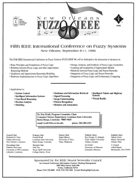
Applications of Tolerance-based Granular Methods

Sheela Ramanna

Applied Computer Science Department
University of Winnipeg
Canada
s.ramanna@uwinnipeg.ca

October 5, 2023

On the Occasion of 80th Birthday of Prof. Skowron



New Orleans FUZZ 96 IEEE

Fifth IEEE International Conference on Fuzzy Systems
New Orleans, September 8-11, 1996

The Fifth IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'96) will be dedicated to the discussion of advances in:

- Basic Principles and Foundations of Fuzzy Logic
- Relations between Fuzzy Logic and other Approximate Reasoning Methods
- Qualitative and Approximate-Reasoning Modeling
- Hardware Implementations of Fuzzy-Logic Algorithms
- Design, Analysis, and Synthesis of Fuzzy-Logic Controllers
- Learning and Acquisition of Approximate Models
- Relations between Fuzzy Logic and Neural Networks
- Integration of Fuzzy Logic and Neural Networks
- Integration of Fuzzy Logic and Evolutionary Computing

- Applications to:
  - System Control
  - Intelligent Information Systems
  - Case-Based Reasoning
  - Decision Analysis
  - Modeling
  - Databases and Information Retrieval
  - Signal Processing
  - Image Understanding
  - Pattern Recognition
  - Robotics and Automation
  - Intelligent Vehicle and Highway Systems
  - Virtual Reality

Dr. Don Kraft, Program Committee Chair
Computer Science Department, Louisiana State University
Baton Rouge, Louisiana 70803-4020

Sponsored by the IEEE Neural Networks Council

Start of a Journey, 1996 –First meetings

- With Prof. Pawlak, New Orleans, USA
- Led to our first visit to Prof. Skowron in 1997 (Warsaw University)

Many more meetings and conferences

- First conference - RSCTC 1998 (Poland)
- 2001 RSTGRC Workshop Japan

Tolerance-based GC

- The notion of tolerance is directly related to the idea of closeness between objects with a tolerable level of difference
- Tolerance Rough Sets
  - Soft granules - overlapping classes via a tolerance relation and approximation operators
- Fuzzy Rough Sets
  - Soft granules - overlapping classes via a fuzzy $\mathcal{T}$-equivalence relation and approximation operators
- Near sets
  - Soft granules - overlapping classes via a tolerance relation

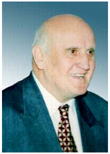
Origins and Motivation

Fuzzy Sets, 1965

- A granule is a clump of objects (points), in the universe of discourse, drawn together by indistinguishability, similarity, proximity, or functionality (Zadeh, 1997)

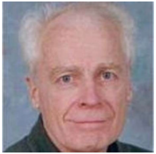
Rough Sets, 1981

- Soft computing methodology based on approximations of sets.



Near Sets, 2007

- Soft computing methodology, derives its origins from rough set theory and descriptive proximity (from sets to families of sets)



Resemblance, Perception and Tolerance

J.H. Poincaré (1894-1902)

- Similarity (resemblance) in sets of sensations
- Perception - Objects in the physical world with characteristics observable to the senses

Jules Henri Poincaré

Jules Henri Poincaré (1854–1912).

Frigyes Riesz

- Proximity or nearness of pairs of sets
- Intl. Congress Mathematicians 1908

Figure 4. Frigyes Riesz, 1880-1956

Tolerance spaces and visual perception

- Zeeman E. C.: 'The Topology of the Brain and Visual Perception', in The Topology of 3-manifolds, Prentice Hall, Englewood, N.J., 1962, pp. 240–248.

Specific Applications - Inspired by the Prof. Skowron and colleagues

- Named Entity Recognition (NER)
  - Skowron, A., Stepaniuk, J.: Tolerance approximation spaces. Fundam. Inf. 27(2,3) (August 1996) 245–253
- Non topic-based classification (Sentiment Analysis)
  - Polkowski, L., Skowron, A., Zytkow, J.: Tolerance Based Rough Sets. In: Lin, T.Y., Wildberger, M. (eds.) Soft Computing: Rough Sets, Fuzzy Logic, Neural Networks, Uncertainty Management, Knowledge Discovery, San Diego, Simulation Councils Inc. (1994) 55–58

Categorizing Linguistic Entities

Started in 2013 with datasets from NELL Corpora (CMU)

- ClueWeb09 (already preprocessed all-pairs data)
- ClueWeb12 (extracted entities from 733,019,372 English web pages)

with Tolerance Rough Sets and Fuzzy Rough Sets

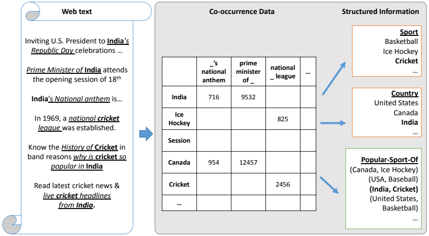


Figure: Common Entities

BioNER in 2021

- Annotating biomedical entities on CORD-19 Open dataset (29,000 articles)
- Extracted, 6,222,196 contextual patterns, 465,250 entities, Co-occurrence Matrix (2.76GB)

with Tolerance Rough Sets

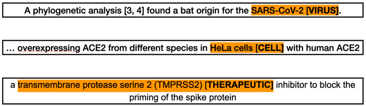


Figure: Biomedical Entities

Linguistic Patterns - common entity types

Fact
Ice Hockey is popular in Canada

Unary Relations
Sport(Ice Hockey), Country(Canada)

Binary Relations
Popular-Sport-Of(Canada, Ice Hockey)

- **contextual extraction patterns**

  e.g. "_ league", "_ and other sports", "_ is popular in _"

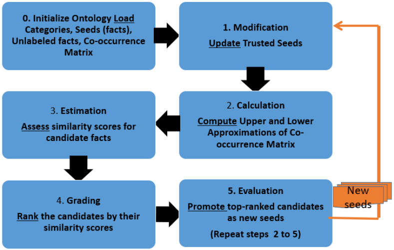
- **co-occurrence statistics**

  e.g. f("Ice Hockey", "_ league") $= n$
  e.g. f("Ice Hockey", "Canada", "_ is popular in _") $= n$

TRS Model for Linguistic Entities

Novel TRS model that permits

- Representation of unary, binary relations and contextual patterns
- Computation of tolerance classes of contextual patterns via co-occurrences
- Calculation Lower and Upper Approximations



Semi-Supervised or weakly supervised algorithms

- Tolerant Pattern Learner:TPL 1.0 and 2.0, Fuzzy Rough Set Pattern Learner (FRL)
- Similarity scoring based on upper and lower approximations
- Benchmarked against CBS and CPL Algorithms (from NELL)

Concept Drift Issues

16 top-ranked instances by TPL 1.0

Iteration 1			Iteration 10		
Phys'Terms	**Soc'politics**	**Vegetables**	**Phys'Terms**	**Soc'politics**	**Vegetables**
inertia	socialism	zucchini	density	humanism	zucchini
acceleration	democracy	spinach	conductivity	pluralism	cabbage
gravity	dictatorship	cucumber	intensity	federalism	kale
buoyancy	monarchy	tomato	viscosity	interna'lism	celery
velocity	independence	broccoli	permeability	nationalism	cauliflower
momentum	justice	lettuce	velocity	rationality	eggplant
magnetism	equality	celery	brightness	liberalism	carrots
resonance	pluralism	cabbage	attenuation	secularism	asparagus
curvature	interna'lism	kale	luminosity	individualism	tomatoes
electromagnet.	federalism	cauliflower	reflectance	democracy	spinach
density	secularism	asparagus	sensitivity	environ'ism	squash
elasticity	liberalism	carrots	amplitude	morality	cucumber
surface tension	hegemony	tomatoes	thickness	pragmatism	melon
polarization	self-determ.	avocado	frequency	spirituality	chicken
vibration	unification	eggplant	water cont.	regionalism	tofu
entropy	capitalism	carrot	salinity	subjectivity	shrimp

TRS Model - Unary Relations

Noun-Context Tolerance Model*

$\mathcal{A} = (\mathcal{C}, \mathcal{N}, I, \omega, \nu)$

- $\mathcal{N}$ and $\mathcal{C}$ are the universes
- $I = I_\theta(c_i) = \{c_j : \omega(N(c_i), N(c_j)) \geq \theta\}$ describes tolerance classes for contexts
- $\omega(A, B) = \frac{2|A \cap B|}{|A|+|B|}$ is the overlap index
- $\nu(X, Y) = \frac{|X \cap Y|}{|X|}$ measures degree of inclusion

- $\mathcal{L}_{\mathcal{A}}(n_i) = \{c_j \in \mathcal{C} : \nu(I_\theta(c_j), C(n_i)) = 1\}$
- $\mathcal{U}_{\mathcal{A}}(n_i) = \{c_j \in \mathcal{C} : \nu(I_\theta(c_j), C(n_i)) > 0\}$

- *C. Sengoz and S. Ramanna. A Semi-supervised Learning Algorithm for Web Information Extraction with Tolerance Rough Sets. Proc.of Active Media Technology, 2014 Web Intelligence Congress, LNCS 8610, 1–10, 2014
- Skowron, A., Stepaniuk, J.: Tolerance approximation spaces. Fundam. Inf. 27(2,3), 1996
- S. Kawasaki, N.B. Nguyen, and T. Ho. Hierarchical Document Clustering Based on Tolerance Rough Set Model, *Proc. of the 4th European Conf. on Principles of Data Mining and Knowledge Discovery*, 458–463, 2000

Similarity Calculation

$$micro(n_i, n_j) = \omega(C(n_i), C(n_j))\alpha + \omega(\mathcal{U}_{\mathcal{A}}(n_i), C(n_j))\beta + \omega(\mathcal{L}_{\mathcal{A}}(n_i), C(n_j))\gamma$$

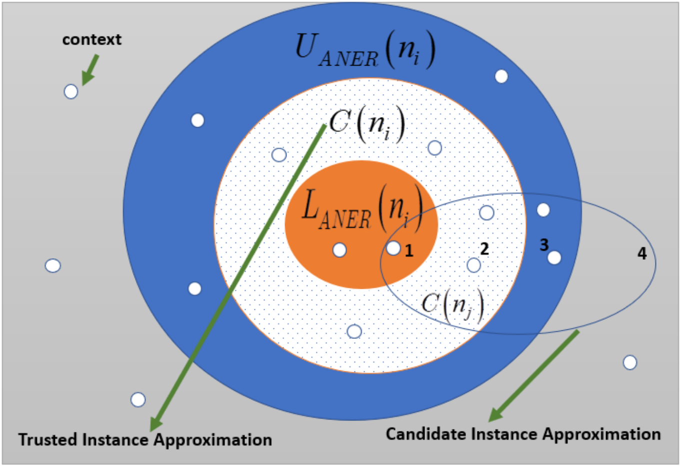


Figure: Zones of Approximation induced by trusted entity (noun (n_i)) and candidate entity (noun (n_j)) and contexts $C(n_i)$ and $C(n_j)$ respectively for a certain category.

TPL Algorithm - Annotating Nouns

Algorithm 1: Tolerant Pattern Learner for Entities

Input : An ontology O defining categories and a small set of seed examples; a large corpus U

Output: Trusted instances for each category

for $r = 1 \rightarrow \infty$ **do**
    for each category *cat* **do**
        for each new trusted noun phrase n_i of *cat* **do**
            Calculate the approximations $\mathcal{U}_{\mathcal{A}}(n_i)$ and $\mathcal{L}_{\mathcal{A}}(n_i)$;
            for each candidate noun phrase n_j **do**
                Calculate $micro(n_i, n_j)$;
        for each candidate noun phrase n_j **do**
            $macro_{cat}(n_j) = \sum_{\forall n_i \in cat} micro(n_i, n_j)$;
        Rank instances by $macro_{cat}/|cat|$;
        Promote top instances as trusted;

TRS Model - Binary Relations

- $\mathcal{R} = \{r_1, r_2, ..., r_Q\}$ is the universe of relational (binary) contexts.
- $\mathcal{T} = \{t_{ij} = (n_i, n_j) \in \mathcal{N}^2 : \exists r_k \in \mathcal{R} \mid f_{\mathcal{T}}(t_{ij}, r_k) > 0\}$ is the universe of co-occurring noun phrase pairs (i.e. tuples)

Then, we define the cross-mapping functions :

- $R : \mathcal{T} \to \mathbb{P}(\mathcal{R})$ maps each noun phrase pair to its set of co-occurring relational contexts: $R(t_{ij}) = \{r_k : f_{\mathcal{T}}(t_{ij}, r_k) > 0\}$
- $T : \mathcal{R} \to \mathbb{P}(\mathcal{T})$ maps each relational context to its set of co-occurring noun phrase pairs: $T(r_k) = \{t_{ij} : f_{\mathcal{T}}(t_{ij}, r_k) > 0\}$

Approximation Operators

Relation-Context Tolerance Model*

$\mathcal{A} = (\mathcal{R}, \mathcal{T}, I, \omega, \nu)$

- $\mathcal{T}$ and $\mathcal{R}$ are the universes defined previously
- $I = I_\theta(r_i) = \{r_j : \omega(T(r_i), T(r_j)) \geq \theta\}$ describes tolerance classes for contexts
- $\omega(A, B) = \frac{2|A \cap B|}{|A| + |B|}$ is the overlap index
- $\nu(X, Y) = \frac{|X \cap Y|}{|X|}$ measures degree of inclusion

- $\mathcal{L}_\mathcal{A}(t_i) = \{r_j \in \mathcal{R} : \nu(I_\theta(r_j), R(t_i)) = 1\}$
- $\mathcal{U}_\mathcal{A}(t_i) = \{r_j \in \mathcal{R} : \nu(I_\theta(r_j), R(t_i)) > 0\}$

- *Sengoz, C., Ramanna, S., Learning Relational Facts From the Web: A Tolerance Rough Set Approach, Pattern Recognition Letters , Elsevier, 2015, 67(P2):130-137.

Dataset: Unary Relations

- Original source is ClueWeb09 [1]. (50+ million web documents.)
- We used the all-pairs treatment [2] by Andy Carlson.
- Sub-sampled 70,000 noun phrases and 60,000 contexts in the form of a matrix.
- Implemented in MATLAB®

1 Jamie Callan and Mark Hoy. Clueweb09 Data Set, 2009

2 Carlson, A.: All-pairs data set (2010)

3 A. Carlson, J. Betteridge, R. C. Wang, E. R. Hruschka, Jr., and T. M. Mitchell. Coupled Semi-supervised Learning for Information Extraction. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, pages 101–110, 2010

4 S. Verma and E. R. Hruschka, Jr. Coupled Bayesian Sets Algorithm for Semi-supervised Learning and Information Extraction. In *ECML PKDD Part II LNCS 7524*, pages 307–322, 2012

TPL 1.0 Results: Precision@30 for Unary Relations

Categories	Iteration 5		Iteration 10	
	TPL	CBS	TPL	CBS
Company	100%	100%	100%	100%
Disease	100%	100%	100%	100%
KitchenItem	100%	94%	100%	94%
Person	100%	100%	100%	100%
PhysicsTerm	93%	100%	90%	100%
Plant	100%	100%	97%	100%
Profession	100%	100%	100%	87%
Sociopolitics	100%	48%	100%	34%
Sport	97%	97%	100%	100%
Website	90%	94%	90%	90%
Vegetable	93%	83%	63%	48%
Average	97.5%	**92%**	94.5%	**87%**

Sengoz, C., Ramanna, S.: A semi-supervised learning algorithm for web information extraction with tolerance rough sets. In: Proc. of Web Intelligence Congress, Active Media Technology 2014, LNCS 8610. Springer, 1-10

Dataset: Binary Relations

Experimental Setup

- Sub-sampled 13 million noun phrase pairs and 11 million contexts in form of a matrix.
- Implemented in C++.
- 10 Categories, 5 Seeds per Category, 10 Iterations

Evaluation

1. **Ranking-based** Precision@30: In any iteration, after noun phrases are scored and ranked for a relation, the percentage of the correct pairs in the set of the top 30-ranked pairs is calculated.
2. **Promotion-based** Precision@30: From the set of all promoted pairs for a given relation, we sampled 30 pairs to be evaluated and we calculated the percentage of the correct pairs within that set.

TPL 1.0 Results: Precision@30 for Binary Relations

Evaluation	**Ranking-based**			**Promotion-based**			
	TPL			**TPL**			**CPL**
Iterations	**1**	**5**	**10**	**1**	**5**	**10**	**10**
Categories							
Athlete-Team	100	90	87	100	96	87	100
CEO-Company	100	100	100	100	100	100	100
City-Country	100	100	100	100	100	100	93
City-State	100	100	100	100	100	100	100
Coach-Team	93	93	93	100	100	93	100
Company-City	83	90	93	40	84	97	50
Stadium-City	97	93	80	80	92	70	100
State-Capital	100	97	73	100	100	63	60
State-Country	100	100	100	100	100	100	97
Team-vs-Team	93	83	80	100	84	80	100
Average	96.6	94.6	90.6	92.0	95.6	89.0	**90.0**

Fuzzy Rough Pattern Learner (FRL)*

Motivation

- Fuzzy Rough Sets permit overlapping or soft similarity classes
- Gain insights into the strengths and weakness of integration of fuzzy and rough sets for categorization of linguistic entities
- Previously applied to query expansion problem for document retrieval**
- Study the effects of concept drift by using the same dataset, iterations and evaluation measures of TPL, CBS and CPL

Solution

1. Instead of a crisp co-occurrence matrix, create a fuzzy (graded) co-occurrence matrix
2. Approximate fuzzy contextual patterns (with rough set operators)
3. Create a new scoring mechanism

*Bharadwaj, A and Ramanna, S. Categorizing Relational Facts from the Web with Fuzzy Rough Sets, Knowledge and Information Systems Journal, Springer, 2019, Volume. 61, Issue 3, 1695-1713.

**M. De Cock and C. Cornelis. Fuzzy rough set based web query expansion. In Proceedings of Rough Sets and Soft Computing in Intelligent Agent and Web Technology, pages 9–16, 2005.

Fuzzyfying co-occurrence information: Binary relations

first step is to normalize the co-occurrence statistics.

$$\vartheta(h_{ij}, r_k) = \frac{f_R(h_{ij}, r_k)}{f_R(h_{ij}, r_k), \forall k{:}1..Q}$$

second step is to fuzzifying the normalized data.

$$S(\vartheta; \alpha, \beta) = \begin{cases} 1 & \text{if } \vartheta \ \geq \ \beta \\ \frac{\vartheta - \alpha}{\beta - \alpha} & \text{if } 0.005 \ \leq \ \vartheta \ < \ \beta \\ 0, & \text{otherwise} \end{cases}$$

$\alpha = 0.001$ and $\beta = 0.02$

Lower And Upper Approximations: Binary Relations

$I = (\mathcal{H}, \mathcal{R}, CO_F)$

- $\mathcal{H}$ denotes the universe of relations.
- $\mathcal{R}$ represents the co-occurring contextual patterns.
- CO_F is a fuzzy set in $\mathcal{H} \times \mathcal{R}$.

- The *upper* and *lower* approximations of the fuzzy set $\mathcal{H}_\mathcal{F}$ in I is denoted by $\mathcal{H}_\mathcal{F} \uparrow CO_F$ and $\mathcal{H}_\mathcal{F} \downarrow CO_F$

- $\mathcal{H}_\mathcal{F} \uparrow CO_F = \sup\limits_{h_{ij} \in \mathcal{H},\ h_{xy} \in \mathcal{TR}} (CO_F(R(h_{ij}), h_{xy}), \mathcal{H}_\mathcal{F}(h_{xy}) : CO_\mathcal{F}(h_{ij}) \geq CO_\mathcal{F}(h_{xy}))$
- $\mathcal{H}_\mathcal{F} \downarrow CO_F = \inf\limits_{h_{ij} \in \mathcal{H},\ h_{xy} \in \mathcal{TR}} (CO_F(R(h_{ij}), h_{xy}), \mathcal{H}_\mathcal{F}(h_{xy}) : ((h_{ij}, h_{xy}) | R(h_{xy}) \cap R(h_{ij}) \neq \emptyset)))$

Tight Upper Approximation and Similarity Score

Tight Upper Approximation

$$CO_F \downarrow\uparrow \mathcal{H}_F(h_{ij}) = CO_F \downarrow (CO_F \uparrow \mathcal{H}_F(h_{ij}))$$

Similarity Score

$$micro(h_{ij}) = \omega_1(\mathcal{H}_{\mathcal{F}} \uparrow CO_F) + \omega_2(\mathcal{H}_{\mathcal{F}} \downarrow CO_F))$$

ω_1 and ω_2 are application dependent.

Fuzzy Rough Learner Algorithm: Binary Relations

Input : An ontology O defining categories; a large corpus $\mathcal{H}$, CO co-occurrence matrix, a small set of trusted relations $\mathcal{TR}$

Output: Trusted instances h_{xy} for $\mathcal{TR}'$, where $\mathcal{TR}'$ is a set of all new promoted trusted noun pair(relation) phrases

for $r = 1 \to$ end of file **do**
    for each category *cat* **do**
        for each new trusted relations h_{xy} belonging to *cat* **do**
            for each candidate relation h_{ij} **do**
                Calculate Fuzzy Relation $\mathcal{CO}_{\mathcal{F}}$;
                Calculate Upper Approximation $U_{\mathcal{H}_F}(h_{ij})$;
                Calculate score ω_1;
                for each candidate relation h_{ij} **do**
                    Calculate Lower Approximation $L_{\mathcal{H}_F}(h_{ij})$;
                    Calculate score ω_2;
            Calculate $micro_{cat}(h_{ij})$;
        Sort trusted instances h_{xy} by $micro_{cat}/|cat|$;
        Promote top trusted instances, such that $\mathcal{TR}' = \mathcal{TR} \cup \{h_{xy}\}$;

FRL Results: Precision@30 for Unary Relations

Categories	Iteration 5			Iteration 10		
	TPL	**CBS**	**FRL**	**TPL**	**CBS**	**FRL**
Company	100	100	100	100	100	100
Disease	100	100	100	100	100	100
KitchenItem	100	94	97	100	94	73
Person	100	100	100	100	100	100
PhysicsTerm	93	100	67	90	100	77
Plant	100	100	77	97	100	100
Profession	100	100	100	100	87	100
Sociopolitics	100	48	93	100	34	87
Sport	97	97	100	100	100	100
Website	90	94	97	90	90	93
Vegetable	93	83	83	63	48	47
Average	97.5	**92**	**92**	94.5	**87**	**89**

Bharadwaj, A., Ramanna, S.: Fuzzy rough set-based unstructured text categorization. Proceedings of 30th Canadian Artificial Intelligence Conference, LNAI 10233, pp. 335-340, 2017

FRL promotion-based results: Precision@30 for Binary Relations

Categories	**TPL**			**FRL**			**CPL**
	1	**5**	**10**	**1**	**5**	**10**	**10**
Athlete||Team	100	96	87	100	100	83	100
CEO||Company	100	100	100	100	100	100	100
City||Country	100	100	100	100	93	96	93
City||State	100	100	100	100	100	100	100
Coach||Team	100	100	93	100	100	100	100
Company||City	40	84	97	100	100	100	50
Stadium||City	80	92	70	80	92	90	100
State||Capital	100	100	63	100	88	43	60
State||Country	100	100	100	100	100	100	97
Team|vs|Team	100	84	80	100	96	100	100
Average	92.0	95.6	**89.0**	98.0	96.6	91.2	**90.0**

FRL ranking-based results: Precision@30 for Binary Relations

Categories	**TPL**			**FRL**		
	Iter. 1	**Iter. 5**	**Iter.10**	**Iter. 1**	**Iter. 5**	**Iter. 10**
Athlete||Team	100	90	87	97	100	97
CEO||Company	100	100	100	100	100	100
City||Country	100	100	100	93	100	100
City||State	100	100	100	97	100	100
Coach||Team	93	93	93	100	100	100
Company||City	83	90	93	97	100	100
Stadium||City	97	93	80	93	70	93
State||Capital	100	97	73	93	83	77
State||Country	100	100	100	90	100	100
Team|vs|Team	93	83	80	100	100	100
Average	96.6	94.6	**90.6**	96	95.3	96.7

Bharadwaj, A., Ramanna, S. Categorizing Relational Facts from the Web with Fuzzy Rough Sets, Knowledge and Information Systems Journal, Springer, 2019, Volume. 61, Issue 3, 1695-1713.

*Mutual Exclusion Constraints were applied with FRL.

TPL 2.0- Next Version*

Motivation

- Explore Scalability of TPL 1.0
- Handling of concept drift in a larger dataset
- Question: Do we need to define additional constraints?

Solution

1. **Extract** categorical information from a large noisy dataset of crawled web pages (733,019,372 English web pages of ClueWeb2012- 6TB)
2. **Prepare** contextual co-occurrence matrix
3. **Extend** the number of iterations

*Moghaddam, H and Ramanna, S., Harvesting Patterns from Textual Web Sources with Tolerance Rough Sets, Patterns Journal, Cell Press, 2020, https://doi.org/10.1016/j.patter.2020.100053

Results with TPL 2.0

Categories	Iteration 5				Iteration 10			
	TPL 2.0	**TPL 1.0**	**CBS**	**FRL**	**TPL 2.0**	**TPL 1.0**	**CBS**	**FRL**
Company	100	100	100	100	100	100	100	100
Disease	100	100	100	100	100	100	100	100
KitchenItem	97	100	94	97	97	100	94	73
Person	100	100	100	100	100	100	100	100
PhysicsTerm	97	93	100	67	97	90	100	77
Plant	94	100	100	77	97	97	100	100
Profession	100	100	100	100	100	100	87	100
Sociopolitics	94	100	47	93	97	100	34	87
Sport	100	97	97	100	100	100	100	100
Website	97	90	94	97	97	90	90	93
Vegetable	74	93	83	83	90	63	48	47
Average	95.7	97.5	92	92	97.7	**94.5**	**87**	**89**

130,536 noun phrases and 118,648 contextual patterns. TPL 2.0 results for iteration 20 was 96.2%

Summary of our experiments

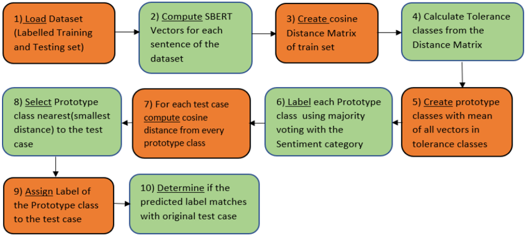
CPL	CBS	TPL 1.0	FRL	TPL 2.0
Core component of NELL	Based on Bayesian Sets	Based on Tolerance Rough Sets	Based on Fuzzy Rough Sets	Based on Tolerance Rough Sets
Corpus of 200-million webpages	Subset from ClueWeb09	Subset from ClueWeb09	Subset from ClueWeb09	Subset from ClueWeb12 (larger data set)
Concept drift- three constraints	Mutual Exclusion constraint	No constraints	Mutual Exclusion constraint	No constraints
Learning Relational (binary) Facts	Learning Relational (unary) Facts	Learning Relational (unary and binary) Facts	Learning Relational (unary and binary) Facts	Learning Relational (unary) Facts
	Outperforms CBS (10^{th} iteration)	Outperforms CBS and CPL (10^{th} iteration)	Outperforms TPL 1.0 on relational (binary) and comparable with unary facts	Outperforms TPL 1.0, FRL, CBS and explores concept drift (20^{th} iteration)

Ramanna, S., Peters, J., Sengoz, C.: Application of tolerance rough sets in structured and unstructured text categorization: A survey. In: G.Wang et al. (eds.),Thriving Rough Sets, Studies in Computational Intelligence, 708, Springer, pp. 119-173 (2017)

TNS Model for Non-topic classification

Tolerance Near Sets-based Classifier that leverages

- Pre-trained birectional transformer encoders
- Efficient feature vector embeddings from textual units
- Sentiment Classification and News Categorization Tasks



Supervised learning algorithm

- Tolerance Classes are induced directly from feature vectors
- Similarity scoring based on distance function and a predefined tolerance level
- First applied to Solar Flare Images*

*G. Poli, E.Llapa, J.R. Cecatto, J.H. Saito, J.F. Peters, S. Ramanna, M.C. Nicoletti: Solar Flare Detection System Based on Tolerance Near Sets in a GPU-CUDA Framework, *Knowledge-based Systems Journal*, Elsevier, 70(1):345–360, 2014

Preliminaries*

Definition

Text-based Tolerance Relation $\cong_{\mathcal{T},\epsilon}$
Let $\langle T, F \rangle$ be a universe of nonempty set of objects T and F be the feature set. Let $\mathcal{T} \subseteq F$ where $\mathcal{T}$ represents textual features. A tolerance space $\langle T, \cong_{\mathcal{T},\epsilon} \rangle$ is defined as:

$$\cong_{\mathcal{T},\epsilon} = \{(t_i, t_j) \in T \times T : \ dist(t_i, t_j) \leq \varepsilon\} \tag{1}$$

where *dist* is the cosine distance given as follows:

$$dist(t_i, t_j) = 1 - \frac{\phi(t_i).\phi(t_j)}{\|\phi(t_i)\| \, \|\phi(t_j)\|} \tag{2}$$

The tolerance relation $\cong_{\mathcal{T},\epsilon}$ induces a tolerance class TC where ε is a user-defined tolerance level.

*Vrushang Patel, Sheela Ramanna, Ketan Kotecha, and Rahee Walambe, Short Text Classification with Tolerance-Based Soft Computing Method, Algorithms, MDPI 2022, https://www.mdpi.com/1999-4893/15/8/267

TSC - Training Phase to generate representative vectors

Input : $TV = \{TV_1, \ldots, TV_M\}$, // Transformer Vectors
$\varepsilon > 0$, // Tolerance level parameter
Output: $(NT, \{(R_1, TextCat_1), \ldots, (R_{NT}, TextCat_{NT})\})$
NT is the size of the Tolerance class set

for $p \leftarrow 1$ **to** M **do**
    for $q \leftarrow p + 1$ **to** M **do**
        computeCosineDist(TV_p, TV_q, Cos_{pq})

for $i \leftarrow 1$ **to** M **do**
    for $j \leftarrow i + 1$ **to** M **do**
        ObjectPairs $\leftarrow$ generatetolerantpairs($\mathrm{Cos}_{ij}, \varepsilon$);
        $N_i \leftarrow$ createobjectneighbour(ObjectPairs, i, TV);
        for ***all***, $o_1, o_2 \in N_i$ **do**
            if $o_1, o_2 \in ObjectPairs$ **then**
                $TC_i \leftarrow \{o_2\}$;
    $T \leftarrow T \cup \{TC_i\}$;
    $TextCat_i \leftarrow$ computeMajorityPol(T_i); //
$NT \leftarrow |T|$; // Number of tolerance classes in T
$\{(R_1, TextCat_1), \ldots, (R_{NT}, TextCat_{NT})\} \leftarrow$
GenerateClassRepresentative(NT);

TSC - Classification Phase

Algorithm 2: TSC Classification Phase: Assigning Sentiment Classes

Input : $\varepsilon > 0$, // Tolerance level parameter
, NT // Size of the Tolerance class set T
, $TV' = \{TV'_1, \ldots, TV'_M\}$, // Transformer Vectors for testing
$\{(R_1, TextCat_1), \ldots, (R_{NT}, TextCat_{NT})\}$ // Representative class vectors generated in the training phase and their associated classes

Output: $(TV' = \{(TV'_1, TextCat_1), \ldots, (TV'_M, TextCat_M)\})$ // Transformer Vectors with assigned categories

for $i \leftarrow 1$ **to** M **do**
  for $j \leftarrow i + 1$ **to** NT **do**
    computeCosineDist(TV'_i, R_j, Cos_{ij});

$TV' \leftarrow$ DetermineClass(Cos_{ij}) // Computes min. distance and assigns classes to the test set vector

Vrushang Patel and Sheela Ramanna, Tolerance-based short text Sentiment Classifier, Proceedings of International Joint Rough Sets Conference 2021, Bratislava, Slovakia, LNAI 12872, pp 259-265.

Dataset: Sentiment Classification

Dataset	Type	Size	Positive	Negative	Neutral	Irrelevant
Covid-Sentiment	Train Test	7000 1003	22.02% 23.53%	30.35% 37.29%	47.63% 39.18%	- -
U.S. Airline Sentiment	Train Test	12000 1000	16.79% 13%	61.02% 67.5%	22.19% 19.5%	- -
IMDB Movie Review	Train Test	20000 2000	50.27% 50.35%	49.73% 49.65%	- -	- -
SST-2	Train Test	15000 1500	55.37% 55.53%	44.63% 44.47%	- -	- -
Sentiment140	Train Test	15000 1000	50% 50%	50% 50%	- -	- -
SemEval 2017	Train Test	17001 3546	40.67% 41.54%	15% 15.76%	44.33% 42.70%	- -
Sanders corpus	Train Test	4059 1015	10.24% 9.85%	11.38% 10.54%	45.26% 47.68%	33.12% 31.93%
UCI Sentence	Train Test	2700 300	49.11% 58%	50.89 42%	- -	- -
Dataset	**Type**	**Size**	**World**	**Sports**	**Business**	**Science**
AG-News	Train Test	12000 1150	3000 300	3000 250	3000 300	3000 300

TSC Results - Weighted F1-score

Table: SBERT vector-based weighted F1-score (rounded) results for six classifiers. Best results are in bold-face.

Dataset	**TSC-mean**	**RF**	**ME**	**SVM**	**SGD**	**LGBM**
Covid-Sentiment (3)	55	44	57	57	**57**	56
U.S. Airline (3)	**77**	77	77	77	75	77
IMDB (2B)	**76**	69	73	73	72	72
SST-2 (2B)	85	85	85	**86**	85	85
Sentiment140 (2)	70	68	**72**	72	66	70
SemEval (3) 2017	60	54	**64**	63	63	60
Sanders corpus (4)	69	70	**76**	74	76	75
UCI Sentence (2B)	**89**	84	86	87	87	83
AG-News (4B)	82	79	88	81	**88**	83
20-Newsgroups (B)	**66**	41	58	52	52	53

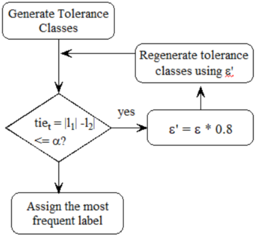
Issues

- PRC-AUC, ROC-AUC and Weighted F1 scores were examined
- Balanced Tolerance Classes
- Number of Sentiment Categories
- Length of words (short and long)
- Quality of Vector Embeddings

Impact of different Embeddings on TSC

Assessing Impact of the following

- DistilBERT, MiniLM, and Word2Vec Word Embeddings
- Examining labelling of Prototype Vectors
- Examining imbalanced tolerance classes



TSC 2.0* Supervised learning algorithm

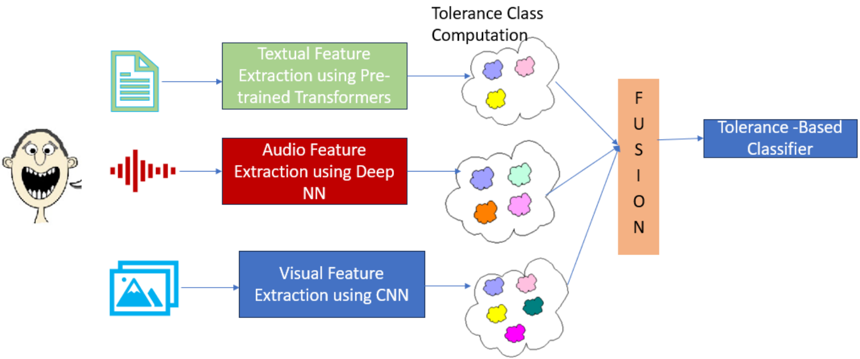
- Includes a tie-breaking and variance classification method
- Includes feature vectors drawn from combination of embedding methods

*T.Hegde, K. S. Sanjay, S. M. Thomas, R. Kambhammettu, A.Kumar M, S. Ramanna, Impact of Vector Embeddings on the Performance of Tolerance Near Sets-based Sentiment Classifier for Text Classification, Proc. of KES, 2023 [to appear].

Observations - F1 Scores

- DistilBERT was the most effective embedding for IMDb, US Airline and Sentiment 140 datasets
- DistilBERT + MiniLM with tie-breaking condition gave the best score for AG news dataset
- Adding additional word embedding did not work for most datasets
- Overall TSC 2.0 has better F1-scores than TSC 1.0
    - IMDB: TSC 2.0 (79.8%) vs. TSC 1.0 (76%)
    - US Airline: TSC 2.0 (78.5%) vs. TSC 1.0 (77%)
    - Sentiment 140 : TSC 2.0 (71%) vs. TSC 1.0 (70%)
    - AG news : TSC 2.0 (88%) vs. TSC 1.0 (82%) - 4 classes

Tolerance-based Multimodal Sentiment Classifier



Multimodal Information Processing**

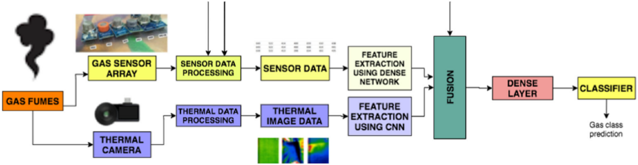


Figure: Case Study- Multimodal Co-learning of sensor fusion in Gas Detection*

**Anil Rahate, Rahee Walambe, Sheela Ramanna, Ketan Kotecha, Multimodal Co-learning: Challenges, applications with datasets, recent advances and future directions, Information Fusion Journal, Elsevier, Volume 81, 2022, Pages 203-239,ISSN 1566-2535, https://doi.org/10.1016/j.inffus.2021.12.003

*Anil Rahate, Shruti Mandaokar, Pulkit Chandel , Rahee Walambe, Sheela Ramanna, Ketan Kotecha, Employing Multimodal Co-learning to Evaluate the Robustness of Sensor Fusion for Industry 5.0 Tasks, Soft Computing Journal, Springer, volume 27, pages,4139–4155 (2023)

Concluding Remarks

- Novel models for representing Linguistic Entities
- Tolerance-based framework for semi-supervised machine learning
- Demonstrated efficacy with benchmark datasets (ClueWeb) and algorithms (CPL and CBS - NELL)
- Acknowledgments

  - Natural Sciences and Engineering Council of Canada (NSERC) Discovery Grants Program
  - Colleagues: Prof. Anand Kumar, National Institute of Technology (NIT) Karnataka, India
  - Prof. Ketan Kotecha and Prof. Rahee Walambe, SIT, Pune, India
  - Graduate Students: Cenker Sengoz, Aditya Bharadwaj, Hoora Rezai Moghaddam, Vrushang Patel (Msc)
  - Anil Rahate (PhD)- SIT Pune
  - Undergraduate Students: NIT, UWinnipeg, IIT-Kharagpur