Using Common-Sense knowledge-base for Detecting Word Obfuscation in Adversarial Communication

Swati Agarwal, Ashish Sureka

IIIT-Delhi | Indraprastha Institute of Information Technology

swatia@iiitd.ac.in, ashish@iiitd.ac.in

January 10, 2015

Table of Contents

Table of Contents

Message Interception - Intelligence and Security Agencies

- Intelligence and security agencies intercepts and scans billions of messages and communications every day to identify dangerous communications between terrorists and criminals
- Law enforcement agencies use message interception to combat criminal and illicit acts
- Terrorist and criminals use textual or word obfuscation to prevent their messages from getting intercepted by the law enforcement agencies
- Automatic word obfuscation detection is natural language processing problem that has attracted several researcher's attention

Message Interception - Intelligence and Security Agencies

- Intelligence and security agencies intercepts and scans billions of messages and communications every day to identify dangerous communications between terrorists and criminals
- Law enforcement agencies use message interception to combat criminal and illicit acts
- Terrorist and criminals use textual or word obfuscation to prevent their messages from getting intercepted by the law enforcement agencies
- Automatic word obfuscation detection is natural language processing problem that has attracted several researcher's attention

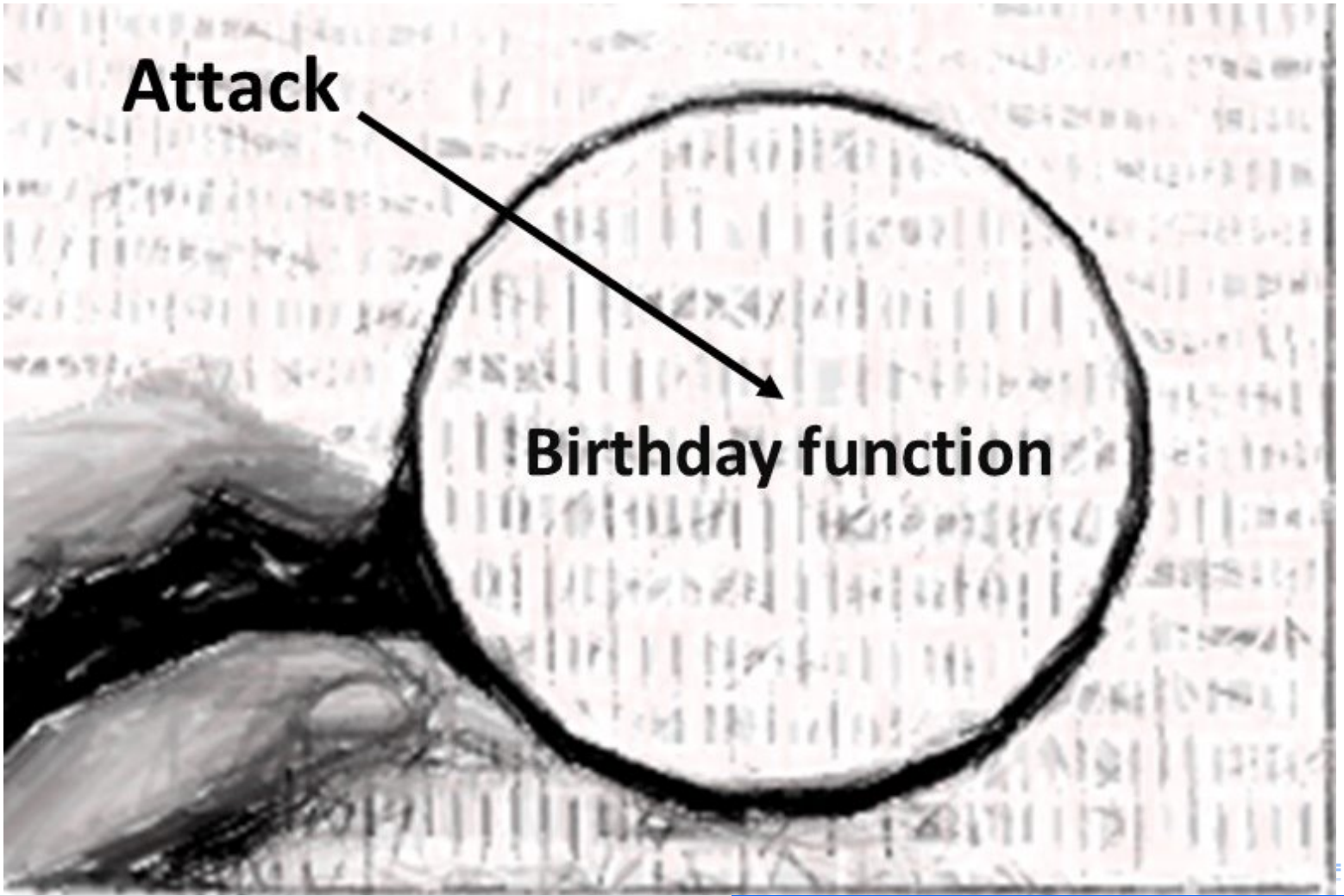
Message Interception - Intelligence and Security Agencies

- Intelligence and security agencies intercepts and scans billions of messages and communications every day to identify dangerous communications between terrorists and criminals
- Law enforcement agencies use message interception to combat criminal and illicit acts
- Terrorist and criminals use textual or word obfuscation to prevent their messages from getting intercepted by the law enforcement agencies
- Automatic word obfuscation detection is natural language processing problem that has attracted several researcher's attention

Message Interception - Intelligence and Security Agencies

- Intelligence and security agencies intercepts and scans billions of messages and communications every day to identify dangerous communications between terrorists and criminals
- Law enforcement agencies use message interception to combat criminal and illicit acts
- Terrorist and criminals use textual or word obfuscation to prevent their messages from getting intercepted by the law enforcement agencies
- Automatic word obfuscation detection is natural language processing problem that has attracted several researcher's attention

Term Obfuscation



Commonsense Knowledge-base - Term Obfuscation

- ConceptNet is a semantic network consisting of nodes representing concepts and edges representing relations between the concepts
- We hypothesize that ConceptNet can be used as a semantic knowledge-base to solve the problem of textual or word obfuscation
- To investigate the application of a commonsense knowledge-base such as ConceptNet for solving the problem of word or textual obfuscation
- To conduct an empirical analysis on large and real-word datasets for the purpose of evaluating the effectiveness of the application of ConceptNet

Commonsense Knowledge-base - Term Obfuscation

- ConceptNet is a semantic network consisting of nodes representing concepts and edges representing relations between the concepts
- We hypothesize that ConceptNet can be used as a semantic knowledge-base to solve the problem of textual or word obfuscation
- To investigate the application of a commonsense knowledge-base such as ConceptNet for solving the problem of word or textual obfuscation
- To conduct an empirical analysis on large and real-word datasets for the purpose of evaluating the effectiveness of the application of ConceptNet

Commonsense Knowledge-base - Term Obfuscation

- ConceptNet is a semantic network consisting of nodes representing concepts and edges representing relations between the concepts
- We hypothesize that ConceptNet can be used as a semantic knowledge-base to solve the problem of textual or word obfuscation
- To investigate the application of a commonsense knowledge-base such as ConceptNet for solving the problem of word or textual obfuscation
- To conduct an empirical analysis on large and real-word datasets for the purpose of evaluating the effectiveness of the application of ConceptNet

Commonsense Knowledge-base - Term Obfuscation

- ConceptNet is a semantic network consisting of nodes representing concepts and edges representing relations between the concepts
- We hypothesize that ConceptNet can be used as a semantic knowledge-base to solve the problem of textual or word obfuscation
- To investigate the application of a commonsense knowledge-base such as ConceptNet for solving the problem of word or textual obfuscation
- To conduct an empirical analysis on large and real-word datasets for the purpose of evaluating the effectiveness of the application of ConceptNet

ConceptNet - Common Sense - Semantic Network

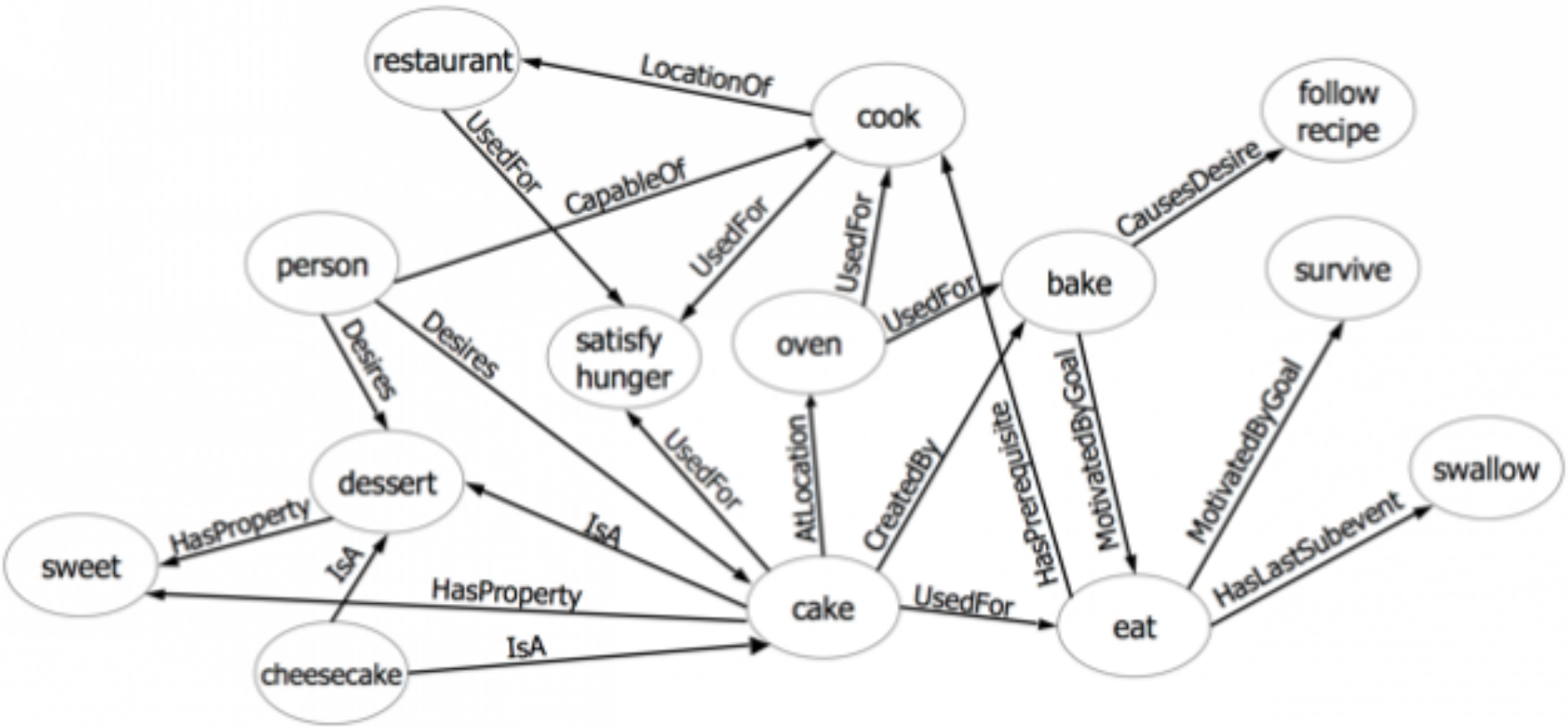


Table of Contents

List of Previous Work - Reverse Chronological Order

Table: ED: Evaluation Dataset, RS: Resources Used in Solution Approach, SA: Solution Approach

Deshmukh et al. 2008 [1]	
ED	Google News
RS	Google search engine
SA	Measuring sentence oddity, enhance sentence oddity and k-grams frequencies
Jabbari et al. 2008 [2]	
ED	British National Corpus (BNC)
RS	1.4 billion words of English Gigaword v.1 (newswire corpus)
SA	Probabilistic or distributional model of context
Fong et al. 2008 [3]	
ED	Enron e-mail dataset, Brown corpus
RS	British National Corpus (BNC), WordNet, Yahoo, Google and MSN search engine
SA	Sentence oddity, K-gram frequencies, Hypernym Oddity (HO) and Pointwise Mutual Information (PMI)
Fong et al. 2006 [4]	
ED	Enron e-mail dataset
RS	British National Corpus (BNC), WordNet, Google search engine
SA	Sentence oddity measures, semantic measure using WordNet, and frequency count of the bigrams around the target word

Research Contributions

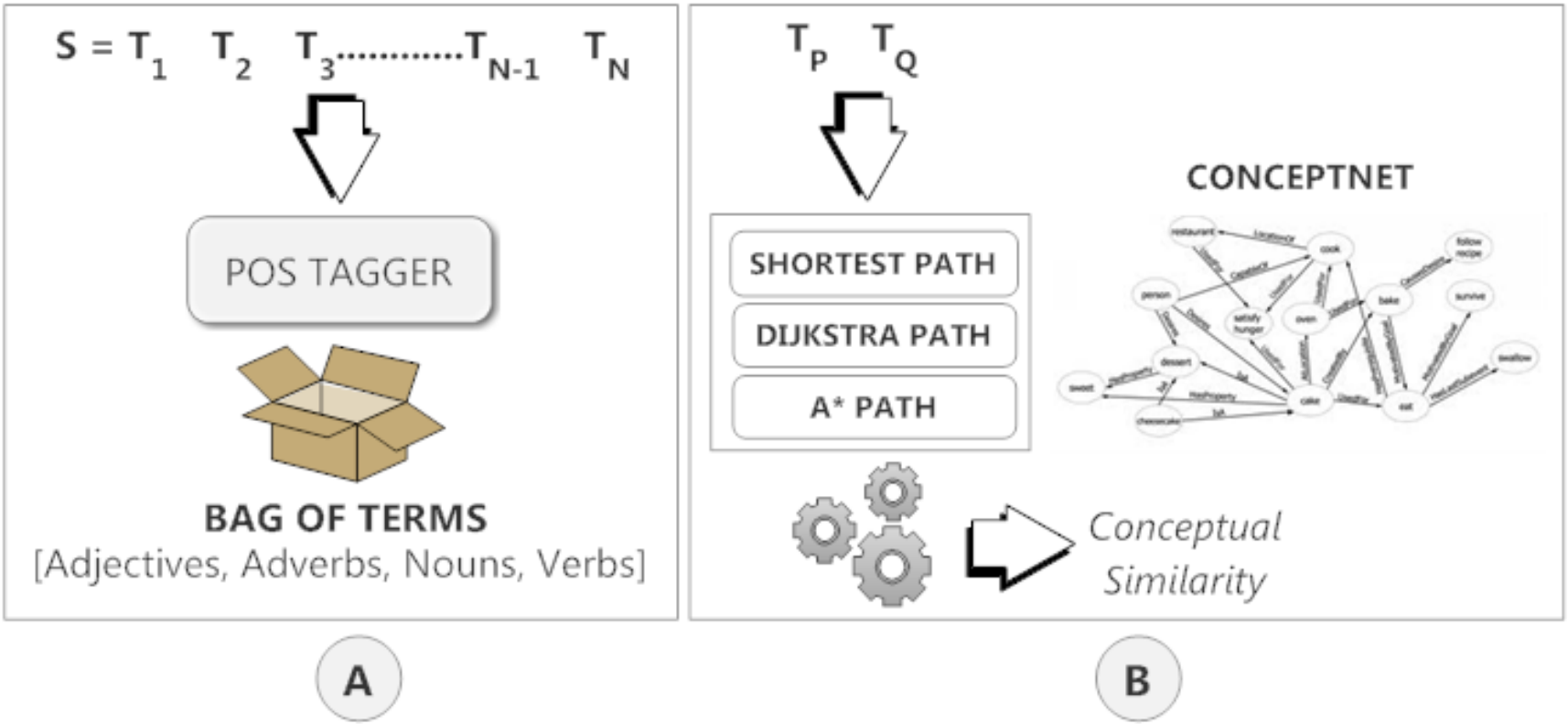
- First focused research investigation on the application of ConceptNet common sense knowledge-base for solving the problem of textual or term obfuscation
- We conduct an in-depth empirical analysis to examine the effectiveness of the proposed approach

Research Contributions

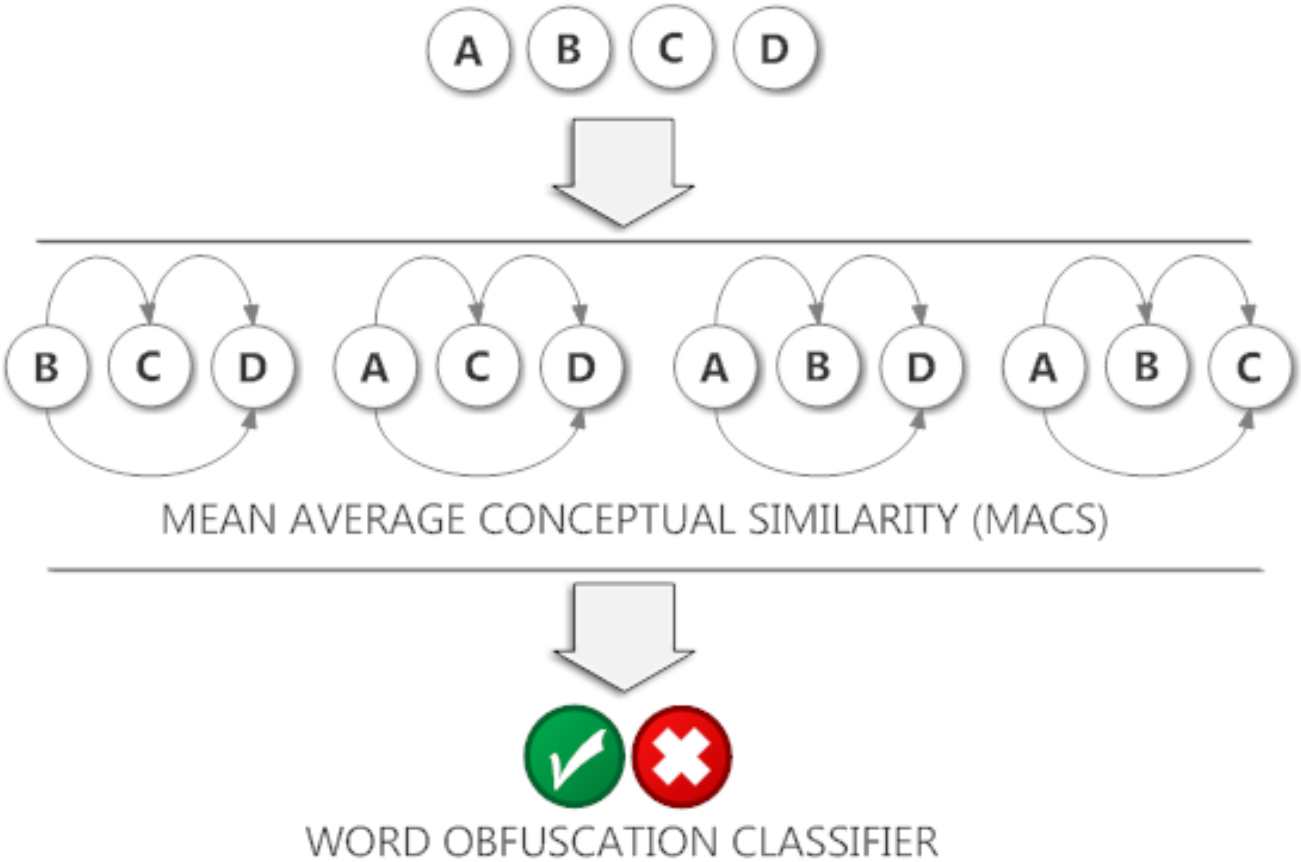
- First focused research investigation on the application of ConceptNet common sense knowledge-base for solving the problem of textual or term obfuscation
- We conduct an in-depth empirical analysis to examine the effectiveness of the proposed approach

Table of Contents

Multiple phase - Processing Pipeline.



Mean Average Conceptual Similarity (MACS) Score



Worked-out Examples

Example 1

Original Sentence: "We will attack the airport with bomb"
Red-flagged term: bomb
Replacement Word: flower
Replaced Sentence: "We will attack the airport with flower"
Bag-of-terms: attack, airport, flower

Example 1

Conceptual similarity between airport and flower is 3
The number of edges between airport and flower is 3
The number of edges between flower and airport is 3
Conceptual similarity between attack and flower is 3
The number of edges between attack and flower is 3
The number of edges between flower and attack is 3
Conceptual similarity between attack and airport is 2.5
The number of edges between attack and airport is 2
The number of edges between airport and attack is 3
MACS: $(3 + 3 + 2.5)/3 = 2.83$

Worked-out Examples

Example 1

Original Sentence: "We will attack the airport with bomb"
Red-flagged term: bomb
Replacement Word: flower
Replaced Sentence: "We will attack the airport with flower"
Bag-of-terms: attack, airport, flower

Example 1

Conceptual similarity between airport and flower is 3
The number of edges between airport and flower is 3
The number of edges between flower and airport is 3
Conceptual similarity between attack and flower is 3
The number of edges between attack and flower is 3
The number of edges between flower and attack is 3
Conceptual similarity between attack and airport is 2.5
The number of edges between attack and airport is 2
The number of edges between airport and attack is 3
MACS: $(3 + 3 + 2.5)/3 = 2.83$

Worked-out Examples

Example 2

Original Sentence: "Pistol will be delivered to you to shoot the president"
Red-flagged term: pistol
Replacement Word: pen
Replaced Sentence: "Pen will be delivered to you to shoot the president"
Bag-of-terms: pen, shoot, president

Example 2

Conceptual similarity between shoot and president is 2.5
The number of edges between president and shoot is 2
The number of edges between shoot and president is 3
Conceptual similarity between pen and president is 3
The number of edges between president and pen is 3
The number of edges between pen and president is 3
Conceptual similarity between pen and shoot is 3
The number of edges between shoot and pen is 3
The number of edges between pen and shoot is 3
MACS: $(2.5 + 3 + 3)/3 = 2.83$

Worked-out Examples

Example 2

Original Sentence: "Pistol will be delivered to you to shoot the president"
Red-flagged term: pistol
Replacement Word: pen
Replaced Sentence: "Pen will be delivered to you to shoot the president"
Bag-of-terms: pen, shoot, president

Example 2

Conceptual similarity between shoot and president is 2.5
The number of edges between president and shoot is 2
The number of edges between shoot and president is 3
Conceptual similarity between pen and president is 3
The number of edges between president and pen is 3
The number of edges between pen and president is 3
Conceptual similarity between pen and shoot is 3
The number of edges between shoot and pen is 3
The number of edges between pen and shoot is 3
MACS: $(2.5 + 3 + 3)/3 = 2.83$

Obfuscated Term Detection

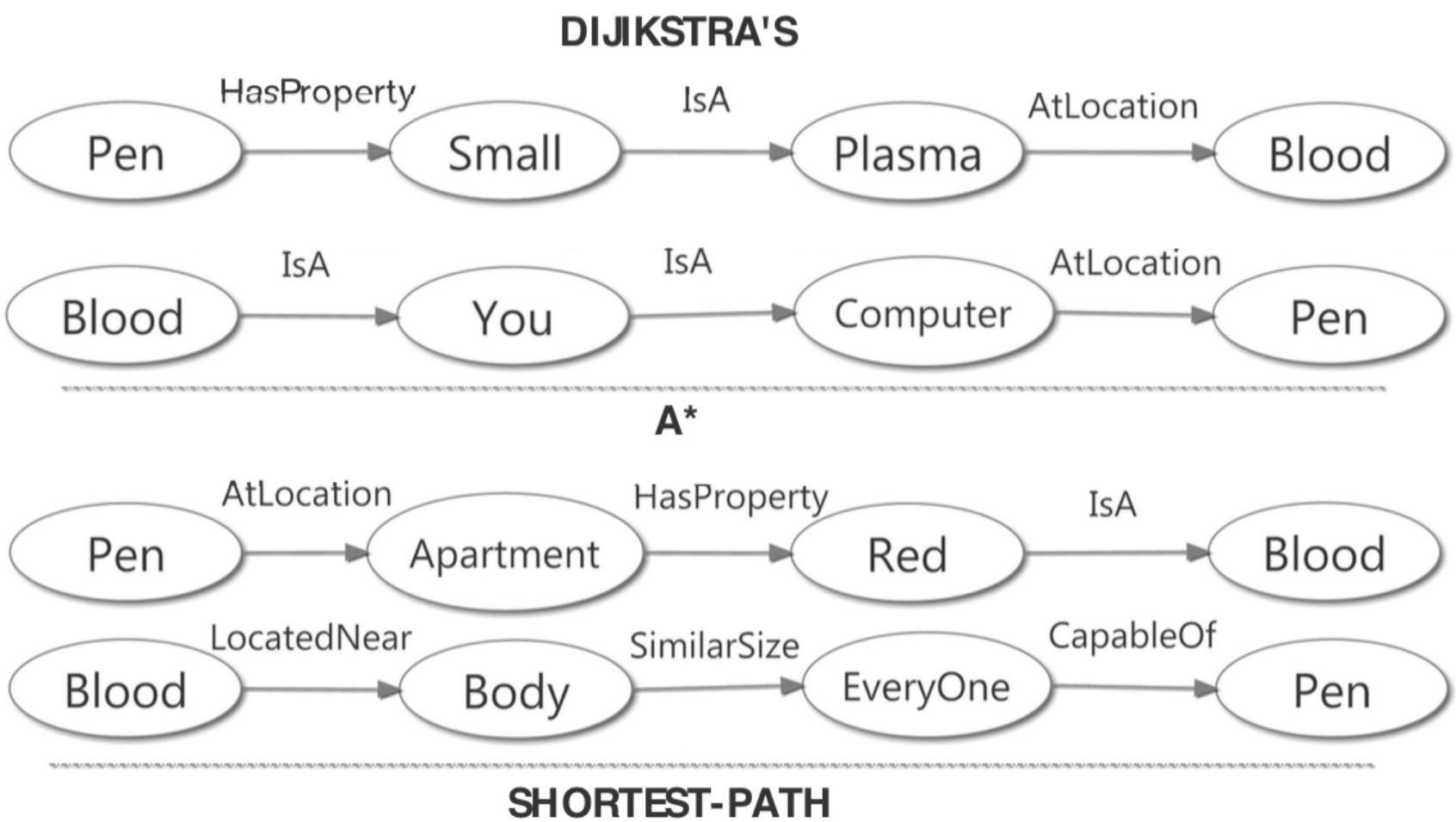
Data: Substituted Sentence S , Conceptnet Corpus C
Result: Obfuscated Term O_T

```
for all record r ∈ C do
    Edge E.add(r.node_1 , r.node_2 , r.relation)
    Graph G.add(E)

tokens = S'.tokenize()
pos.add(pos_tag(tokens))
for all tag ∈ pos and token ∈ tokens do
    if tag is in (verb, noun, adjective, adverb) then
        BoW.add(token.lemma)

for iter = 0 to BoW.length do
    concepts = BoW.pop(iter)
    for i = 0 to concepts.length − 1 do
        for j = i to concepts.length do
            if (i! = j) then
                pathc_{i,j} = Dijikstrapathlen(G, i, j)
                pathc_{j,i} = Dijikstrapathlen(G, j, i)
                avg.add(Average(c_{i,j}, c_{j,i})
    mean.add(Mean(avg))
O_T = BoW.valueAt(min(mean))
```

Solution Pseudo-Code and Algorithm



Solution Pseudo-Code and Algorithm

Concrete Examples of Computing Conceptual Similarity between Two Given Terms Using Three Diferent Distance Metrics or Algorithms:

			Dijikstra's Algo			A-Star Algo		
S.N.	**Term 1**	**Term 2**	**T1-T2**	**T2-T1**	**Mean**	**T1-T2**	**T2-T1**	**Mean**
1	Tree	Branch	1	1	1	1	1	1
2	Pen	Blood	3	3	3	3	3	3
3	Paper	Tree	1	1	1	1	1	1
4	Airline	Pen	4(NP)	4	4	4(NP)	4	4
5	Bomb	Blast	2	4(NP)	3	2	4(NP)	3

Concrete Examples of Conceptually and Semantically Unrelated Terms and their Path Length to Compute the Default Value for No-Path:

Term 1	Bowl	Wire	Coffee	Office	Feather	Driver
Term 2	Mobile	Dress	Research	Festival	Study	Sun
Path Length	3	3	3	3	3	3

Solution Pseudo-Code and Algorithm

Concrete Examples of Computing Conceptual Similarity between Two Given Terms Using Three Diferent Distance Metrics or Algorithms:

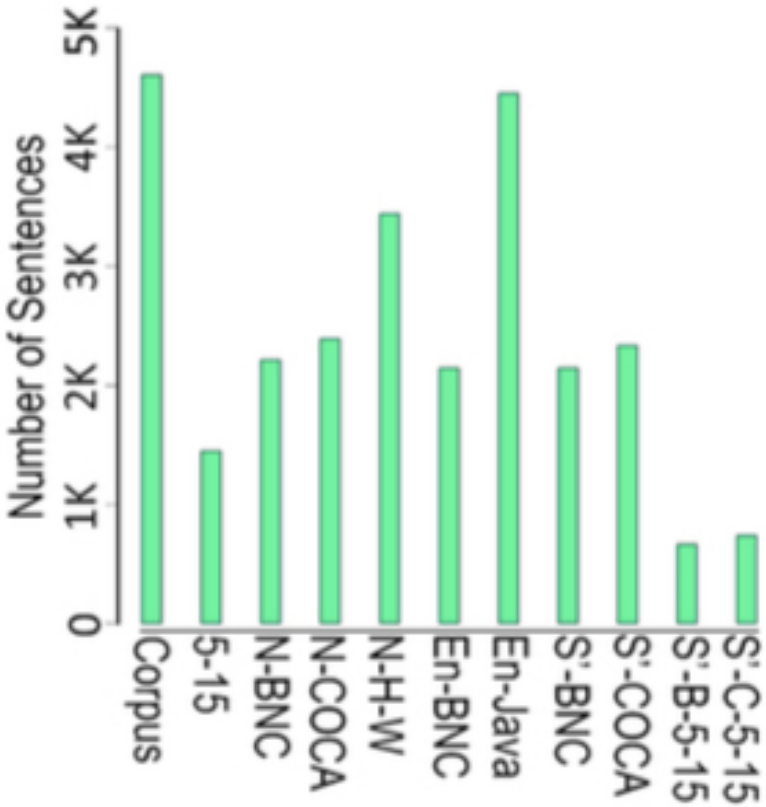
			Dijikstra's Algo			A-Star Algo		
S.N.	**Term 1**	**Term 2**	**T1-T2**	**T2-T1**	**Mean**	**T1-T2**	**T2-T1**	**Mean**
1	Tree	Branch	1	1	1	1	1	1
2	Pen	Blood	3	3	3	3	3	3
3	Paper	Tree	1	1	1	1	1	1
4	Airline	Pen	4(NP)	4	4	4(NP)	4	4
5	Bomb	Blast	2	4(NP)	3	2	4(NP)	3

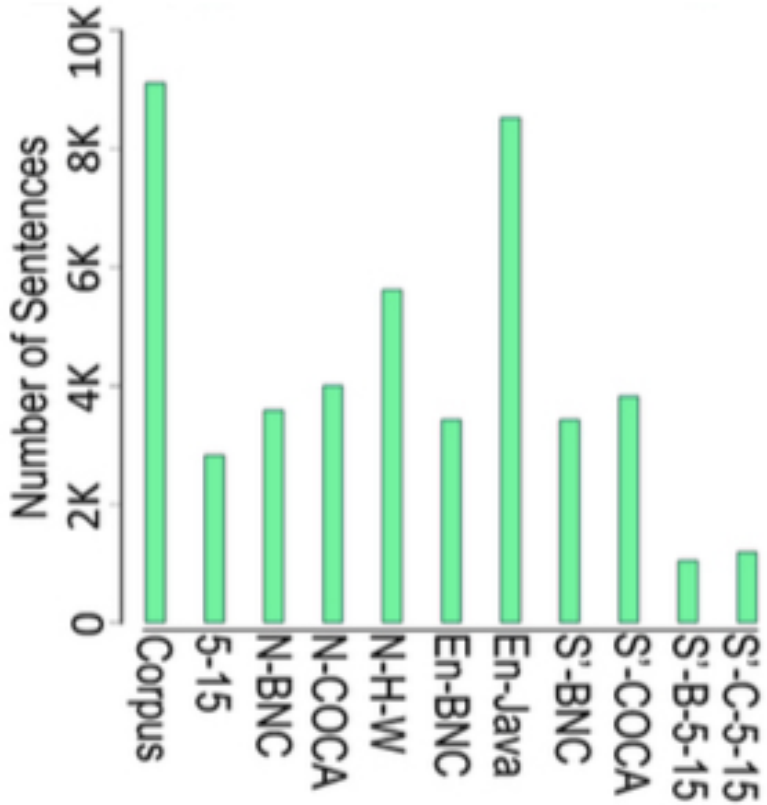
Concrete Examples of Conceptually and Semantically Unrelated Terms and their Path Length to Compute the Default Value for No-Path:

Term 1	Bowl	Wire	Coffee	Office	Feather	Driver
Term 2	Mobile	Dress	Research	Festival	Study	Sun
Path Length	3	3	3	3	3	3

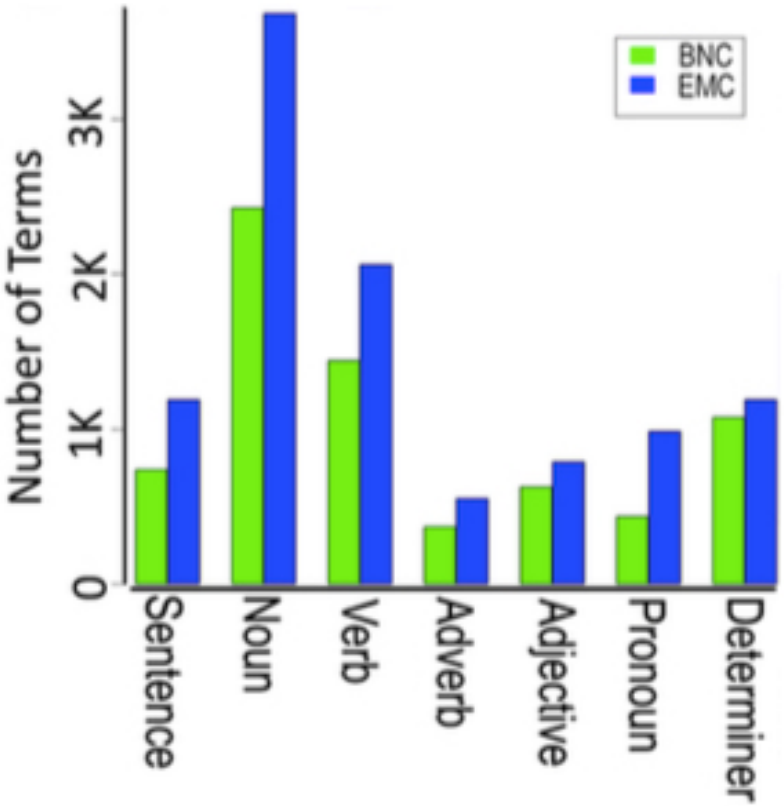
Table of Contents

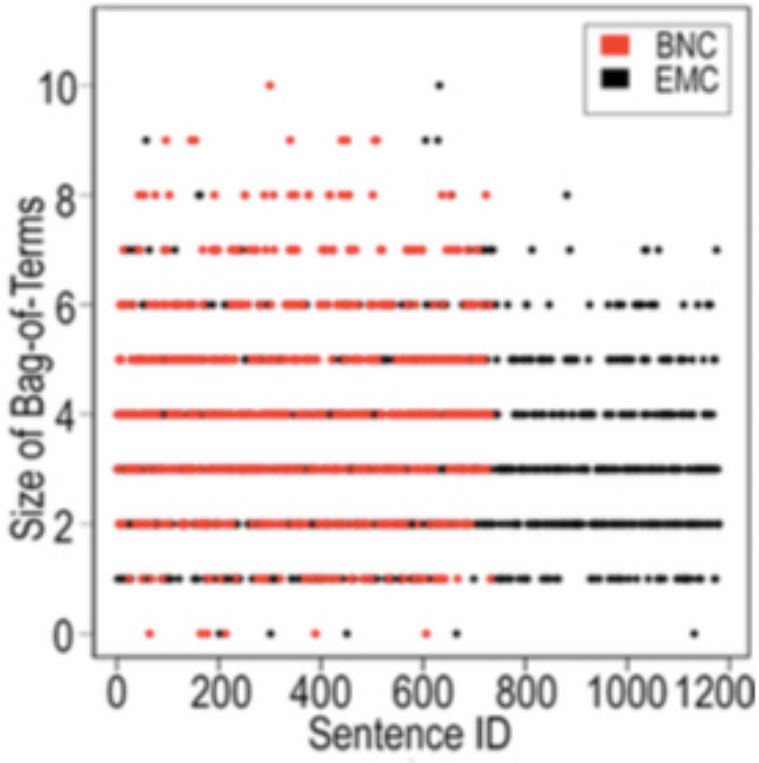
Bar Chart - Experimental Dataset Statistics





Part-of-Speech Tags, Size of Bag-of-Terms





BNC and EMC Dataset

Abbr	Description	BNC	EMC
Corpus	Total sentences in brown news corpus	4607	9112
5-15	Sentences that has length between 5 to 15	1449	2825
N-BNC	Sentences that has their first noun in BNC (british national corpus)	2214	3587
N-COCA	Sentences that has their first noun in 100 K list (COCA)	2393	4006
N-H-W	If first noun has an hypernym in WordNet	3441	5620
En-BNC	English sentences according to BNC	2146	3430
En- Java	English sentences according to Java language detection library	4453	8527
S'-BNC	#Substituted sentences using BNC list	2146	3430
S'-COCA	#Substituted sentences using COCA (100K) list	2335	3823
S'-B-5-15	#Substituted sentences (between length of 5 to 15) using BNC list	666	1051
S'-C-5-15	#Substituted sentences (between length of 5 to 15) using COCA list	740	1191

Data Cleaning

Examples of Sentences Discarded While Word Substitution:

Corpus	Sentence	Reason
EMC	Since we're ending 2000 and going into a new sales year I want to make sure I'm not holding resource open on any accounts which may not or should not be on the list of focus accounts which you and your team have requested our involvement with.	Sentence length is not between 5 to 15
EMC	next Thursday at 7:00 pm Yes yes yes.	First noun is not in BNC/COCA list
BNC	The City Purchasing Department the jury said is lacking in experienced clerical personnel as a result of city personnel policies	Sentence length is not between 5 to 15
BNC	Dr Clark holds an earned Doctor of Education degree from the University of Oklahoma	First noun does not have a hypernym in WordNet

Term Substitution using COCA Frequency List

Example of Term Substitution using COCA Frequency List. NF= First Noun/Original Term, ST= Substituted Term:

Sentence	NF	Freq	ST	Freq	Sentence'
Any opinions expressed herein are solely those of the author.	Author	53195	Television	53263	Any opinions expressed herein are solely those of the television.
What do you think that should help you score women.	Score	17415	Struggle	17429	What do you think that should help you struggle women.
This was the coolest calmest election I ever saw Colquitt Policeman Tom Williams said	Election	40513	Republicans	40515	This was the coolest calmest republicans I ever saw Colquitt Policeman Tom Williams said
The inadequacy of our library system will become critical unless we act vigorously to correct this condition	Inadequacy	831	Inevitability	831	The inevitability of our library system will become critical unless we act vigorously to correct this condition

Text Substitution Technique

Data: Sentence S, Frequency List COCA, WordNet DataBase W_{DB}
Result: Substituted Sentence S'
if $(5 < S.length < 15)$ **then**
    $tokens \leftarrow S.tokenize()$
    $POS \leftarrow S.pos\_tag()$
    $NF \leftarrow token[POS.indexOf(\text{"}NN\text{"})]$
    if $(COCA.has(NF)\ AND\ W_{DB}.has(NF.hypernym))$ **then**
        $lang \leftarrow S.Language\ Detection$
        if $(lanf == \text{"}en\text{"})$ **then**
            $F_{NF} \leftarrow COCA.freq(NF)$
            $F_{NF'} \leftarrow COCA.nextHigherFreq(F_{NF})$
            $NF' \leftarrow COCA.hasFrequency(F_{NF'})$
            $S' \leftarrow S.replaceFirst(NF, NF')$
            $return\ S'$

Examples from Research Papers

	Original Sentence	Substituted Sentence	Paper	Result
1	the bomb is in position	the alcohol is in position	Fong2006 [5]	alcohol
2	copyright 2001 south-west airlines co all rights reserved	toast 2001 southwest airlines co all rights reserved	Fong2006 [5]	southwest
3	please try to maintain the same seat each class	please try to maintain the same play each class	Fong2006 [5]	try
4	we expect that the attack will happen tonight	we expect that the campaign will happen tonight	Fong2008 [4]	campaign
5	an agent will assist you with checked baggage	an vote will assist you with checked baggage	Fong2008 [4]	vote
6	my lunch contained white tuna she ordered a parfait	my package contained white tuna she ordered a parfait	Fong2008 [4]	package
7	please let me know if you have this information	please let me know if you have this men	Fong2008 [4]	know
8	It was one of a series of recommendations by the Texas Research League	It was one of a bank of recommendations by the Texas Research League	Fong2008 [4]	recomm.
9	The remainder of the college requirement would be in general sub jects	The attendance of the college requirement would be in general sub jects	Fong2008 [4]	attendance
10	A copy was released to the press	An object was released to the press	Fong2008 [4]	released

Examples from Research Papers

	Original Sentence	Substituted Sentence	Paper	Result
11	works need to be done in Hydrabad	works need to be done in H	Deshmukh14 [3]	H
12	you should arrange for a preparation of blast	you should arrange for a preparation of daawati	Deshmukh14 [3]	daawati
13	my friend will come to deliver you a pistol	my friend will come to deliver you a CD	Deshmukh14 [3]	CD
14	collect some people for work from Gujarat	collect some people for work from Musa	Deshmukh14 [3]	Musa
15	you will find some bullets in the bag	you will find some pen drives in the bag	Deshmukh14 [3]	pen drives
16	come at Delhi for meeting	come at Sham for meeting	Deshmukh14 [3]	Sham
17	send one person to Bangalore	send one person to Bagu	Deshmukh14 [3]	Bagu
18	Arrange some riffles for next operation	Arrange some DVDs for next operation	Deshmukh14 [3]	DVDs
19	preparation of blast will start in next month	preparation of Daawati work will start in next month	Deshmukh14 [3]	Daawati
20	find one place at Hydrabad for operation	find one place at H for operation	Deshmukh14 [3]	H
21	He remembered sitting on the wall with a cousin, watching the German bomber fly over	He remembered sitting on the wall with a cousin, watching the German dancers fly over	Jabbari08 [10]	German

Brown News Corpus and Enron Email Corpus

Concrete Examples of Sentences with Size of Bag-of-terms Less Than 2:

Corpus	Sentence	Bag-of-terms	Size
BNC	That was before I studied both	[]	0
BNC	The jews had been expected	[jews]	1
BNC	if we are not discriminating in our cars	[car]	1
EMC	What is the benefits?	[benefits]	1
EMC	Who coined the adolescents?	[adolescents]	1
EMC	Can you help? his days is 011 44 207 397 0840 john	[day]	1

Concrete Examples of Sentences with the Presence of Technical Terms and Abbreviations:

Sentence	Tech Terms	Abbr
#4. artifacts 2004-2008 maybe 1 trade a day.	Artifacts	-
We have put the interview on IPTV for your viewing pleasure.	Interview, IPTV	IPTV
Will talk with KGW off name.	-	KGW
We are having males backtesting Larry May's VaR.	backtesting	VAR
Internetworking and today American Express has surfaced.	Internetworking	-
I do not know their particles yet due to the Enron PRC meeting conflicts.	Enron	PRC
The others may have contracts with LNG consistency owners.	-	LNG

Brown News Corpus and Enron Email Corpus

Concrete Examples of Sentences with Size of Bag-of-terms Less Than 2:

Corpus	Sentence	Bag-of-terms	Size
BNC	That was before I studied both	[]	0
BNC	The jews had been expected	[jews]	1
BNC	if we are not discriminating in our cars	[car]	1
EMC	What is the benefits?	[benefits]	1
EMC	Who coined the adolescents?	[adolescents]	1
EMC	Can you help? his days is 011 44 207 397 0840 john	[day]	1

Concrete Examples of Sentences with the Presence of Technical Terms and Abbreviations:

Sentence	Tech Terms	Abbr
#4. artifacts 2004-2008 maybe 1 trade a day.	Artifacts	-
We have put the interview on IPTV for your viewing pleasure.	Interview, IPTV	IPTV
Will talk with KGW off name.	-	KGW
We are having males backtesting Larry May's VaR.	backtesting	VAR
Internetworking and today American Express has surfaced.	Internetworking	-
I do not know their particles yet due to the Enron PRC meeting conflicts.	Enron	PRC
The others may have contracts with LNG consistency owners.	-	LNG

Brown News Corpus and Enron Email Corpus

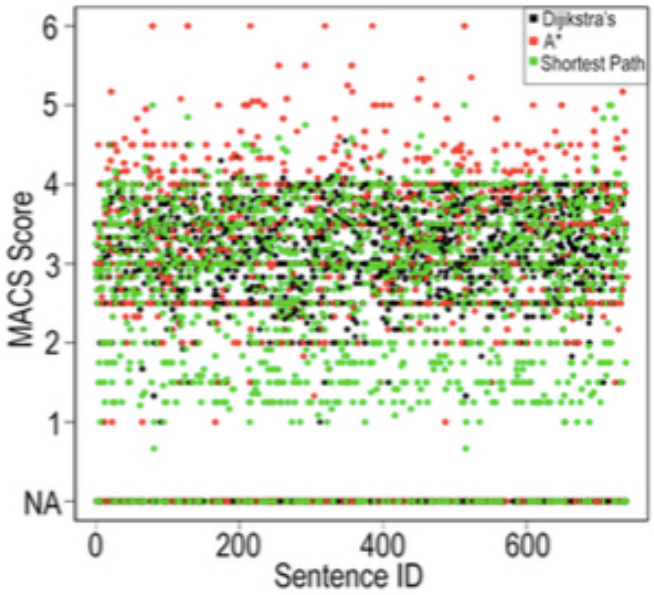
Concrete Examples of Long Sentences (Length of Bag-of-terms >= 5) Where Substituted Term is Identified Correctly:

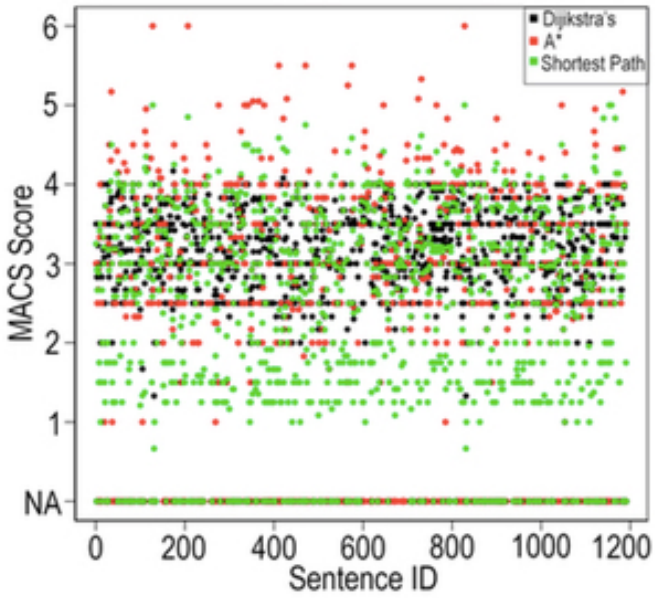
Corpus	Sentence	Original	Bag-of-Terms
BNC	He further proposed grants of an unspecified input for experimental hospitals	Sum	[grants, unspecified, input, experimental, hospitals]
BNC	When the gubernatorial action starts Caldwell is expected to become a campaign coordinator for Byrd	Campaign	[gubernatorial, action, Caldwell, campaign, coordinator, Byrd]
BNC	The entire arguments collection is available to patrons of all members on interlibrary loans	Headquarters	[entire, argument, collection, available, patron, member, interlibrary, loan]
EMC	Methodologies for accurate skill-matching and pilgrims ecien-cies=20 Key Benefits ?	Fulfillment	[methodologies, accurate, skill, pilgrims, eciencies, benefits]
EMC	PERFORMANCE REVIEW The measurement to provide feedback is Friday November 17.	Deadline	[performance, review, measurement, feedback, friday, november]

Brown News Corpus and Enron Email Corpus

Accuracy Results for Brown News Corpus (BNC) and Enron Mail Corpus (EMC):

	Total Sentences	Correctly Identified	Accuracy Results	NA
BNC	740	573	77.4%	46
EMC	1191	629	62.9%	125

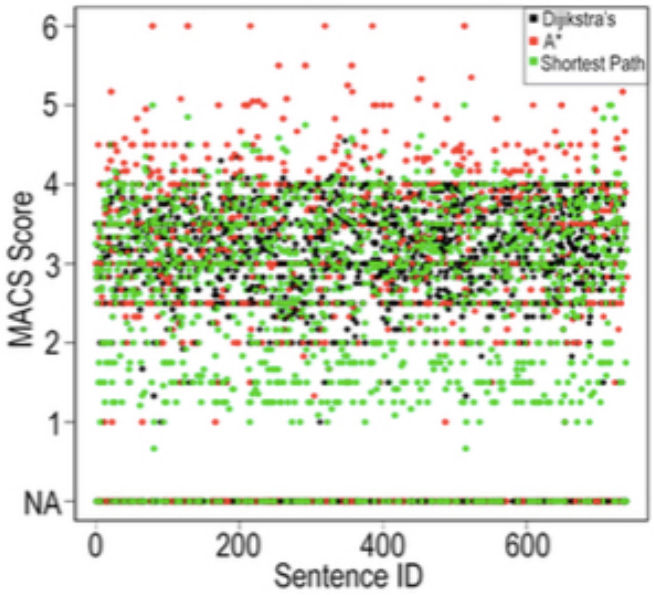


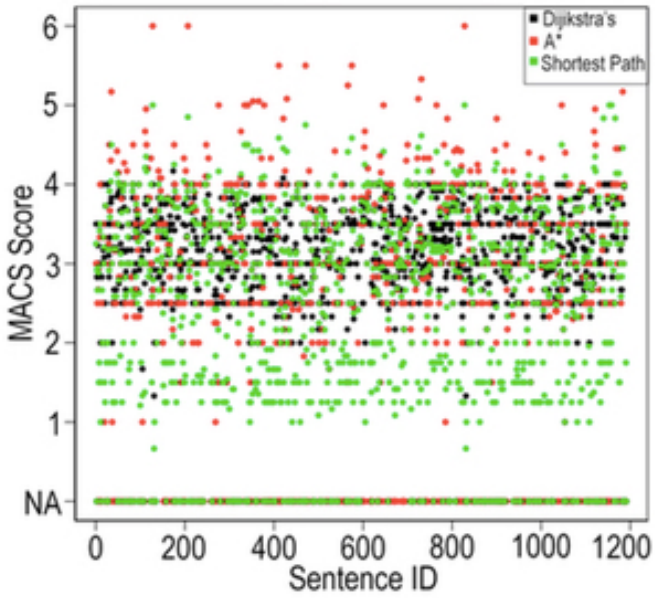


Brown News Corpus and Enron Email Corpus

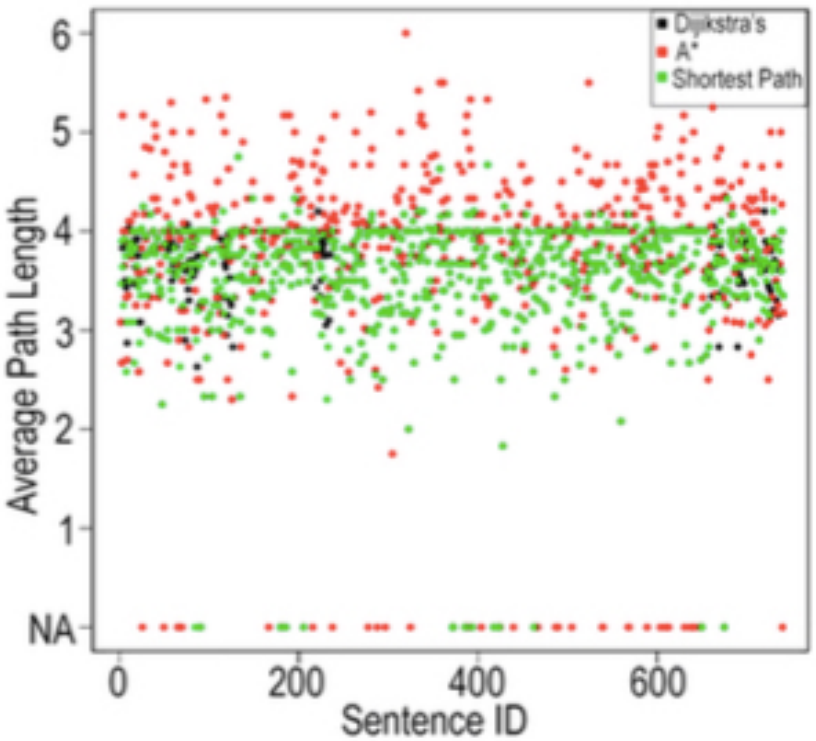
Accuracy Results for Brown News Corpus (BNC) and Enron Mail Corpus (EMC):

	Total Sentences	Correctly Identified	Accuracy Results	NA
BNC	740	573	77.4%	46
EMC	1191	629	62.9%	125





Brown News Corpus and Enron Email Corpus



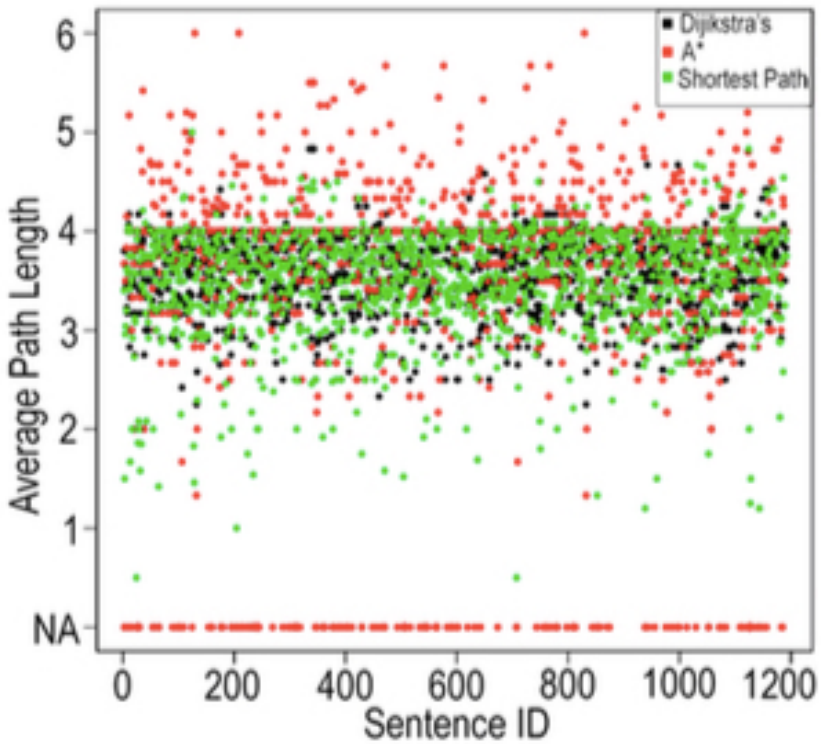


Table of Contents

Conclusion

- We present an approach to detect term obfuscation in adversarial communication using ConceptNet common-sense knowledge-base
- Experimental results reveal an accuracy of 72.72%, 77.4% and 62.0% respectively on the three dataset
- Experimental results demonstrate that our approach is able to detect term obfuscation in long sentences containing more than 5 - 6 concepts
- We demonstarate that the proposed approach is generalizable

Conclusion

- We present an approach to detect term obfuscation in adversarial communication using ConceptNet common-sense knowledge-base
- Experimental results reveal an accuracy of 72.72%, 77.4% and 62.0% respectively on the three dataset
- Experimental results demonstrate that our approach is able to detect term obfuscation in long sentences containing more than 5 - 6 concepts
- We demonstarate that the proposed approach is generalizable

Conclusion

- We present an approach to detect term obfuscation in adversarial communication using ConceptNet common-sense knowledge-base
- Experimental results reveal an accuracy of 72.72%, 77.4% and 62.0% respectively on the three dataset
- Experimental results demonstrate that our approach is able to detect term obfuscation in long sentences containing more than 5 - 6 concepts
- We demonstarate that the proposed approach is generalizable

Conclusion

- We present an approach to detect term obfuscation in adversarial communication using ConceptNet common-sense knowledge-base
- Experimental results reveal an accuracy of 72.72%, 77.4% and 62.0% respectively on the three dataset
- Experimental results demonstrate that our approach is able to detect term obfuscation in long sentences containing more than 5 - 6 concepts
- We demonstarate that the proposed approach is generalizable

Table of Contents

References I


Sonal N. Deshmukh, Ratnadeep R. Deshmukh, and Sachin N. Deshmukh.
Performance analysis of different sentence oddity measures applied on google and google news repository for detection of substitution.
International Refereed Journal of Engineering and Science (IRJES), 3(3):20–25, 2014.


Ben Allison Sanaz Jabbari and Louise Guthrie.
Using a probabilistic model of context to detect word obfuscation.
Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08), 2008.


SW. Fong, D. Roussinov, and D.B. Skillicorn.
Detecting word substitutions in text.
IEEE Transactions on Knowledge and Data Engineering, 20(8):1067–1076, 2008.


SW Fong, DB Skillicorn, and D Roussinov.
Detecting word substitution in adversarial communication.
In *6th SIAM Conference on Data Mining. Bethesda, Maryland*, 2006.

References II


Chi-En Wu and Richard Tzong-Han Tsai.
Using relation selection to improve value propagation in a conceptnet-based sentiment dictionary.
Knowledge-Based Systems, 2014.


Arbi Bouchoucha, Xiaohua Liu, and Jian-Yun Nie.
Integrating multiple resources for diversified query expansion.
Advances in Information Retrieval, pages 437–442, 2014.


R Akileshwari, S Revathi, and A Grace Selvarani.
A novel approach for similarity based video annotation utilizing commonsense knowledgebase.


Soujanya Poria, Basant Agarwal, Alexander Gelbukh, Amir Hussain, and Newton Howard.
Dependency-based semantic parsing for concept-level text analysis.
Computational Linguistics and Intelligent Text Processing, pages 113–127, 2014.

References III


Hugo Liu and Push Singh.
Conceptnet a practical commonsense reasoning tool-kit.
BT technology journal, 22(4):211–226, 2004.


Catherine Havasi, Robert Speer, and Jason Alonso.
Conceptnet 3: a flexible, multilingual semantic network for common sense knowledge.
In *Recent Advances in Natural Language Processing*, pages 27–29, 2007.


Robert Speer and Catherine Havasi.
Conceptnet 5: A large semantic network for relational knowledge.
In *The People s Web Meets NLP*, pages 161–176. Springer, 2013.


Dmitri Roussinov, SzeWang Fong, and David Skillicorn.
Detecting word substitutions: Pmi vs. hmm.
In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '07, pages 885–886, 2007.

References IV


Nakatani Shuyo.
Language detection library for java, 2010.

Thank you!

Contact:
swatia@iiitd.ac.in
ashish@iiitd.ac.in