NLP Technologies Against Cyber Crime

Dr. Shantipriya Parida

Agenda

- Cyber Crime
- Natural Language Processing
- NLP in Cyber Crime
  - Topic Modelling
  - Text Classification
  - Named Entity Recognition
  - Authorship Attribution
  - Machine Translation
  - Fake News Detection
- Fusion of Technologies
- Conclusion

What is Cyber Crime ?

- Any illegal activity carried out using computers or the internet.
- Types of Cyber Crimes:
  - Hacking
  - Child Sexual Abuse and Grooming
  - Phishing
  - Illegal Drug and Weapons Trafficking
  - Human Trafficking and Sex Crimes

CSA

- Child sexual abuse (CSA) and grooming[1] are a real threat for children and young people of all ages and backgrounds.
- Sexual Predators[2], attract victims that either have no sense of who they interact with neither the scope of interaction.
- The effects from "Child grooming" could cause psychological damages to the children that may never cure.

Image source: https://violenceagainstchildren.un.org/news/european-union-must-close-legal-gap-protect-children-sexual-abuse-online-un-experts

[1] *Grooming means forming a relationship or connection with a child for the purpose of sexually abusing them. (Source: https://slideplayer.com/slide/5719384/)*

[2] *Person seen as obtaining or trying to obtain sexual contact with another person in a metaphorically "predatory" manner"* (Wikipedia)

Grooming Stages

- Friendship — Asking Profile, Asking picture, Giving compliment
- Forming a Relationship — Talking about activity, favourite, hobby, school
- Risk assessment — Asking questions to know risk of conversation, Asking if the child is alone or under adult or friend supervision
- Exclusivity — Trying to build mutual trust, Using falling in love words, Using word to express feeling
- Sex talk — Using word about biology, body, intimate parts, and sexual category, Asking hot picture, Arranging further contact and meetings

Source: https://www.theguardian.com/technology/2003/jul/17/childprotection.childrensservices

Challenges

- Parents or law enforcements agents can’t watch over the children all the time.
- An officer has to read hundreds or thousands of pages of chat-texts in order to provide accusative evidence.
- Thus it is prone to error or may lead to a biased decision.
- Currently, there is no system that can automatically identify the elements of child exploitation on text chats (on-line, off-line).
- A difficult task for any single agency, authority, ministry, or NGO, or company to tackle this problem.
- It require strong cooperation to fight against the same.
- One of the key challenges involve is the lack of data for building tools/technologies.

Challenges

- Interpol report[1] highlights impacts of COVID19 on child sexual abuse.
  - Social and economic factors impacted child sexual exploitation and abuse (CSEA) across the world.
  - Closure of schools, move to virtual learning, online time spent (entertainment, social)

Image source: https://www.europol.europa.eu/covid-19/covid-19-child-sexual-exploitation

[1]https://www.interpol.int/en/News-and-Events/News/2020/INTERPOL-report-highlights-impact-of-COVID-19-on-child-sexual-abuse

Challenges

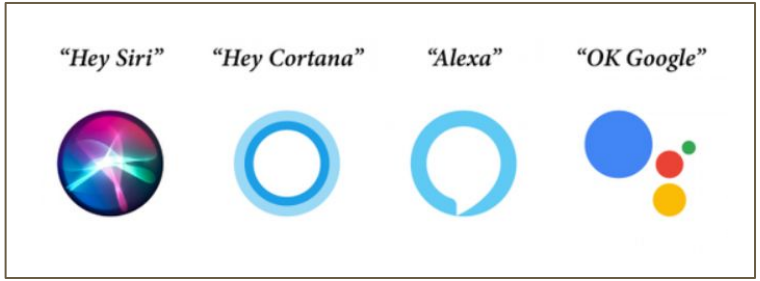
- Roxanne

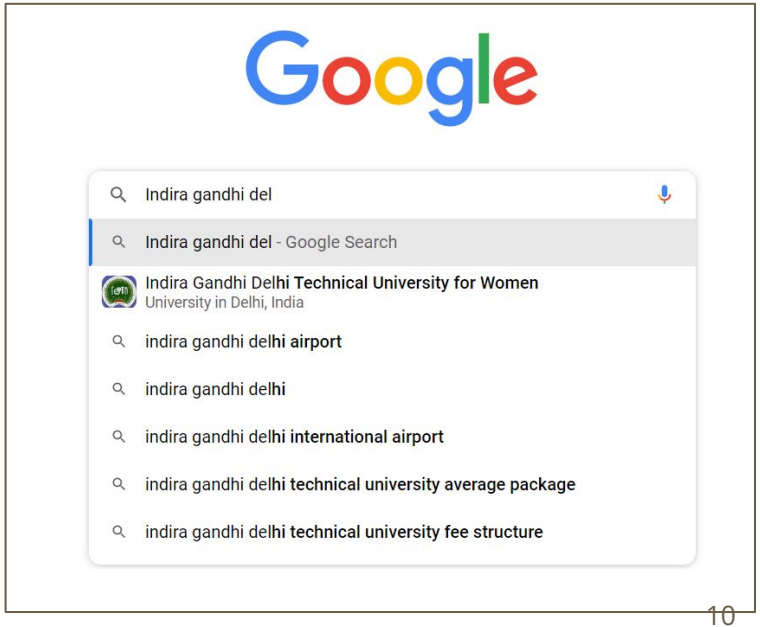
What is NLP

- Natural language processing (NLP) helps computers communicate with humans in their own language and scales other language-related tasks.
- NLP makes it possible for computers to read text, hear speech, interpret it, measure sentiment and determine which parts are important.

NLP in Daily Life

- Personal assistants: Siri, Cortana, and Google Assistant.
- Auto-complete: In search engines (*e.g.* Google).
- Spell checking: Almost everywhere, in your browser, your IDE (*e.g.* Visual Studio), desktop apps (*e.g.* Microsoft Word).
- Machine Translation: Google Translate.
- Chat bots.





NLP in Cyber Crime

- The web and social media applications and platforms are an overall complex and multidimensional data landscape where NLP can be used.
- Criminal investigations require manual intervention of several investigators and translators.
- NLP techniques help criminal investigators handle large amounts of textual information in a more efficient and faster way.
- Commonly used techniques:
  - Optical Character Recognition (OCR)
  - Machine Translations
  - Text Summarization

Topic Modelling (1)

- What is Topic Modelling ?.
  - Topic modeling is a statistical modeling approach to discover the abstract “topics” occurs in a collection of documents.
- Types of Topic Modeling
  - Unsupervised, and Semi-supervised
- Application of Topic Modeling
  - text mining, text classification, machine learning, information retrieval, and recommendation engines.

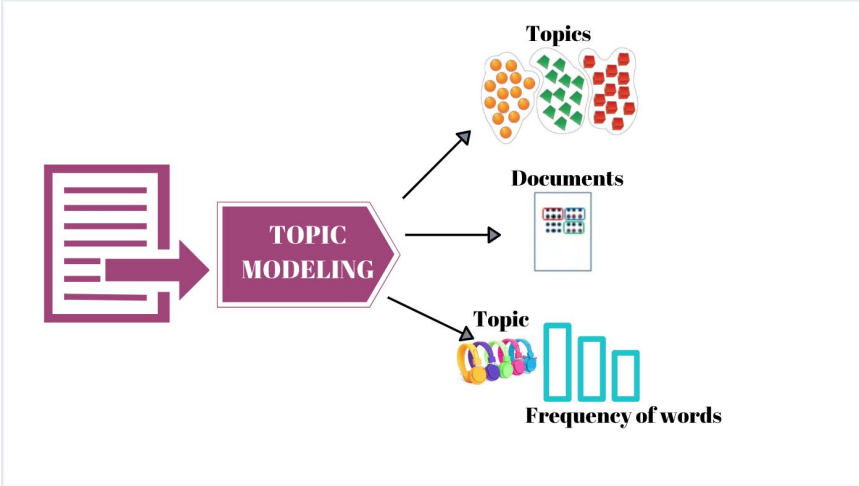


Fig: Topic Modeling

Topic Modeling (2)

- Extracting topics from a large scale grooming dataset.
- Data source : Liveme streaming platform (https://www.liveme.com)
- Number of messages : 39.38 Million
- Number of user : 1.42 Million
- Short messages in multiple languages (e.g. chinese, russian, indonesian, japanese, German) contain text, emojis, numbers
- Messages contain remarks, questions

Fig: LiveMe

Fig:Top collocates of sexual words in LiveMe dataset (Lykousas et al.)

Topic Modeling (3)

- Topic extraction using Latent Dirichlet Allocation (LDA) method

Topic	Keywords	#Docs
18	CLOTH_TERM, show, open, SEX_TERM, nice, dare, dance, hot, stand, leg, put, kiss, turn, pull, wear, cam, camera, remove, foot, top, snapchat, gift, girl, rub, low, hand, lift, finger, message, tease	12,209
11	love, nice, pretty, kiss, cute, eye, girl, gorgeous, hot, SEX_TERM, sweet, lip, hair, dear, tattoo, smile, single, dance, friend, number, stand, beauty, lovely, cutie, face, boyfriend	8,477
1	sleep, phone, bed, tired, cool, car, wake, cold, drive, smoke, hour, fall, hear, talk, asleep, stay, high, long, house, game, guess, chill, goodnight, sound, money, iphone, fun, pay	5,731
19	talk, hear, happen, friend, leave, wrong, true, cool, mad, sad, care, sound, hurt, smile, fight, stay, fine, dude, person, funny, break, hard, nice, head, long, boy, army, problem, lose, girl, yep	4,432
7	block, admin, girl, message, leave, report, show, talk, account, kid, creep, young, shut, ban, rude, fake, nasty, send, perv, truth, boy, lie, hater, wrong, police, child, unblock	3,328
16	drink, cat, food, pizza, laugh, eat, funny, face, water, put, chicken, dead, hair, head, challenge, roast, cream, leave, SEX_TERM, apple, taco, chocolate, pet, bob, hand, candy, mouth, cheese, nose	3,180
12	cute, snapchat, send, instagram, rate, clown, dab, hot, number, insta, hair, play, text, friend, pretty, love, put, single, phone, kik, profile, eye, cutie, chat, ghost, boy, girl, fake, girlfriend	3,080
3	send, gift, spam, castle, diamond, share, top, level, broadcast, win, giveaway, broadcaster, wand, number, stream, boat, enter, entry, star, love, feature, join, porsche, coin, awesome, comment, fan	2,953
5	coin, drop, coindrop, follower, send, win, feature, shout, fan, castle, love, wand, dab, thot, number, gift, shoutout, giveaway, diamond, stream, iphone, lag, goal, dude, pumpkin, andy, light, level	2,798
4	song, play, sing, love, voice, rap, singing, amazing, nice, dance, awesome, beat, put, hear, panda, listen, cool, singer, sound, closer, juju, black, job, girl, talent, guitar, boy, heart, hit, drake	2,687
8	love, stream, friend, accent, talk, remember, guess, cool, speak, leave, sleep, skype, long, cute, funny, nice, number, meet, hair, mate, lot, person, dad, class, cat, joke, jenni, kat, join, change	2,682
20	light, turn, gang, love, stay, queen, squad, hit, chill, number, king, slay, fact, thot, level, savage, rock, party, dead, boy, mad, play, homie, ight, lot, black, nun, show, petty, dope, top, sum	2,366
2	hola, mami, como, cute, hermosa, eres, show, spanish, amor, SEX_TERM, pretty, bella, donde, hot, bonita, bien, lip, kiss, speak, tienes, espanol, gorgeous, stand, rico, jada	2,142
13	girl, love, cute, play, blue, twin, pretty, red, hot, black, dance, snapchat, green, pink, makeup, hair, lady, white, friend, cool, color, game, face, team, texas, nice, CLOTH_TERM, batman, favorite	1,954
14	kate, love, kid, nice, awesome, cool, tree, country, santa, dad, boy, level, show, broadcast, send, amazing, hear, wolf, talk, lot, son, king, falcon, grim, happen, stream, matt, house, long, rock	1,831
9	beam, love, lag, send, king, cris, stream, castle, fletch, broadcast, show, level, dude, awesome, nick, game, amazing, feature, remember, joey, gift, beem, roll, diamond, join, happen, rip, rackbar	1,663
15	ready, love, spam, feature, stay, game, number, win, boy, tre, read, letter, chat, duck, turtle, greg, cat, spamme, fun, red, ugh, play, controller, send, coin, hehe, cool, high, comment, gift, party	1,027
17	love, fan, favorite, youtube, shout, meet, dab, channel, pickle, song, shoutout, canada, movie, awesome, fav, twerk, vote, magic, food, subscribe, notice, tattoo, cool, texas, win, vid, hair, ily	908
6	race, love, family, human, unity, amen, put, country, draw, earth, whiskey, broadcast, peace, block, thre, lucky, princess, spam, britt, join, general, respect, coin, barbie, send, level, lag, brit	642
10	president, kira, criticize, article, essay, literary, loco, fard, natur, fward, lag, foard, riot, ward, folard, kilo, follrd	14

Fig: Sample Topic from "Large-scale analysis of grooming in modern social networks" dataset

Text Classification (1)

- Text classification is a task of NLP where the model needs to predict the classes of the text documents.
- In the traditional process, we are required to use a huge amount of labelled data to train the model, and also they can't predict using the unseen data.
- Adding zero-shot learning with text classification has taken NLP to the extreme.
- Zero-shot text classification technique classify the text documents without using any single labelled data or without having seen any labelled text.

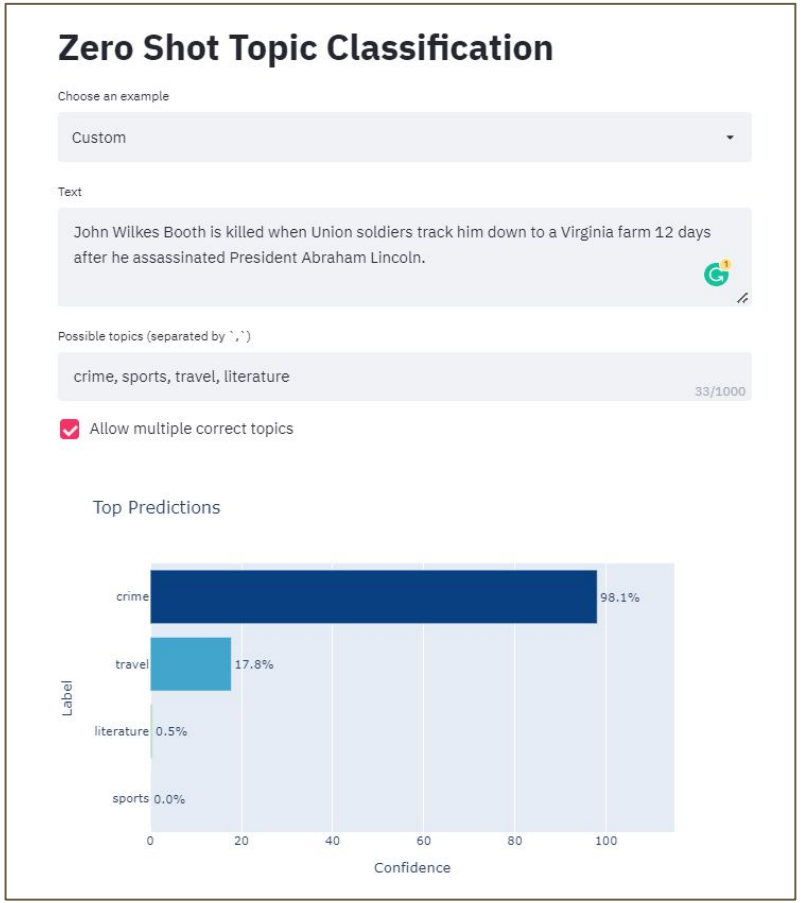


Fig: Zero-shot Text Classification

Text Classification (2)

Use Case: Predator Detection

- Most research in detection of cyber-predators is based on the chat log transcripts provided by Perverted Justice.
- A group of volunteers who posed as children in chatrooms in order to lure predators.
- Chats with predators have been transcribed, anonymized and made available to the public.

Image source: http://www.perverted-justice.com/

Text Classification (3)

- Proposed method for detection of misbehaving users in chats is based on two main hypotheses
  - i) Suspicious Conversations Identification (SCI) stage
  - ii) Victim From Predator disclosure (VFP) stage

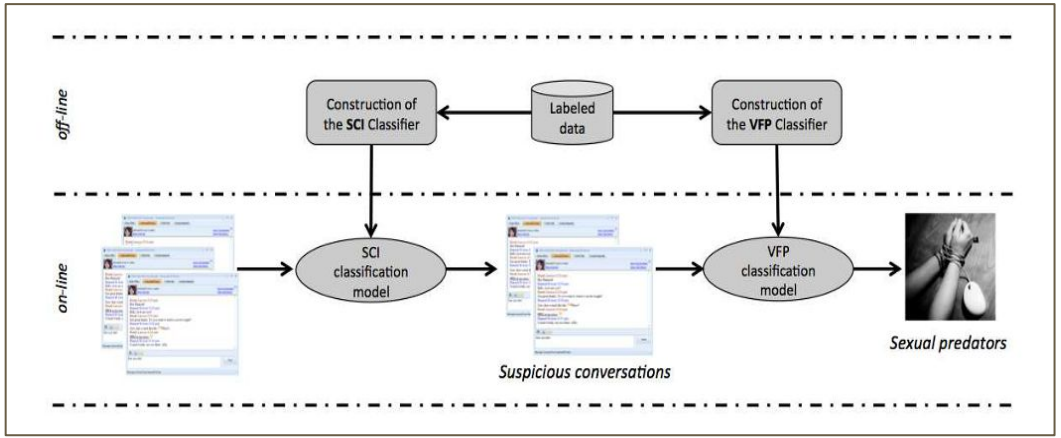
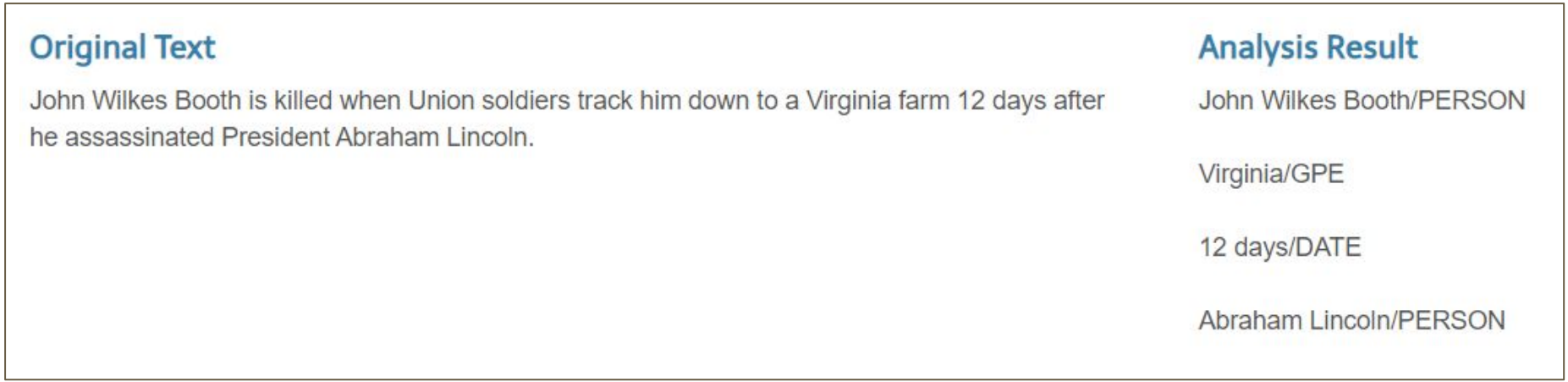


Fig: General overview of the proposed sexual predators identification system (Villatoro-Tello et Al)

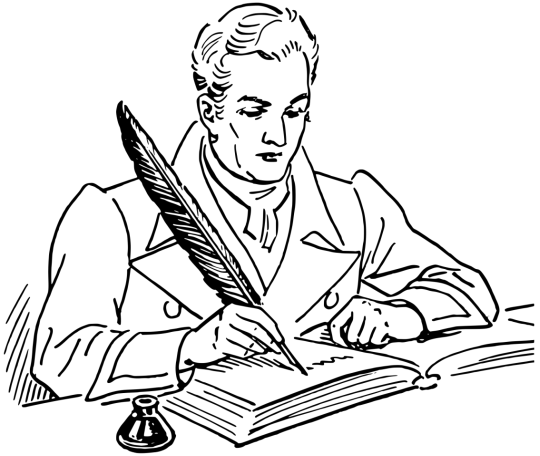
Named Entity Recognition

- Named entity recognition is a natural language processing technique that can automatically scan entire articles and pull out some fundamental entities in a text and classify them into predefined categories.
- Entities can be: Organizations, Quantities, Monetary values, Percentages, People's names, Company names, Geographic locations (Both physical and political), Product names, Dates and times, Amounts of money,Names of events.



Authorship Attribution (1)

- Authorship identification is a process in which the author of a text is identified.
- Internet technologies and online social network, web services (e.g., emails, blogs, forums and micro-blogs) become a means by which new ideas and information spread rapidly.
- People on the virtual space do not need to provide their real identities, accurate automatic authorship attribution of anonymous documents is increasingly requisite
- Authorship attribution techniques can assist law enforcement to discover criminals who supply false information in their virtual identities, and collect digital evidence for cybercrime investigation.



Authorship Attribution (2)

- Authorship attribution approaches fall into two major categories:
  - profile based approaches
  - instance-based approaches
- The basic assumption of forensic linguistics is the notion of idiolect.
-
- Users show individual linguistic features which make them recognisable by their usage of language.
- More recently, deep learning-based approaches have been explored for Authorship Attribution (AA).

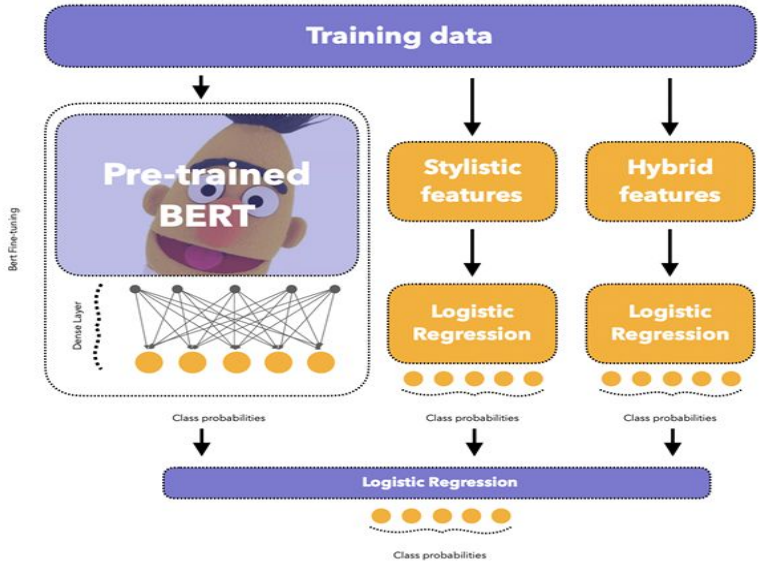
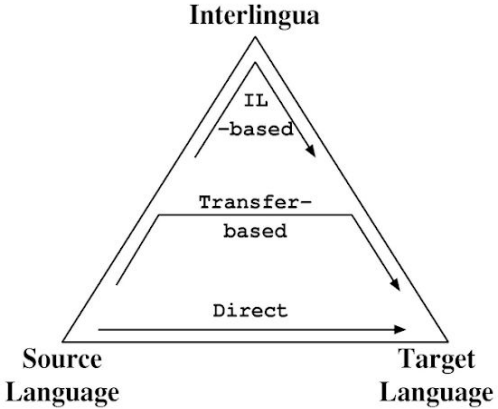


Fig: BertAA + Style + Hybrid architecture (Mael et al.)

Machine Translation (1)

- Automatic conversion of text/speech from one natural language to another.

- Machine translation approaches:
  - Grammar-based
    - Interlingua-based
    - Transfer-based
- Direct
  - Example-based
  - Statistical
  - Neural



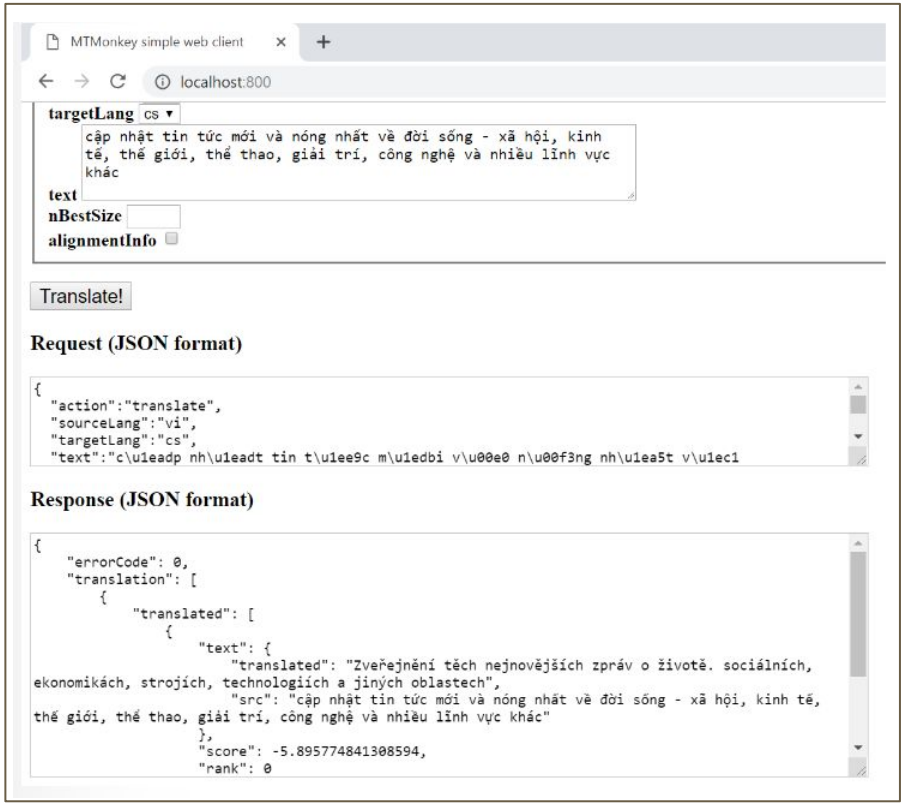
Google Translate

SYSTRAN
beyond language

Machine Translation (2)

MTViet

- Machine Translation from Vietnamese into Czech for the Purposes of the Police of the Czech Republic
- The goal of the project was to develop an off-line machine translation system from Vietnamese into Czech that could be used by the Police of the Czech Republic internally.
- Czech police can automatically translate confidential materials without any concerns about compromising the security of the data.
- The system was optimized particularly for short text messages (SMS) which can often lack diacritic marks and suffer from other types of noise.



Fake News Detection (1)

- Fake news provides information that aims to manipulate people for different purposes: terrorism, political elections, advertisement, satire, among others
- In social networks, misinformation extends in seconds among thousands of people
- A fake news detection system aims to help users detect and filter out potentially deceptive news

Fake News Detection (2)

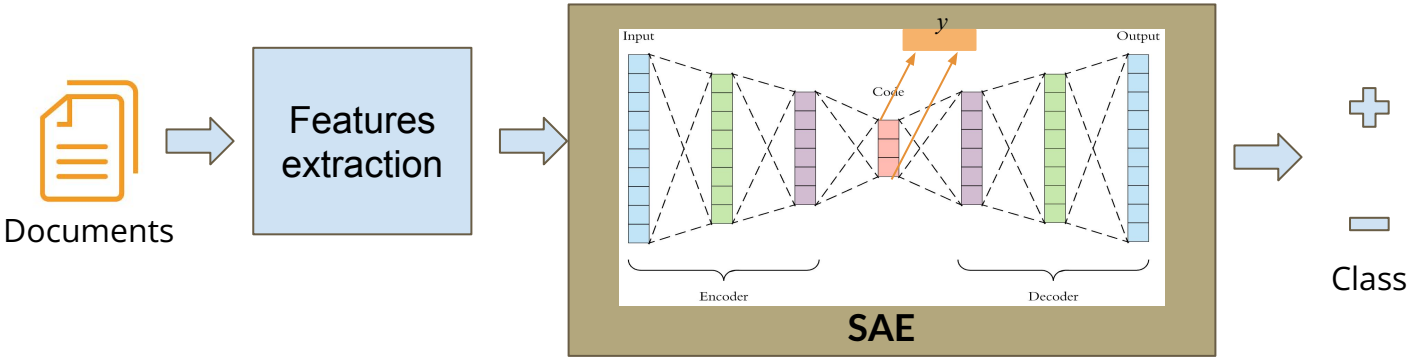
Shared tasks at MexAT3

- Fake news detection
  - Information that aims to manipulate people for different purposes
  - 971 documents, 676 for training and 295 for test
- Aggressiveness detection
  - Accurately identifying significant threats to users who are exposed in social media domain
  - 10,475 anonymised tweets, 7332 for training 3143 for testing
- Both datasets are in Mexican Spanish

Fake News Detection (3)

Methodology

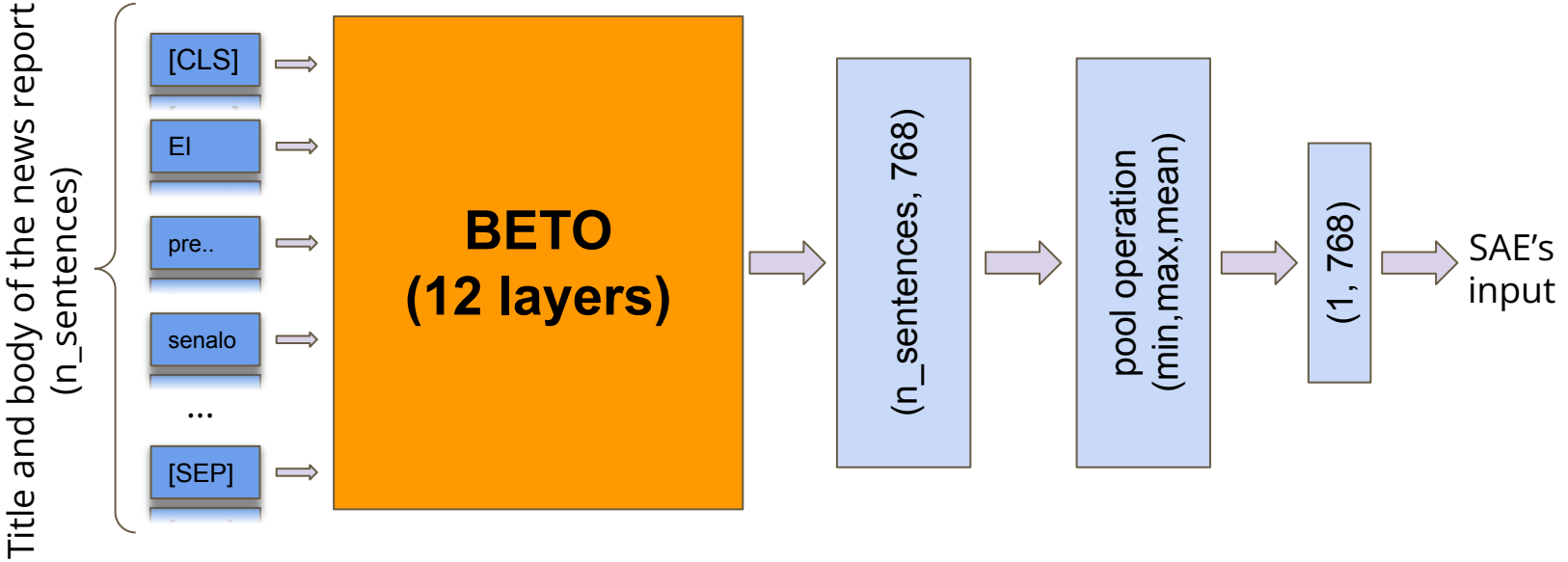
- Our goal was to evaluate the pertinence of deep SAE in these tasks
- As input features we used:
  - Spanish pre-trained BERT encodings (BETO[2])
  - Traditional text representation techniques such as word and char n-grams (ranges 1-2 and 1-3)
  - Combinations of BETO encodings plus traditional words/char n-grams vectors



[2] Cañete, J., Chaperon, G., Fuentes, R., Ho, J. H., Kang, H., & Pérez, J. (2020). Spanish pre-trained bert model and evaluation data. In *Practical ML for Developing Countries Workshop@ ICLR 2020*.

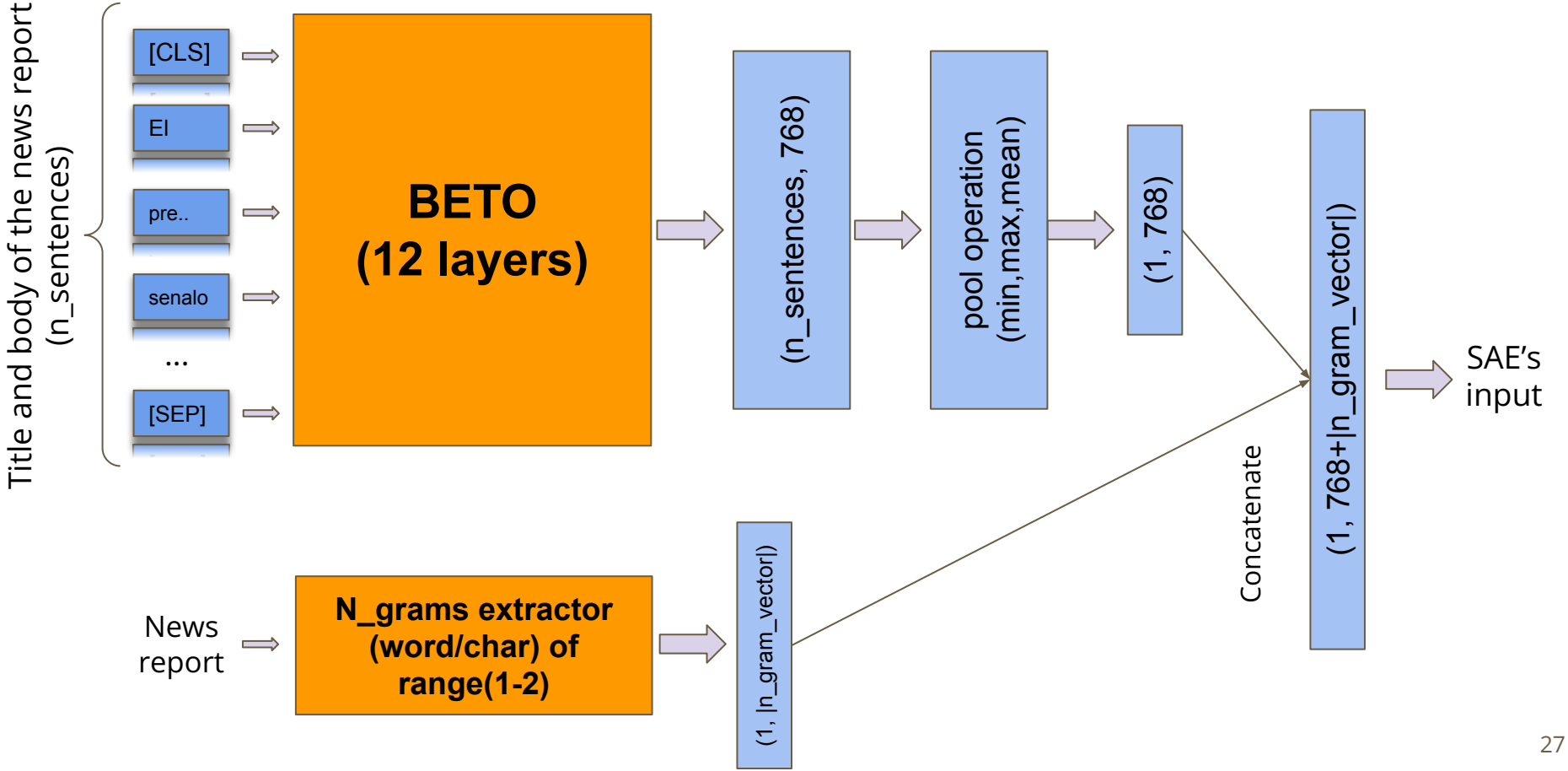
Fake News Detection (4)

Methodology (Features extraction)



Fake News Detection (5)

Methodology (Features extraction)



Results (fake news task)

Table 3. Results in validation and test phases reported in F-score for fake-news (F+), real-news (F-), and macro average of F-score (Fm).

		Validation phase			*Test phase*			
Input features	`min-df,max-df`	Fm	F+	F-	ID	Fm	F+	F-
W(1,2)	0.01, 0.5	0.775	0.793	0.758	-	-	-	-
W(1,3)	0.01, 0.5	0.778	0.798	0.758	-	-	-	-
C(1,2)	0.01, 0.5	0.697	0.719	0.674	-	-	-	-
C(1, 3)	0.01, 0.5	0.757	0.768	0.745	-	-	-	-
B(min-pooling)		**0.843**	0.842	0.845	2	**0.856**	**0.844**	**0.868**
B(max-pooling)		0.830	0.830	0.830	-	-	-	-
B(mean-pooling)		0.833	0.831	0.835	-	-	-	-
C(1, 3)+W(1,2)	0.01, 0.5	0.805	0.807	0.802	-	-	-	-
B+W(1,2)	0.01, 0.3	**0.845**	0.846	0.844	1	0.850	0.840	0.859
B+C(1,3)	0.01, 0.3	0.834	0.834	0.835	-	-	-	-
B+W(1,2)+C(1,3)	0.01, 0.3	0.833	0.831	0.835	-	-	-	-
B+W(1,2)+C(1,3)	0.01, 0.5	**0.848**	0.846	0.850	-	-	-	-
Third best system (in the track)						0.817	0.819	0.817
BOW-RF (baseline-given by track organizers)						0.786	0.785	0.787

Fusion of Technologies

- Automating the process of criminal investigations to speed it up can have a large impact on the daily work of police practitioners.
- It remains a complex task, because of the multimodality of data (e.g. intercepted telephone calls, text mostly, but also CCTVs), the plurality of languages to handle, and the lack of realistic training data matching this domain.
- The integrated platform processes intercepted phone calls, runs state-of-the-art components such as speaker identification, automatic speech recognition or named entity detection, and builds a knowledge graph of the extracted information.
- AutoCrime

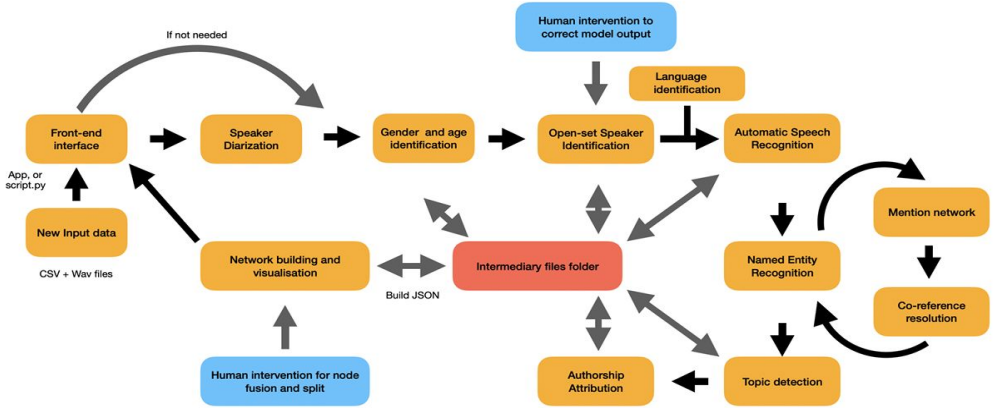


Fig: AutoCrime Data processing in the platform (Mael et al.)

Conclusion

- NLP technologies helps to handle large amount of data and able to extract meaningful information to speedup crime investigation.
- The fusion of technologies (NLP, Speech, Video, Network Analysis) new research direction in crime investigation and able to generate state-of-the-art result.

References

[1] Yang, M., & Chow, K. P. (2014, June). Authorship attribution for forensic investigation with thousands of authors. In *IFIP International Information Security Conference* (pp. 339-350). Springer, Berlin, Heidelberg.

[2] Belvisi, N. M. S., Muhammad, N., & Alonso-Fernandez, F. (2020, April). Forensic authorship analysis of microblogging texts using n-grams and stylometric features. In 2020 8th International Workshop on Biometrics and Forensics (IWBF) (pp. 1-6). IEEE.

[3] Maël Fabien, Esau Villatoro-Tello, Petr Motlicek, and Shantipriya Parida. 2020. BertAA : BERT fine-tuning for Authorship Attribution. In Proceedings of the 17th International Conference on Natural Language Processing (ICON), pages 127–137, Indian Institute of Technology Patna, Patna, India. NLP Association of India (NLPAI).

[4] Banerveld, M. V., Kechadi, M. T., & Le-Khac, N. A. (2016). A natural language processing tool for white collar crime investigation. In Transactions on Large-Scale Data-and Knowledge-Centered Systems XXIII (pp. 1-22). Springer, Berlin, Heidelberg.

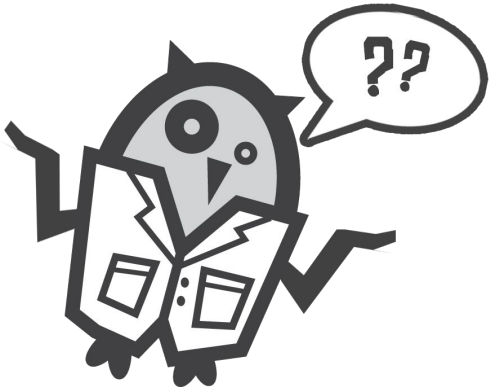
[5] Sohrabi, B., Vanani, I. R., & Shineh, M. B. (2018). Topic modeling and classification of cyberspace papers using text mining. Journal of Cyberspace Studies, 2(1), 103-125.

[6] Lykousas, N., & Patsakis, C. (2021). Large-scale analysis of grooming in modern social networks. Expert Systems with Applications, 176, 114808.

[7] Villatoro-Tello, E., Juárez-González, A., Escalante, H. J., Montes-y-Gómez, M., & Pineda, L. V. (2012, September). A Two-step Approach for Effective Detection of Misbehaving Users in Chats. In CLEF (Online Working Notes/Labs/Workshop) (Vol. 1178).

[8] Fabien, M., Parida, S., Motlícek, P., Zhu, D., Krishnan, A., & Nguyen, H. H. (2021). ROXANNE Research Platform: Automate Criminal Investigations. In Interspeech (pp. 962-964).

Q&A



Contact information:

- Twitter: @Shantipriyapar3
- Web : shantipriya.me

Thank you