CAIR CENTRE FOR ARTIFICIAL INTELLIGENCE RESEARCH

Word-level interpretable scoring mechanism for novel text using Tsetlin Machine

BIMAL BHATTARAI

ICT

University of Agder

Novelty Detection Example

Trained topics: Baseball and Football events
Users Queries: Rugby events (New topic)

Example:
Intent – Who won the rugby match yesterday?
Answer- (can be anything)

Trained topics: Account and Card queries
Users Queries: Loan (New topic)

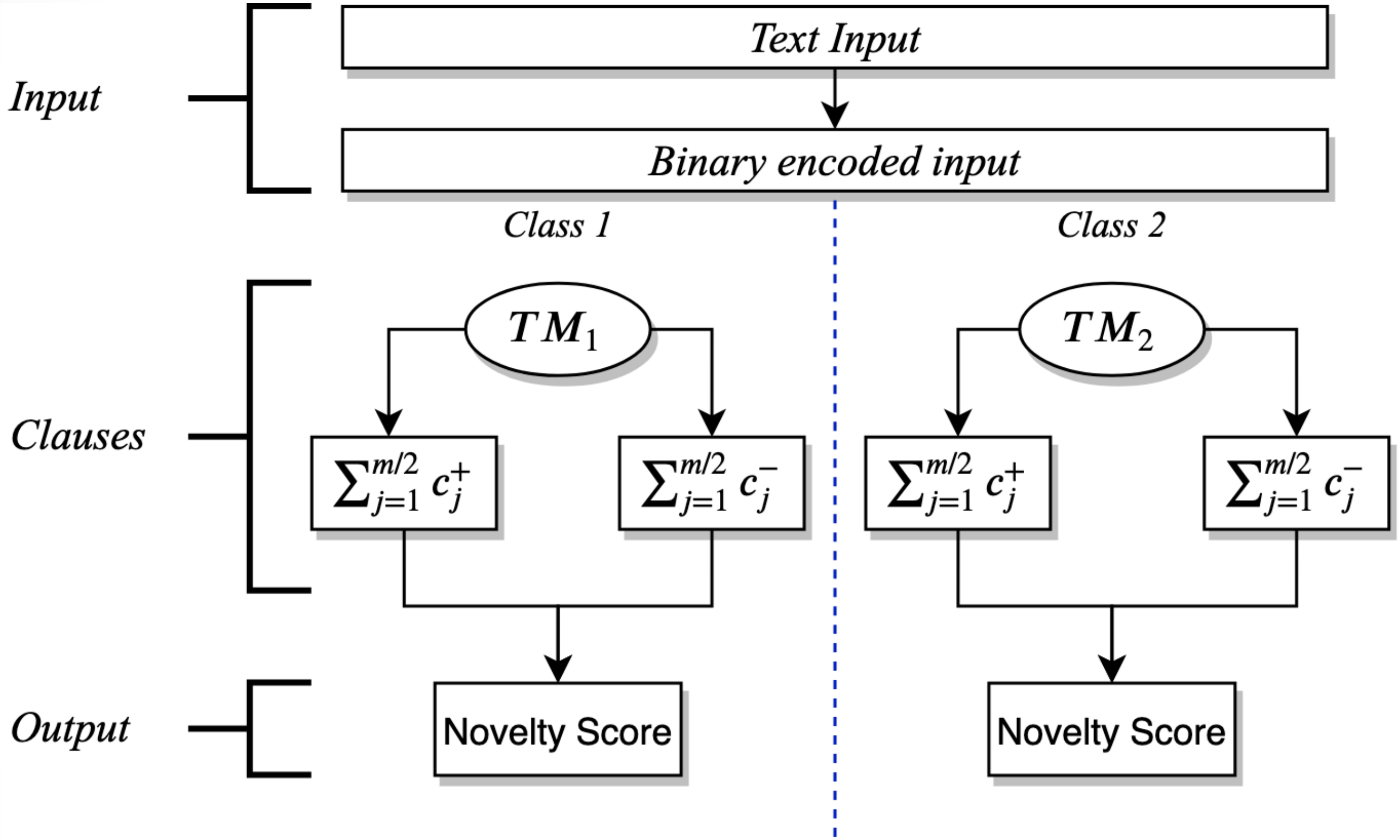
Example:
Intent – I have an account in this bank, How can I get a loan?
Answer- (can be anything)

Novelty Detection

Example:
Intent – Who won the rugby match yesterday?
TM Model- *IF* {(won) *and* (rugby) *and* (match) *and* (yesterday)} *then* New topic
Novelty Score- Very low

Example:
Intent – I have an account in this bank, How can I get a loan?
TM Model- *IF* {(account) *and* (bank) *and* (get) *and* (loan)} *then* New topic
Novelty Score- Very low

TM Framework



Results

Algorithms	20 Newsgroup	Spooky action author	CMU movie	BBC sports	WOS
LOF	52.51 %	50.66 %	48.84 %	47.97 %	55.61 %
Feature Bagging	67.60 %	62.70 %	64.73 %	54.38 %	69.64 %
HBOS	55.03 %	48.55 %	48.57 %	49.53 %	55.09 %
Isolation Forest	52.01 %	48.66%	49.10 %	49.35%	54.70 %
Average KNN	76.35 %	57.76 %	56.21 %	55.54 %	**79.22** %
K-Means clustering	81.00 %	61.30 %	49.20 %	47.70 %	41.31 %
One-class SVM	**83.70** %	43.56 %	51.94 %	83.53 %	36.32 %
TM framework	82.50 %	**63.15**%	**68.15** %	**89.47** %	70.37 %

Table 3: Performance comparison of proposed TM framework with cluster and outlier based novelty detection algorithms.

New Problem Statement

Problems :

- Detecting Novel text is a challenging task, which we have addressed using Tsetlin Machines.
- Understanding what part of sentence is novel.

Goal :

- Rank each word in a sentence for novelty.

Example: Novelty Description

Sentence: The apple that he ate yesterday was good**.**

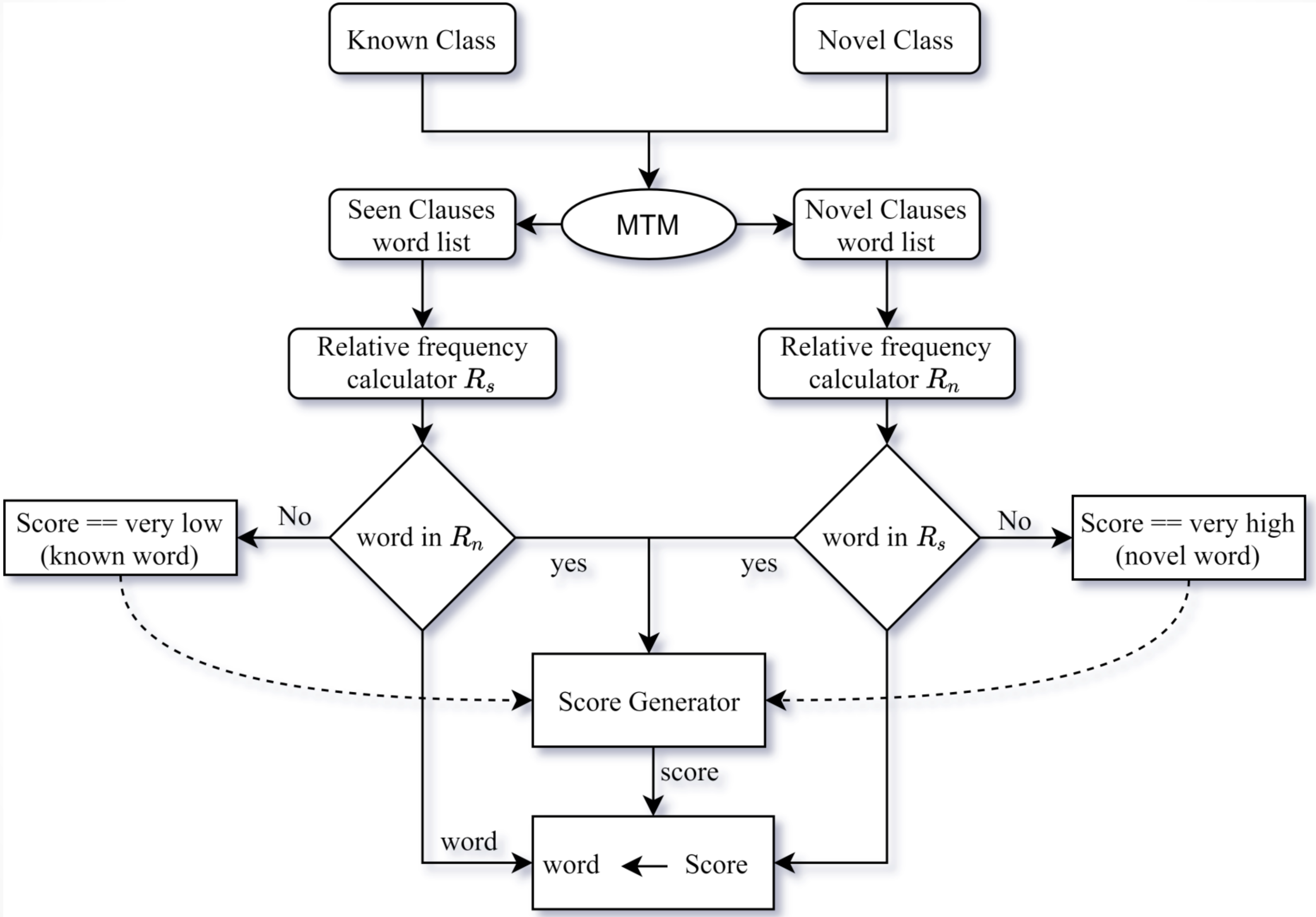
Category- Known

Sentence : The apple shares decreased over this week.

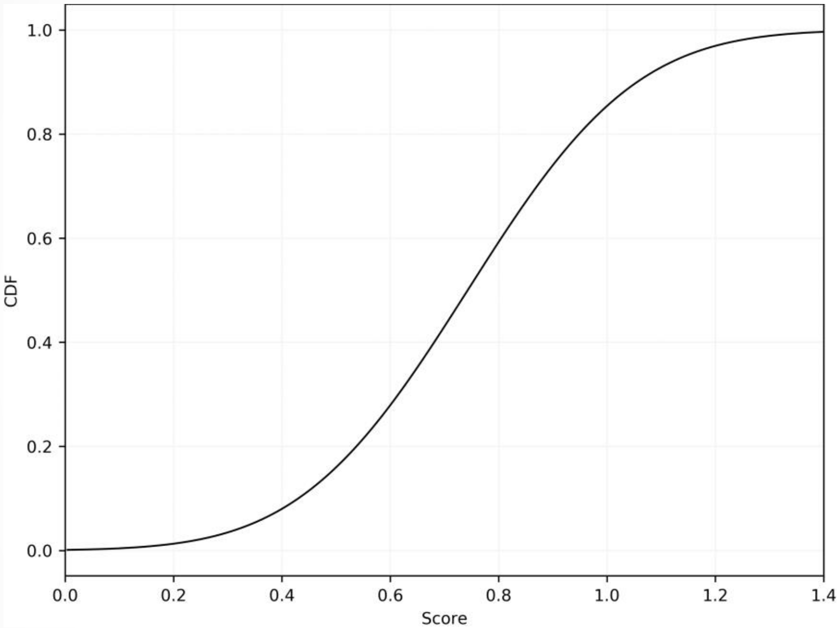
Category- Novel

- Different scores for "apple": low for "apple" fruit, high for "apple" company/phone

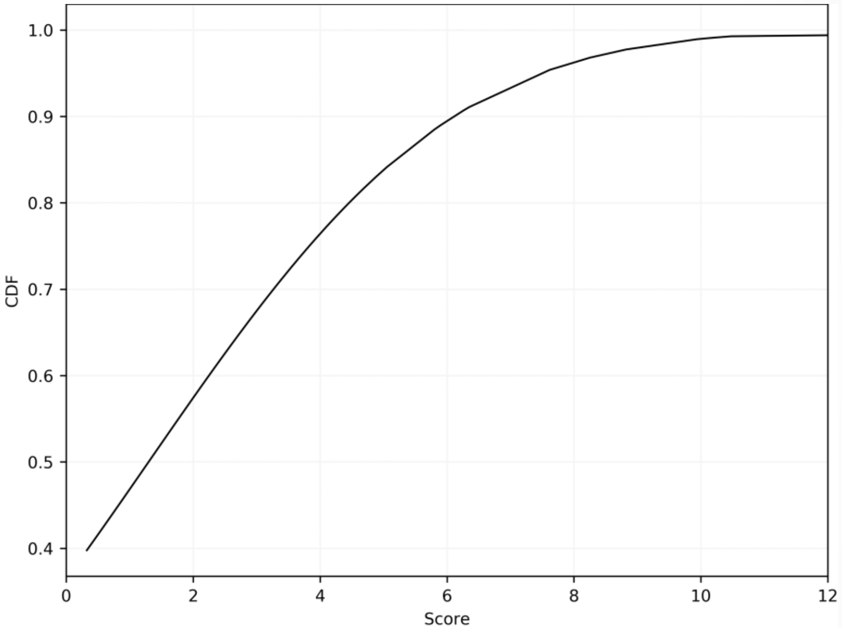
Mechanism


Known Class
Novel Class
MTM
Seen Clauses word list
Novel Clauses word list
Relative frequency calculator R_s
Relative frequency calculator R_n
Score == very low (known word)
No
word in R_n
yes
yes
word in R_s
No
Score == very high (novel word)
Score Generator
score
word
word ← Score

Result



For known



For novel

Context Correlations

	Rugby	Centre	Scrumhalf	Flanker	Flyhalf
Rugby	88.80	2.60	2.80	2.63	2.59
Centre	2.60	278.73	3.54	3.66	6.23
Scrumhalf	2.80	3.54	1718.11	4.05	2.96
Flanker	2.63	3.66	4.05	1558.95	3.55
Flyhalf	2.59	6.23	2.96	3.55	960.91

Confusion Matrix showing the correlation score

References

[1] M. L. Tsetlin, "On behaviour of finite automata in random medium," *Avtomat. i Telemekh*, vol. 22, no. 10, pp. 1345, 1961.
[2] O.-C. Granmo, "The Tsetlin Machine - A Game Theoretic Bandit Driven Approach to Optimal Pattern Recognition with Propositional Logic," Apr. 2018.
[3] O. C. Granmo and B. J. Oommen, "Solving stochastic nonlinear resource allocation problems using a hierarchy of twofold resource allocation automata," *IEEE Trans. Comput.*, vol. 59, no. 4, pp. 545–560, 2010.
[4] K. D. Abeyrathna, O.-C. Granmo, L. Jiao, and M. Goodwin, "The Regression Tsetlin Machine: A Tsetlin Machine for Continuous Output Problems," May 2019.
[5] G. T. Berge, O.-C. Granmo, T. O. Tveit, M. Goodwin, L. Jiao, and B. V. Matheussen, "Using the Tsetlin Machine to Learn Human-Interpretable Rules for High-Accuracy Text Categorization with Medical Applications," Sep. 2018.
[6] Z. Yan *et al.*, "DocChat: An information retrieval approach for chatbot engines using unstructured documents," in *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers*, 2016, vol. 1, pp. 516–525.
[7] A. Ritter, C. Cherry, and W. B. Dolan, "Data-Driven Response Generation in Social Media."
[8] A. Kannan *et al.*, "Smart Reply: Automated Response Suggestion for Email."

CENTRE FOR ARTIFICIAL INTELLIGENCE RESEARCH

Thank you!