in those topics. For instance, in Topic 3 and Topic 6, physical evidences such as bullet, DNA, and saliva that are likely found at the scene are identified and clustered, which can help CIP officials identify the appearance of physical evidence. Semantic NMF with keyword highlighting results can show CIP officials what details they can expect from the collection of documents such as different crime types or evidence, which can potentially accelerate the decision-making. Comparing results from the two semantic NMF methods, keyword highlighting leads to more interpretable and demonstrably better topics for the AOB data set.

Topic 1	**Topic 2**	**Topic 3**	**Topic 4**	**Topic 5**	**Topic 6**	**Topic 7**	**Topic 8**	**Topic 9**	**Topic 10**
appellant	stopped	involving	going	primary	help	instruct	improper	member	transcript
people	identified	object	imprisonment	police	drove	plaintiff	perpetrator	anything	without
evidence	violent	minutes	imposed	photographs	verdicts	including	self	established	regarding
trial	past	mind	daughter	photo	however	imposed	including	instant	inadmissible
jury	introduced	propensity	passenger	never	including	might	actions	separate	proof
would	turned	plus	brief	next	asked	specifically	holding	establish	word
supra	arrived	generally	plaintiff	caliber	error	notice	accused	failure	respondent
murder	estrada	identity	acts	position	noted	asked	thus	verdict	present
also	gang	whether	timely	committing	certain	dated	honorable	discretion	plus
years	legal	come	answer	arguments	ibid	related	included	provides	victim

Table (9) The top 10 keywords learned by semantic NMF on the AOBs from both categories

Topic 1	**Topic 2**	**Topic 3**	**Topic 4**	**Topic 5**	**Topic 6**	**Topic 7**	**Topic 8**	**Topic 9**	**Topic 10**
appellant	snitch	caliber	semen	executed	saliva	ligature	notice	mixture	submitted
people	informant	fired	period	follows	sperm	around	timely	dna	certificate
evidence	lineup	bullet	caused	fully	dna	agreed	filed	least	authorities
trial	provides	recovered	might	addressed	inside	allegation	january	consistent	word
jury	shown	shots	exhibit	clerk	object	shot	amended	excluded	appealability
supra	photo	body	envelope	perjury	get	presented	february	blood	dated
would	simply	scene	injury	warrant	analysis	help	december	juror	document
murder	eyewitness	direction	document	envelope	reference	killed	judgment	recovered	respectfully
apartment	imprisonment	weapon	addressed	foregoing	taken	go	motion	ballistics	words
penal	photographs	feet	fully	february	apartment	talk	abstract	taken	statement

Table (10) The top 10 keywords learned by semantic NMF with keyword highlighting on the AOBs from both categories

4.2.3 HNMF

For both the top-down (Section 3.3.1) and bottom-up (Section 3.3.2) HNMFs, we begin from an initial factorization of rank 10. For top-down HNMF, each of the 10 super-topics in the first layer is split into three sub-topics. As a result, we discovered a total of 30 sub-topics from AOBs. While, in the bottom-up HNMF, the 10-topic layer is regarded as the sub-topic layer, where 2 super-topic layers are built to combine the 10 sub-topics subsequently.

In Table 11, we display the topic keywords in each of the 2 layers formed by top-down NMF on the AOBs. These topics and the hierarchical structure are interesting and meaningful to explore. Super-topic 3 concerning murder, for example, is split into 3 more specific sub-topics: the first concerning premeditated murder, the second concerning aiding and abetting, and the third concerning gang related murder. Super-topic 4 related to "burglary" is branched into 3 sub-topics, one involving specific charges, one describing case details, and the other specifying injuries.

In Tables 12 through 14 we display a three layer bottom-up HNMF on the AOBs with topic numbers 10, 4, and 2. The hierarchical factorization at the second layer combines the layer one topics related to murder (topics 1 through 3), burglary (topics 4 and 5), and some of the topics related to gangs and courtroom trials (topics 8 through 10). The third layer of hierarchical factorization then combines together topics at the second layer related to murder and burglary (topics 1 and 2).

Primary Topics	Subtopic 1	Subtopic 2	Subtopic 3
Super-topic 1: prosecutor, misconduct, prejudicial, witness, prejudice, admission, statements, uncharged, prosecutorial, probative	*prosecutor*, witness, *misconduct*, *prejudicial*, *statements*, prejudice, *admission*, discretion, burglary, robbery	gang, prosecutor, members, member, beer, expert, shooting, witness, attempted, officer	instruction, identity, eating, stabbed, tank, wearing, fight, scene, perpetrator, citation
Super-topic 2: gang, members, member, expert, shooting, crips, gangs, enhancement, car, territory	*gang*, *member*, *members*, *shooting*, *expert*, murder, *crips*, enhancement, *gangs*, prosecutor	sex, male, murder, *car*, dna, injury, *shooting*, trigger, robbery, september	*gang*, beer, trunk, *expert*, men, intent, *car*, attempted, *members*, attempt
Super-topic 3: murder, intent, aider, abettor, shooting, premeditation, degree, killing, attempted, premeditated	*murder*, *premeditation*, deliberation, *killing*, *shooting*, *intent*, *attempted*, *premeditated*, shot, finding	*murder*, *aider*, *abettor*, abetting, aiding, perpetrator, probable, instruction, car, *intent*	gang, expert, members, beer, *intent*, member, *shooting*, *murder*, prosecutor, tattoo
Super-topic 4: strike, enhancement, felony, burglary, injury, discretion, strikes, bodily, serious, robbery	*strike*, *burglary*, *strikes*, *felony*, convictions, *discretion*, threat, *robbery*, priors, *serious*	mayhem, instruction, unanimity, eating, beer, stabbed, identity, citation, *injury*, tank	*injury*, *bodily*, *enhancement*, *strike*, enhancements, year, *felony*, *serious*, assault, personally
Super-topic 5: sexual, sex, rape, duress, vagina, touching, lewd, penis, offenses, penetration	*sex*, *sexual*, *rape*, propensity, *offenses*, falsetta, molestation, *penis*, raped, admission	*duress*, *lewd*, *sexual*, *penetration*, *touching*, acts, *vagina*, touched, penis, occurred	lines, *sexual*, hearsay, aunt, abuse, spontaneous, expert, exam, testify, declarant
Super-topic 6: accomplice, robbery, corroboration, special, burglary, murder, statements, circumstance, instruction, conspiracy	*accomplice*, *murder*, *corroboration*, *special*, *robbery*, *circumstance*, apartment, *statements*, instruct, accomplices	gang, members, expert, threat, beer, member, attempted, shooting, car, witness	*instruction*, abuse, sexual, *robbery*, testify, strike, witnesses, sex, injury, discretion
Super-topic 7: juror, jurors, prospective, misconduct, deliberations, motion, verdict, dna, excused, shooting	sexual, abuse, injury, instruction, discretion, strike, bodily, robbery, probation, prosecutor	gang, members, member, vargas, *shooting*, murder, expert, car, beer, threat	*juror*, *jurors*, *misconduct*, *prospective*, *shooting*, prosecutor, eyewitness, *motion*, *deliberations*, photo
Super-topic 8: car, phone, detective, apartment, officer, shooting, shot, motion, going, plea	*car*, *phone*, murder, *detective*, *apartment*, *shooting*, shot, *officer*, robbery, *going*	vargas, suv, gang, *car*, blood, men, declaration, pants, bills, partner	probation, *plea*, report, request, unknown, per, conditions, violation, november, file
Super-topic 9: suggestive, witness, eyewitness, photo, lineup, shooting, identifications, photographic, suspect, pack	*suggestive*, *witness*, *photo*, *eyewitness*, *pack*, *lineup*, *identifications*, *photographic*, *suspect*, procedure	instruction, eating, stabbed, identity, tank, wearing, fight, scene, assailant, perpetrator	gang, expert, lineup, *suggestive*, beer, members, *shooting*, member, car, special
Super-topic 10: passion, manslaughter, heat, voluntary, instruction, provocation, self, instruct, lesser, murder	*passion*, *manslaughter*, *heat*, *voluntary*, *murder*, *provocation*, *lesser*, *instruct*, *self*, malice	*instruction*, injury, phone, prosecutor, robbery, strike, testify, bodily, weapon, dna	gang, threat, beer, expert, members, member, threats, attempted, car, associate

Table (11) The top 10 keywords learned by Top-down HNMF on the AOBs from both categories. Keywords in primary topics are italicized in subtopics.

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
murder	murder	car	burglary	strike	sexual	prosecutor	suggestive	juror	gang
manslaughter	accomplice	phone	apartment	enhancement	sex	misconduct	witness	jurors	members
passion	aider	detective	intent	injury	rape	witness	eyewitness	prospective	member
voluntary	abettor	shooting	mayhem	felony	duress	prejudicial	photo	misconduct	expert
heat	robbery	apartment	residential	bodily	vagina	statements	lineup	deliberations	shooting
killing	special	murder	threat	robbery	touching	prejudice	identifications	motion	crips
premeditation	intent	shot	instruction	discretion	lewd	admission	shooting	verdict	gangs
malice	abetting	officer	felony	strikes	penis	prosecutorial	photographic	dna	murder
provocation	aiding	going	burglaries	imposed	offenses	objection	suspect	excused	enhancement
deliberation	instruction	got	unanimity	firearm	penetration	errors	pack	instruction	car

Table (12) The top 10 topic keywords learned by the first layer of bottom-up HNMF on AOBs from both categories

Topic 1	Topic 2	Topic 3	Topic 4
murder	burglary	sexual	gang
car	strike	sex	members
shooting	felony	rape	member
phone	robbery	duress	shooting
shot	enhancement	prosecutor	expert
detective	apartment	vagina	prosecutor
instruction	intent	touching	juror
degree	convictions	lewd	murder
apartment	discretion	offenses	witness
killing	serious	penis	crips

Table (13) The top 10 topic keywords learned by the second layer of bottom-up HNMF on AOBs from both categories

Topic 1	Topic 2
murder	gang
car	members
shooting	member
instruction	shooting
prosecutor	expert
burglary	prosecutor
apartment	juror
phone	murder
robbery	witness
detective	crips

Table (14) The top 10 topic keywords learned by the third layer of bottom-up HNMF on AOBs from both categories

4.3 Classification

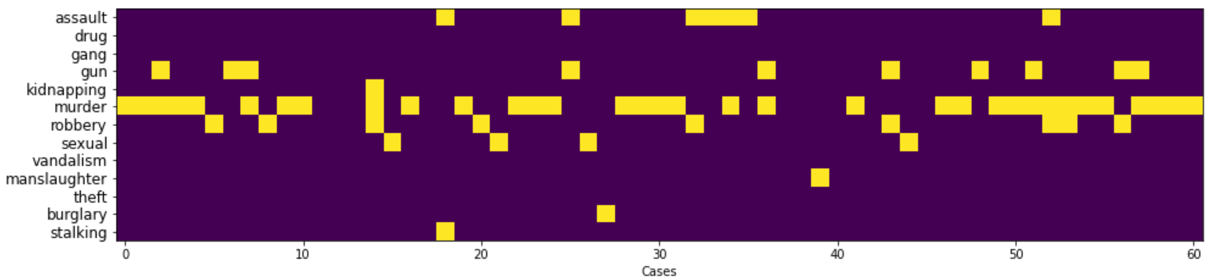
Besides investigating the hidden topics from the initial letters and the AOBs, we take advantage of the supervised nature of SNMF (Section 3.4.1) and SSNMF (Section 3.4.2) to provide some useful insights into classification questions raised by CIP officials. We train models to reconstruct decision type labels from initial letter data and crime type labels from AOB data. Labeling decision type can help to determine whether a new case is worth pursuing by examining only the initial letters. In practice, if the predicted decision type for a case is "*Cases for Investigation*", then this case may be worth pursuing because it is classified as similar to cases that are ready for additional investigation. Labeling crime type can provide an overall understanding of each case before officials take a closer investigation. For each of the following experiments, we calculate the average LAS score (3.4.3) over 10 trials with different training and testing data partitioned by 75% to 25% ratios to measure algorithms' classification accuracy. For building the tf-idf vocabulary, we use the tuning parameters "max_df=0.8" and "min_df=0.2" in the function "TfidfVectorizer".

4.3.1 Decision Type Classification

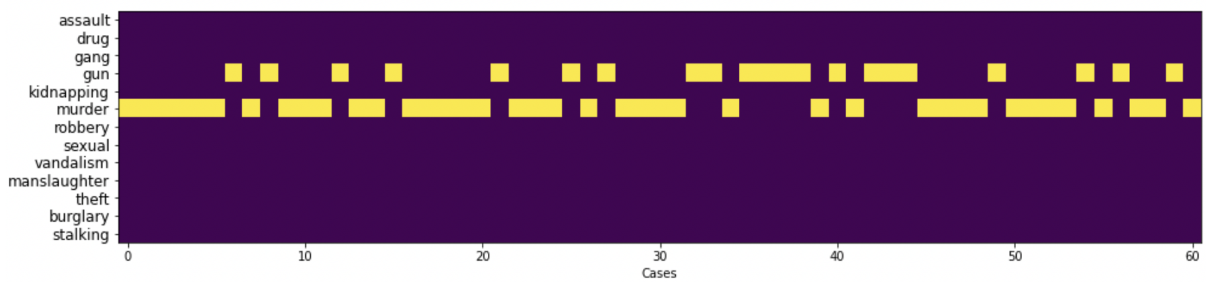
In addition to constructing the training and testing data matrices, we manually create the corresponding decision type label matrices described as follows. Since one case can only belong to one decision type, for each case in the training or testing data, the corresponding column of the label matrix is either $[1, 0]^T$ to indicate that this case belongs to the *Cases for Investigation* type or $[0, 1]^T$ to indicate it belongs to *Cases to be closed*. The SNMF algorithm yields an average LAC score of 65% while the the SSNMF algorithm yields an average LAC score of 55%. The higher rate in SNMF is counter-intuitive since SNMF incorporates fewer cases than SSNMF when training the model. However, the low LAC score of SSNMF can probably be explained by the fact that our training data set is small and adding more cases will lead to over-fitting issues.

4.3.2 Crime Type Classification

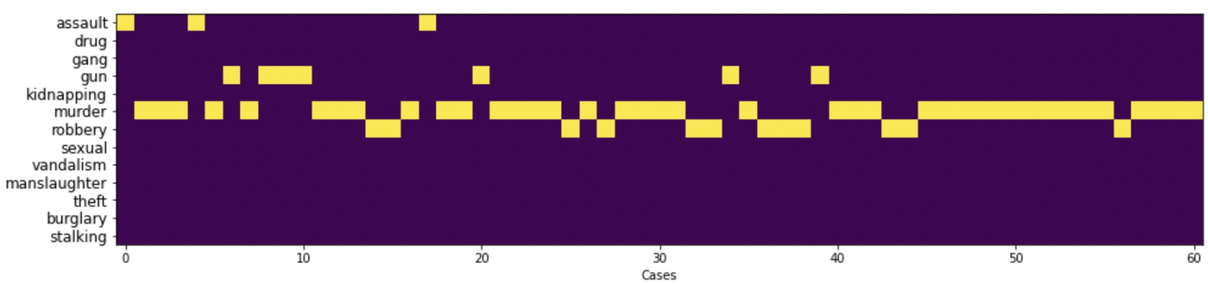
We extract crime labels of each case as described in Section 2.2 and then implement both algorithms. SNMF algorithm results in an average LAC score of 92%, while the SSNMF algorithm yields an average LAC score of 91.8%. Both resulting label matrices are heavily centered towards the dominant labels with the less frequent labels having coefficients very close to 0. Overall, the two algorithms generate promising prediction rates for crime labels, but it can be possibly due to the data set being highly biased towards the crime of murder. Further study on a larger data set is needed to more accurately determine the value of crime type classification. To visualize the results of crime type classification, Figure 2a shows the actual crime labels in the testing data. Figure 2b shows the reconstructed crime labels by SNMF, and Figure 2c shows the reconstructed crime labels by SSNMF.



(a) True crime labels for cases in testing data from both categories.



(b) Reconstructed crime labels for cases in testing data by SNMF from both categories.



(c) Reconstructed crime labels for cases in testing data by SSNMF from both categories.

Figure (2) Comparison of original labels and reconstructed ones. The yellow pixel indicates that the case is assigned to the corresponding crime label on the y-axis, while the dark purple pixel indicates that the case is not assigned.

5 Discussion and Future Works

In this paper, we first provide an exposition of popular variants of Non-negative Matrix Factorization. Then, we discover and analyze meaningful lexical topics from the initial letters and AOBs provided by the California Innocence Project (CIP) through the various NMF methods. We also reconstruct crime type and decision type labels for each case file using (semi)supervised Matrix Factorization methods.

In general, topics generated from the initial letters (discussed in Section 4.1) cover three major themes: seeking assistance, trial information, and evidence. Across all methods, we consistently observe a topic related to seeking assistance, which is the major purpose of the initial letters. Topic results generated from classical NMF are generally trial-related and thus can help CIP identify cases whose initial letters describe their trials comprehensively at first hand. Semantic NMF with keyword highlighting reveals topics related to evidence such as eyewitness, fingerprints, and video footage, which allows CIP to know about the appearance of detailed physical evidence in the document before reading them. Results from top-down and bottom-up HNMF reveal the hidden hierarchical structure, which can potentially help CIP specify or classify trial and evidence information. Identifying the trial information and the corresponding evidence, our analysis of initial letters offers CIP a brief overview of cases.

Topics generated from the AOBs (discussed in Section 4.2) cover two major themes: type of crime and physical evidence. Topic results of classical NMF reveal the diversity of crime types in the AOB documents. Analyzing cases within those topics can potentially help CIP identify special cases belonging to multiple crime types. Top-down HNMF provides additional information and classification about one type of crime, while the bottom-up approach shows how certain crimes can be combined to form a more general category. Besides crime types, CIP also focuses on crucial facts and potential evidence stated in the AOB documents. Semantic NMF with keyword highlighting, by adding more weight on keywords related to physical evidence, generates topics covering various types of evidence. Combining both pieces of knowledge about crime types and physical evidence, those topics can help CIP gain basic understanding (before reading) of the information included in each case, and this pre-knowledge can increase their efficiency in evaluating the case. While results from AOBs are very informative, results from initial letters are also valuable as they depict a general picture of cases from various perspectives such as trial information and evidence when CIP first screens them.

For both initial letters and AOBs, semantic NMF itself generates overlapping and hard-to-interpret topics. A potential explanation is that the extra contextual information semantic NMF adds to the topics does not provide additional information for understanding. Since the SPPMI matrix is trained from the CIP data set itself, bias is unavoidable. In order to reduce the bias, we could train our SPPMI matrix from different data corpus and compare the results. However, its idea of capturing semantic information motivates us to propose a novel keyword highlighting version of semantic NMF. Through emphasizing important keywords (see A.1), this method encourages more topics related to evidence in both initial letters and AOBs. As CIP officials input different sets of highlighting keywords, the topic modeling result will vary correspondingly. As a result, the flexible nature of semantic NMF with keyword highlighting enables CIP officials to gain more control over what kind of information they can expect from case files. Thus, semantic NMF with keyword highlighting could potentially be the most effective method to assist CIP's decision-making due to its flexibility in both highlighting keywords and the SPPMI matrix.

Our second objective is to classify and predict the label of each case using (semi)supervised NMF (discussed in Section 4.3). We perform two classification tasks: the first one on initial letters deciding whether a case should go into investigation or be closed (decision type), and the second one on AOBs deciding the types of crime concerning the case (crime type). Our results have an average prediction accuracy score of 60% for the decision types and 90% for the crime types. The classification accuracy for decision types may suffer from over-fitting due to the limited size of the data set, so more research must be performed on a larger data set to help us draw the conclusion. Potentially, we wish that the first classification can help simplify the screening process since CIP officials can first examine the cases that are classified to be similar to past cases that were investigated. The model trained for the second classification can help CIP officials discern the crime type for each case file when there is a large number of incoming case files. The decreased classification accuracy in predicting decision types may imply that the process of decision making is too complicated to be handled by machine learning algorithms, such as NMF. Aside from those topic modeling results, human judgement is still crucial and decisive. The number of topics in NMF is decided by the user. When, in reality, there are more topics then we asked for, NMF could potentially disregard some of the minor topics or just represent them with only a few words (eg. the topic with Spanish words disappeared in AOB results). In the case of analyzing important legal documents, ignoring details could affect holistic decision making. Meanwhile, people could be wrong and biased against certain details as well but NMF offers an alternative perspective to the data set. Combining both perspectives from machine learning and human judgment could accelerate the process of making a holistic decision.

Therefore, the purpose of machine learning methods, in our case, is more informative rather than conclusive when it comes to decision-making. While applying those methods to other data sets, users should pay attention to privacy and potential bias. Sensitive variables related to people's personal information shall be removed, such as names, addresses, etc. Bias is a constant topic in machine learning algorithms, such as word embeddings [4]. The algorithms should not take people's gender, race, religion, or different dialects they speak into account; To mitigate those biases while preserving the properties of the word embedding, we should also consider some debiasing methods [4, 16]. See also [3] for a nice discussion of the benefits and pitfalls of such debiasing approaches.

Across all the topic modeling results generated from initial letters, we constantly observed topics consisting of Spanish words. Although the way we pre-process the initial letter data set and the implementation of NMF variants is not able to handle multilingual data, these topics of Spanish stop word topics could help CIP notice the existence of multiple languages in the initial letters. In the future, to better include those initial letters in Spanish, or even other languages, into Topic Modeling algorithms, we could apply Multilingual Unsupervised and Supervised Embeddings (MUSE), which take into account the polysemy of words, to translate them into

English [6]. Given the syntax complexity of many languages, we hope to capture contextual phrases instead of a single word to increase the interpretability of the resulting topics. We can achieve this by utilizing n-grams instead of mono-grams when building tf-idf vocabularies. Since, some important keywords may be short phrases, applying n-grams can generate keyphrases so that potentially improve our model's flexibility in capturing meaningful topics. We also wish to apply Non-negative Tensor Factorization methods to the CIP data to honor the often multidimensional structure of the data. For example, utilizing time information of each case can help CIP understand the distribution of cases chronologically. As CIP constantly receives new cases over time, we plan to update our topic clustering results utilizing information from those new cases via an online version of NMF.

Acknowledgments

The authors appreciate Prof. Elizaveta Rebrova and Dr. Denali Molitor for their guidance in this project. The authors also appreciate the support from UCLA Computational and Applied Math REU, NSF BIGDATA #1740325 and NSF DMS #2011140.

References

[1] M. Ailem, A. Salah, and M. Nadif, *Non-negative matrix factorization meets word embedding*, in Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2017, pp. 1081–1084.

[2] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*, O'Reilly Media, Inc., 2009.

[3] S. L. Blodgett, S. Barocas, H. Daumé III, and H. Wallach, *Language (technology) is power: A critical survey of "bias" in NLP*, in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, July 2020, Association for Computational Linguistics, pp. 5454–5476.

[4] T. Bolukbasi, K.-W. Chang, J. Zou, V. Saligrama, and A. Kalai, *Man is to computer programmer as woman is to homemaker? debiasing word embeddings*, in Proceedings of the 30th International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 2016, Curran Associates Inc., p. 4356–4364.

[5] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux, *API design for machine learning software: experiences from the scikit-learn project*, in ECML PKDD Workshop: Languages for Data Mining and Machine Learning, 2013, pp. 108–122.

[6] A. Conneau, G. Lample, M. Ranzato, L. Denoyer, and H. Jégou, *Word translation without parallel data*, arXiv preprint arXiv:1710.04087, (2017).

[7] M. Gao, J. Haddock, D. Molitor, D. Needell, E. Sadovnik, T. Will, and R. Zhang, *Neural nonnegative matrix factorization for hierarchical multilayer topic modeling*, in 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), IEEE, 2019, pp. 6–10.

[8] N. Gillis, *The why and how of nonnegative matrix factorization*, arXiv:1401.5226 [cs, math, stat], (2014). arXiv: 1401.5226.

[9] K. Kenyon-Dean, *Word embedding algorithms as generalized low rank models and their canonical form*, 2019.

[10] D. Kuang and H. Park, *Fast rank-2 nonnegative matrix factorization for hierarchical document clustering*, in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, 2013, pp. 739–747.

[11] D. D. Lee and H. S. Seung, *Learning the parts of objects by non-negative matrix factorization*, Nature, 401 (1999), pp. 788–791.

[12] H. Lee, J. Yoo, and S. Choi, *Semi-Supervised Nonnegative Matrix Factorization*, IEEE Signal Processing Letters, 17 (2010), pp. 4–7.

[13] O. Levy and Y. Goldberg, *Neural word embedding as implicit matrix factorization*, in NIPS, 2014.

[14] J. Li, K. Zhang, and Q. Fan, *Keyword extraction based on tf-idf for chinese news document*, Wuhan University Journal of Natural Sciences, 12 (2007), pp. 917–921.

[15] T. Li and C. Ding, *The relationships among various nonnegative matrix factorization methods for clustering*, in Sixth International Conference on Data Mining (ICDM'06), 2006, pp. 362–371.

[16] O. Papakyriakopoulos, S. Hegelich, J. C. M. Serrano, and F. Marco, *Bias in word embeddings*, in Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, New York, NY, USA, 2020, Association for Computing Machinery, p. 446–457.

[17] J. Ramos, *Using tf-idf to determine word relevance in document queries*, in Proceedings of the first instructional conference on machine learning, vol. 242, New Jersey, USA, 2003, pp. 133–142.

[18] G. Salton and C. Buckley, *Term-weighting approaches in automatic text retrieval*, Information Processing & Management, 24 (1988), pp. 513–523.

[19] D. Tu, L. Chen, M. Lv, H. Shi, and G. Chen, *Hierarchical online nmf for detecting and tracking topic hierarchies in a text stream*, Pattern Recognition, 76 (2018), pp. 203–214.

[20] W. Xu, X. Liu, and Y. Gong, *Document clustering based on non-negative matrix factorization*, in Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, 2003, pp. 267–273.

A Appendix

A.1 Highlighted Keywords

eyewitness, microscopy, shaken baby syndronme, sbs, abusive head trauma, aht, false confession, coerced confession, rampart, comparative bullet lead analysis, cbla, tool mark, toolmark,tread mark comparison, tread mark analysis, fiber comparison, impression comparison, impression analysis, arson, ballistics, blood splatter, handwriting comparison, informat, snitch, strangle, ligature, sodomy, sexual assault, intercourse, digital penetration, penetration by a foreign object, saliva, semen, sperm, amylase, seminal fluid, duct tape, bindings, mixture, bite mark, bitemark, fingerprint, fingernail scrapings, shell casing, blood type