of characters	10	х		x		x	x			
Number of unique words	8	х		x	х					
Number of paragraphs	6	х	x	x			x			
Number of commas	6	х		x			x			
Number of syllables	5	х	x	x	х					
Average clause length	5	х		x			x			
Number of long sentences	4	х		x	х	x				
Number of conjunctions	4	х	x	x						
Fourth root of the number of words	3	х		x						
Average t-unit length	3	х		x						
Average paragraph length	3	х	x	x						
Number of long words	3	х		x						
Number of short sentences	3	х		x		x				
Percentage of long words	3			x		x	x			
Number of words per sentence	3			x		x				
Number of clauses per sentence	3	х		x						

TABLE - Surface and lexical features

naa

Introduction	Methods	NN for AES	Challenges	References
000000	00000000000000	0000000	0000000	00000000

Feature	Count	Language families										
		EN	FR	Ger	Rom	Sin	Jap	Sem	Fin	Sla		
TTR and variants	14	x	x	x		х	x	x				
n-grams	10	x		x	х	х				х		
Frequency	10	x	x	x		x		x				
Lexical density	8	x		x	х				х	х		
MTLD	6		x	x		х						
HDD	5	x	x	x								
Lexical variation	5	x		x	х				х	х		
Lexical diversity	5 5			x	х	х		x	х	х		
Lexical level		x		x			х					
OOV words	4		x	x								
VOCD	3	x	x									
Yule's K	3	x		x			х					
Nominal ratio	3	x		x								
(Cross-)Entropy	3			x								
Complex words	3			x								

TABLE - Surface and lexical features

18/48

Introduction	Methods	NN for AES	Challenges	References
000000	00000000000000	0000000	0000000	00000000

Feature	Count		EN		FR	I	Ger	Ι	Lar Rom		age f Sin		lies Jap	I	Sem	Fi	n	Sla
Error features				-				· ·										
Number of grammar errors	17		х		х	T	х	1	х		х	1	х	1	х	l x	. 1	х
Number of spelling errors	17		х		х		х		х		х				х			х
Punctuation errors	3		х		х		х											
Part-of-speech features																		
PoS distribution	9		х				х	1			х	1	х	1			- 1	
PoS ratio	8		х				х				х				х	x	.	
PoS ngrams	7		х				х		х		х						1	х
Number of different pos tags	6		х				х											
Morphological features																		
Verb morphology	1																	
(tense, mood, voice, number,																		
person)	6				х		х									X		
Noun morphology (cases)	6 5 4						х									x	.	
Percentage passive sentences							х						х					
Affixes	3						х								х			
Prompt-specific features																		
Similarity between	1																	
essay and prompt	5	1	х	1		1	х	1	х	1	х	1	х	1		1	1	

Sar

Introduction	Methods	NN for AES	Challenges	References
0000000	000000000000000	0000000	0000000	00000000

Feature	Count	EN		FR		Ger		Lang Rom		age fa Sin		ilies Jap		Sem		Fin		Sla
Similarity-based features LSA Shared nouns between sentences Comparison of essay to essays	5 3	x x				x x				x x		x				x		
at each grade Comparison of essay to essays at highest grade	3	x x								x x	1							
Syntax features Depth of parse tree Sentence syntax similarity	8	x x				x		x		x x								x
Semantic features Number of meanings per word Hyper- and hyponymy	43	x x				x x												
Readability features Flesch reading ease Flesch-Kincaid grade level LIX	4 3 3	X X X				x x x		x										

Sar

Methods

NN for AES

Challenges

References

Overview discussion

- English is the richer language as regards amount of features, followed by German, then Chinese.
- Still a lot of work to do for the majority of languages
 - \longrightarrow For French, mostly surface features along with errors detection.
 - \longrightarrow With Deep Learning, the need for engineered features has decreased at the moment !
- Not much syntactic features, nor any based on explicit pedagogical knowledge (cf. yesterday workshop and CEFRLex).

590

Methods

NN for AES

FABRA

Challenges

References

Brief discussion about variables for AES

It is "easy" to develop a large set of features for AES → quick ad for our brand new feature computing system : FABRA

Description Papers Demo Documentation

22-04-2022

22/48

FABRA

French Aggregator-Based Readability Assessment toolkit

FABRA is a readability toolkit based on the aggregation of a large number of readability predictor variables targeting French. The toolkit is implemented as a service-oriented architecture, which obviates the need for installation, and simplifies its integration into other projects.

PAPERS

Main reference

Wilkens, R., Alfter, D., Wang, X., Pinkard, P., Tack, A., Yancey, K., François, T. (2022). FABRA: French Aggregator-Based Readability Assessment toolkit. In Proceedings of the thirteenth international conference on language resources and evaluation (LREC 2022) (submitted).

https://cental.uclouvain.be/fabra/

Introduction
0000000

NN for AES

Challenges

FABRA

variable	avg	median	mode	Q1	Q3	sum	min	max	len	var	stdev	RSD	IQR	80P	90P	dolc
LENwrdSTEM	3.82	3	2	2	5	1059	1	10	277	4.05	2.01	0.53	3	5	6	3
LENwrdLETTERS	4.46	4	2	2	6	1235	1	14	277	8.16	2.86	0.64	4	7	9	5
LENsntWRD	30.78	24	0	10.6	53	277	0	74	9	645.19	25.4	0.83	42.4	54	62.8	38.8
LENwrdSYL	1.54	1	1	1	2	386	1	4	250	0.64	0.8	0.52	1	2	3	2
LEXcovLGNO	0.11	0.13	0	0	0.25	1	0	0.25	9	0.01	0.12	1.04	0.25	0.25	0.25	0.13
LEXcovLGAR	0.44	0.63	0.63	0.09	0.63	4	0	0.75	9	0.09	0.31	0.69	0.54	0.63	0.65	0.03
LEXcovLGST	0.47	0.63	0.63	0.18	0.66	4.25	0	0.75	9	0.09	0.3	0.65	0.49	0.68	0.75	0.13
LEXcovLGCO	0.11	0.12	0	0.1	0.14	1.01	0	0.16	9	0	0.05	0.42	0.04	0.14	0.15	0.03
LEXcovLGAL	0.11	0.12	0	0.1	0.14	1.01	0	0.16	9	0	0.05	0.42	0.04	0.14	0.15	0.03
LEXcovLLNO	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0
LEXcovLLAR	0.5	0.5	0	0	1	4	0	1	8	0.29	0.53	1.07	1	1	1	0.5

Sac

NN for AES

Challenges

References

Brief discussion about variables for AES (II)

However, not every variable should be considered !

- 1 Obviously, variables need to be efficient to score essays (high correlation with scores).
- 2 Variables might be redundant with others (collinearity)
- 3 In addition, variables should be fair \rightarrow no information about the candidates
- 4 Similarly, variables should not be systematically biased
 → E.g. should not capture gender, ethnicity, socioeconomic status, etc.
- 5 Features should have construct validity

 \longrightarrow proportion of commas might be very informative, but is directly the cause of a good writing (risk for cheating the system).

Introduction
0000000

NN for AES

Challenges

References



1 Introduction

2 Methods for AES

3 Neural Networks for AES

4 Opportunities and challenges

590

25/48



The era of Deep Neural Networks

Since 2012, NLP has experienced a genuine revolution with the deep neural networks

Number of papers in main NLP conferences using DL models :

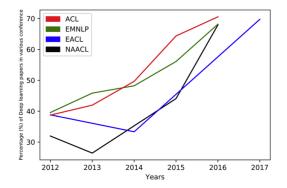


FIGURE - Source :

https://tryolabs.com/blog/2017/12/12/deep-learning-for-nlp-advancements-and-trends-in-2017/

26/48

Introduction	Methods	NN for AES	Challenges	References
000000	00000000000000	0000000	0000000	0000000

What is a neuron?

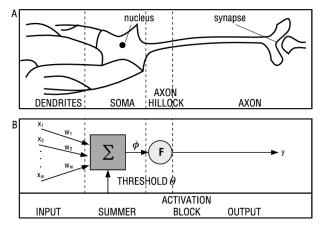


FIGURE - Source [Świetlik et al., 2004]

Depending on the activation function, may be equivalent to linear/logistic regression

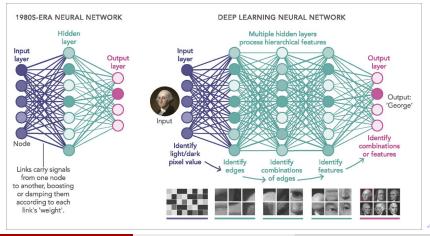
NN for AES

Challenges

References

Deep Learning Principle

Deep Learning = stack various layers of neurons (non-linearity, complex learners)





Deep Learning advantages

1. DL networks can auto-encode text characteristics as variables by themselves

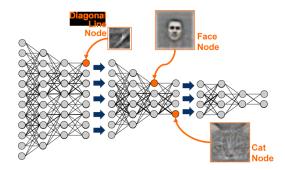


FIGURE - Source : [Glauner, 2015]

Deep Learning advantages

2. Transfer learning

Possibility to train a deep network on a task and to reuse the lower layers (more generic) for problems where there is little data.

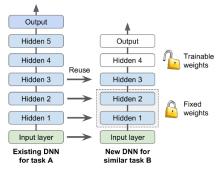


FIGURE - Source : [Géron, 2017, 287]

22-04-2022

Introduction N 0000000 C

NN for AES

Challenges

References

Deep Learning advantages

Among these lower layers, are embeddings, i.e. semantic models aiming at representing the whole language.

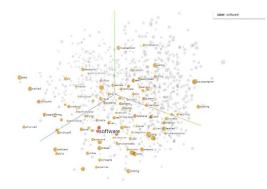


FIGURE - Source: https://medium.com/@aakashchotrani/



Deep Learning for AES

- [Alikaniotis et al., 2016] : one of the 1st approach
 - \longrightarrow design score-specific embeddings
- [Dong and Zhang, 2016] : propose a hierarchical model

 \longrightarrow essays = sequences of sentences, which are sequences of words (two levels of representations).

[Dong et al., 2017] introduce the mechanism of attention to AES

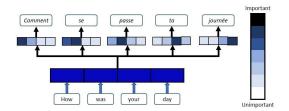


FIGURE - Source: https://blog.floydhub.com/attention-mechanism/

22-04-2022

000000 000000000000 000000 000000 000000	oduction	Methods	NN for AES	Challenges	References
	00000	00000000000000	00000000	0000000	0000000

Plan

1 Introduction

2 Methods for AES

3 Neural Networks for AES

4 Opportunities and challenges

500

NN for AES

Challenges

References

Assessing Speech

- Generally, even more challenging for humans to rate speech production
- There is a rather rich tradition of studies (see [Zechner and Evanini, 2019])

 \longrightarrow Specific challenges related to the automatic recognition of speech (ASR)

 \longrightarrow Set of specific features : pronunciation, fluency, etc.

To my knowlegde, test-makers are not as much advanced for ASE than AWE.

DQC

Deal with cheating

AES systems are prone to be cheated [Klebanov and Madnani, 2021] :

- Overuse of shell language (part of discourse that helps organize the arguments).
 → Good news : humans can handle unnecessary shell language [Beiar et al., 2013].
- Off-topic responses : systems can be trained to detect them, based on similarity between question and answer.
- Plagiarism : test-takers can memorize segments of texts related to known tasks (canned responses).
- Coming issue : artificially generated essays (e.g. GPT-3).

NN for AES

Challenges

References

Example of generated essay

Babel essay generation outputs texts that target known weaknessess of AES systems:http://babel-generator.herokuapp.com/

keywords : automated, essay, scoring

Example of generated text

Marking has not, and likely never will be reclusively incensed. Essay is the most fundamental adherent of human life; some with an arrangement and others at grout. Automatize which enlightenments the exposure lies in the area of philosophy along with the search for literature.

[Cahill et al., 2018] showed that the distribution of some features can be used to distinguish generated essays with genuine ones with 100% accuracy.

Methods

NN for AES

Challenges

References

Model interpretation and fairness

- As DL is pervasive in NLP, interpretation becomes a serious issue, especially for high-stake tests.
- Debate on the use of attention maps as an interpretation tool (see our ACL paper Bibal et al. 2022 for an introduction to the debate)

QUESTION: how much do a pool add to home insurance

ANSWER:

the primary concern of add a pool be the liability exposure if someone not in your household be hurt use the pool you may be hold responsible and/or sue if a judgement be bring against you it can mean 100's thousand in settlement if you live in a typical neighborhood and your yard / pool be fence and secure most insurance company will charge little or no additional dollar for the exposure of the pool if the yard / pool be not fence most company will either require a sign exclusion of coverage for injury arise out of the use of the pool or deny you coverage altogether there be exception to the fencing requirement if the home be in a rural area with no close neighbor

FIGURE - Source : [Santos et al., 2016]

Methods

NN for AES

Challenges

References

Offering feedback

- Maybe due to the success of AES with test-makers, feedback as envisioned by Page –, is not a priority.
- There is work on feedback, but research on effectiveness of automated feedback on writing is inconclusive [Klebanov and Madnani, 2021]
- Feedback heaviliy depends on context (L1 vs. L2 writers, skill level, age, etc.).

NN for AES

Challenges

References

Conclusion

- AES is cost-saving, consistent, and may increase reliability
- Importance of keeping the human in the loop (to detect frauds)
- Most work has been done on English, so other languages should be supported (if relevant)
- Still open challenges for researchers and test-makers.

Methods

NN for AES

Challenges

References

Thank you for your attention

SAC

Introduction	Methods oooooooooooooooo	NN for AES	Challenges 00000000	References

References I

Alikaniotis, D., Yannakoudakis, H., and Rei, M. (2016). Automatic text scoring using neural networks.

In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers), pages 715–725.



Anandarajan, M., Hill, C., and Nolan, T. (2019). Semantic space representation and latent semantic analysis. In *Practical Text Analytics*, pages 77–91. Springer.



Attali, Y. and Burstein, J. (2006).

Automated essay scoring with e-rater® v. 2. The Journal of Technology, Learning and Assessment, 4(3).

Bachman, L., Lynch, B., and Mason, M. (1995). Investigating variability in tasks and rater judgements in a performance test of foreign

language speaking.

Language testing, 12(2) :238–257.

22-04-2022



References II



Bejar, I. I., VanWinkle, W., Madnani, N., Lewis, W., and Steier, M. (2013). Length of textual response as a construct-irrelevant response strategy : The case of shell language.

ETS Research Report Series, 2013(1) :i-39.



Bestgen, Y. (2016).

Using collocational features to improve automated scoring of efl texts.

In Proceedings of the 12th Workshop on Multiword Expressions, pages 84–90.

Burstein, J., Marcu, D., and Knight, K. (2003). Finding the write stuff : Automatic identification of discourse structure in student essays. *IEEE Intelligent Systems*, 18(1) :32–39.

Burstein, J., Tetreault, J., and Andreyev, S. (2010). Using entity-based features to model coherence in student essays. In Human language technologies : The 2010 annual conference of the North American chapter of the Association for Computational Linguistics, pages 681–684.



Cahill, A., Chodorow, M., and Flor, M. (2018). Developing an e-rater advisory to detect babel-generated essays. *Journal of Writing Analytics*, 2 :203–224.

22-04-2022

Introduction	Methods	NN for AES	Challenges
000000	00000000000000	0000000	0000000

References III



Chen, H. and He, B. (2013).

Automated essay scoring by maximizing human-machine agreement. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 1741–1752.



Chen, Y.-Y., Liu, C.-L., Lee, C.-H., Chang, T.-H., et al. (2010). An unsupervised automated essay-scoring system. *IEEE Intelligent systems*, 25(5):61–67.



Dong, F. and Zhang, Y. (2016).

Automatic features for essay scoring-an empirical study. In *Proceedings of EMNLP 2016*, volume 435, pages 1072–1077.



Dong, F., Zhang, Y., and Yang, J. (2017). Attention-based recurrent convolutional neural network for automatic essay scoring. In *Proceedings of CoNLL*, pages 153–162.



Dronen, N., Foltz, P., and Habermehl, K. (2015). Effective sampling for large-scale automated writing evaluation systems. In *Proceedings of the second (2015) ACM conference on learning@ scale*, pages 3–10.

000

References

Introduction	Methods	NN for AES	Challenges	References
000000	0000000000000	0000000	0000000	•••••

References IV



Flor, M., Fried, M., and Rozovskaya, A. (2019).

A benchmark corpus of english misspellings and a minimally-supervised model for spelling correction.

In Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications, pages 76–86.



Foltz, P. W., Laham, D., and Landauer, T. K. (1999).

The intelligent essay assessor : Applications to educational technology. Interactive Multimedia Electronic Journal of Computer-Enhanced Learning, 1(2):939–944.



Géron, A. (2017).

Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems.

" O'Reilly Media, Inc.".



Klebanov, B. B. and Madnani, N. (2021).

Automated essay scoring.

Synthesis Lectures on Human Language Technologies, 14(5) :1–314.

Introduction	Methods	NN for AES	Challenges	References
0000000	00000000000000	0000000	0000000	•••••

References V



Larkey, L. (1998).

Automatic essay grading using text categorization techniques.

In Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval, pages 90–95.



Lemaire, B. and Dessus, P. (2001).

A system to assess the semantic content of student essays. Journal of Educational Computing Research, 24(3):305–320.

Loukina, A., Zechner, K., Bruno, J., and Klebanov, B. B. (2018). Using exemplar responses for training and evaluating automated speech scoring systems. In *Proceedings of the thirteenth workshop on innovative use of NLP for building educational applications*, pages 1–12.



McNamara, T. and Lynch, B. (1997).

A generalisability theory study of ratings and test design in the oral interaction and writing modules.

access : Issues in language test design and delivery, page 197.

500

Introduction
0000000

Methods 000000000000000 NN for AES

Challenges

References

References VI



Page, E. B. (1966).

The imminence of... grading essays by computer. *The Phi Delta Kappan*, 47(5) :238–243.



Santos, C. d., Tan, M., Xiang, B., and Zhou, B. (2016). Attentive pooling networks. *arXiv preprint arXiv :1602.03609*.



Shermis, M. and Burstein, J. (2003).

Automated essay scoring : A cross-disciplinary perspective. Routledge.



Somasundaran, S., Lee, C., Chodorow, M., and Wang, X. (2015). Automated scoring of picture-based story narration.

In Proceedings of the tenth workshop on innovative use of NLP for building educational applications, pages 42–48.



Świetlik, D., Bandurski, T., and Lass, P. (2004). Artificial neural networks in nuclear medicine. *Nuclear Medicine Review*, 7(1) :59–67.

Introduction	Methods 000000000000000	NN for AES	Challenges	References
--------------	----------------------------	------------	------------	------------

References VII



Tack, A., François, T., Roekhaut, S., and Fairon, C. (2017). Human and automated cefr-based grading of short answers. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 169–179.



Vajjala, S. and Rama, T. (2018).

Experiments with universal cefr classification.

In Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications, pages 147–153.



Waldrop, M. M. (2019).

News feature : What are the limits of deep learning? Proceedings of the National Academy of Sciences, 116(4) :1074–1077.



Wang, Y., Wang, Y., Dang, K., Liu, J., and Liu, Z. (2021). A comprehensive survey of grammatical error correction. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 12(5) :1–51.



Williamson, D., Bejar, I., and Hone, A. (1999). 'mental model'comparison of automated and human scoring. *Journal of Educational Measurement*, 36(2) :158–184.

22-04-2022

47/48

Introduction	Methods	NN for AES	Challenges	References
0000000	00000000000000	0000000	0000000	•••••

References VIII

Yannakoudakis, H., Briscoe, T., and Medlock, B. (2011). A new dataset and method for automatically grading esol texts.

In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics : Human Language Technologies-Volume 1*, pages 180–189. Association for Computational Linguistics.

Zechner, K. and Evanini, K. (2019).

Automated speaking assessment : Using language technologies to score spontaneous speech.

Routledge.