

CRIME INCIDENTS EMBEDDING USING RESTRICTED BOLTZMANN MACHINES

SHIXIANG ZHU, YAO XIE
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GA, USA
SHIXIANG.ZHU@GATECH.EDU

CREATING THE NEXT®



INTRODUCTION

WHAT IS CRIME ANALYSIS?

CREATING THE NEXT®

WHAT IS THE GOAL IN CRIME ANALYSIS

A fundamental and one of the most challenging tasks in crime analysis is to find **related crime series**, which are committed by the same individual or group.



Crime incident reports (a.k.a. police reports) are a large source of data that contains rich information for detecting related crime series.

According to the crime analysts, the free-text narrative contains **the most useful information** from their investigation.

SHERIFF'S DEPARTMENT
County of Stanislaus

DATE: 3-23-70 FOLLOWUP REPORT CONTINUATION REPORT FILE NO. 062677

VICTIM: 100 OFFER OR DENIAL REPORT LOCATION OF OCCURRENCE

TOIDERS, Kathleen Ann

All headings of Followup Report, but 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, 1000.

3-23-70 2:30 AM: Received a call from Patterson PD to meet the complainant at their station regarding an incident which occurred in the County area. Arrived at the Patterson PD at 2:50 AM.

Contacted the complainant who stated at approx. 11:15 PM, 3-22-70 she had been driving on Hwy. 132 near Interstate 5 when she observed a late model vehicle, light tan in color, no other description, following her and blinking it's lights as if the operator of the vehicle desired the complainant to stop. Complainant stated she stopped her vehicle, a 1957 Chevrolet S/W bearing Calif. lic. HQJ 518, somewhere in the area of Hwy. 132 and Interstate 5. She advised that the exact area is unfamiliar to her and that she was not sure where the vehicle had been stopped.

Complainant stated that a WMA, approx. 30 yrs., 5-9, 160 lbs., drk. hair wearing blk. rimmed plastic glasses, wearing a drk. ski jacket and drk. blue bell-bottomed pants, approached her vehicle and advised her that her left rear wheel of her vehicle was wobbling and that he would repair the malfunction. Complainant stated she stayed in her vehicle and observed the suspect take a tire tool to the rear of her vehicle and work on the wheel for approx. one to two minutes, at which time the suspect approached the complainant and advised her the wheel was okay at this time and that she could proceed. The complainant then observed the suspect pull away from the roadway and proceed from the area.

Complainant stated she started her vehicle and was attempting to pull away from the curb when the vehicle came to a lurching stop at which time she exited the vehicle in an attempt to ascertain what the problem was. She discovered that the left rear wheel had fallen off the vehicle and it had only been secured by one lug bolt. Stated at that time she observed the suspect return in his vehicle and he advised her he would transport her to a service station in order to have the vehicle repaired. The complainant entered the suspect's vehicle carrying her small daughter, approx. 10 yrs., and left the area under the pretense that the suspect was attempting to locate a service station.

Complainant stated the suspect drove around in the County area, possibly near Tracy for approx. one hour to 1 1/2 hours and several times she had asked the suspect if he intended to stop at a station in order for her to seek help to have her vehicle repaired. (Complainant stated the suspect was quite friendly with her, did not make any advances toward her, or threats toward her, and when asked if he was going to stop he would merely elude the question and start talking about something else.)

She became quite frightened, feeling that possibly the suspect intended to do some physical injury to her and that when the suspect stopped at a Stop sign, the exact area or location unknown, she jumped from the vehicle carrying her daughter, and ran into a field nearby, hiding from the suspect. Complainant stated the suspect merely closed the door that the complainant had opened in order to leave the vehicle, and then had driven away. She stated she stopped a vehicle on the roadway and had advised the occupants of same what had occurred. The occupants of this vehicle had transported her to Patterson PD in order for her to seek help.

The persons transporting the complainant were not present upon this writer's arrival, therefore their names were not obtained.

Complainant advised that she was not injured during the incident, or her daughter had not been.

REPORTING OFFICER: *Chase #22* INVESTIGATING OFFICER: Lovett #160 PAGE (cont.)

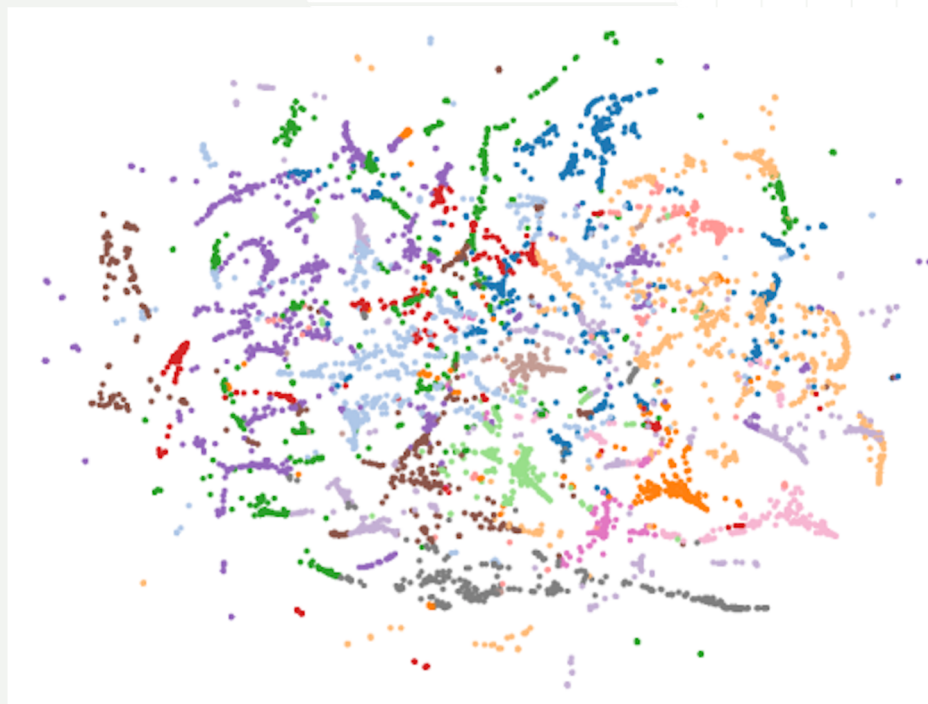
steno/pjh

FOLLOWUP - CONTINUATION REPORT FRED WOODS MAR 23 1970

PROBLEM

Inspired by the idea of *word2vec* (word embeddings).

Find a **distributed representations** for the free-text narratives of the police reports, which map key information into a feature vector space that automatically captures the similarity of incidents.





METHOD

EMBEDDING USING GBRBM

CREATING THE NEXT®



Questioning the victims or the witnesses

DEKALB COUNTY POLICE DEPARTMENT
GA0440200

Case # 17-05147

DATE: 07/1/2017 3:44:50 PM Agency: 00000000

INITIAL REPORT

On 04/20/17 at approximately 1500 hours, I responded to 3384 Glenwood Rd the Chevron Gas station in reference to a female soliciting customers for money inside the location. When I arrived at the scene, I observed the female suspect later identified as Tasha McCarty standing inside the location by the front door. When I arrived the store the suspect attempted to push me out of the way and walk out of the door and stated "be get you look for is over there".

I stopped the suspect at the door and told her to step back. While I was attempting to speak with the complainant behind the counter, the suspect continued to try and walk past me. I responded to her several times to stay right where she was. She then stated that she was a Federal Agent and stated a random badge number which I do not recall.

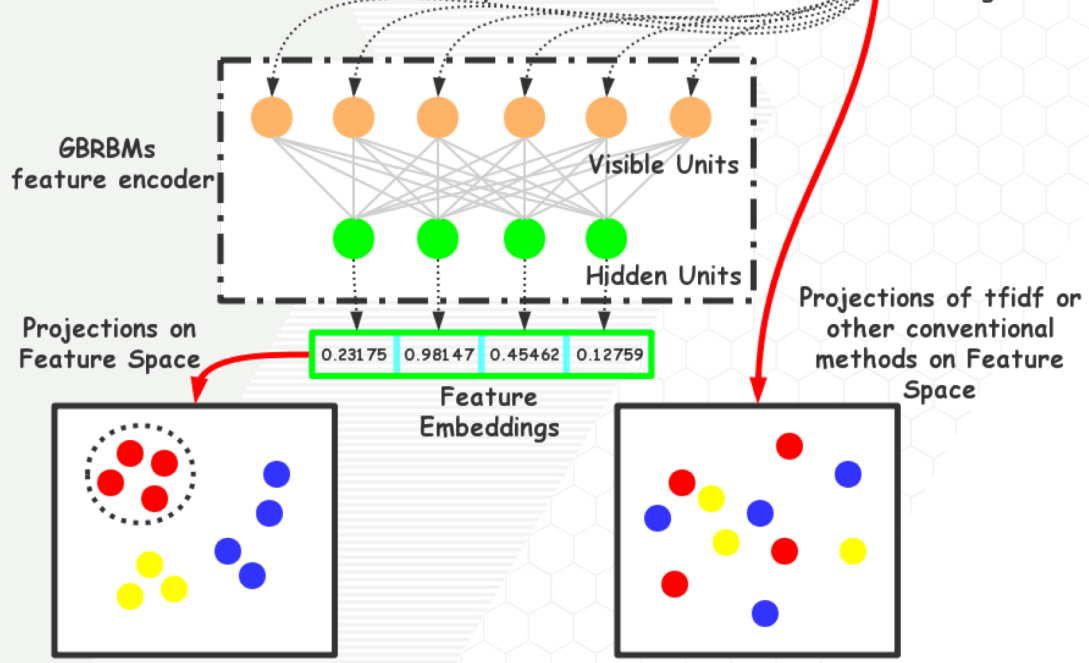
When I told her that she could be arrested for impersonating an officer, she told me that I was impersonating an officer and then reached out and grabbed my badge. I attempted to push the female back and told her not to touch me and she grabbed my vest and radio. I attempted to call for another unit over the radio, however I was unable to get out due to another officer speaking over the air.

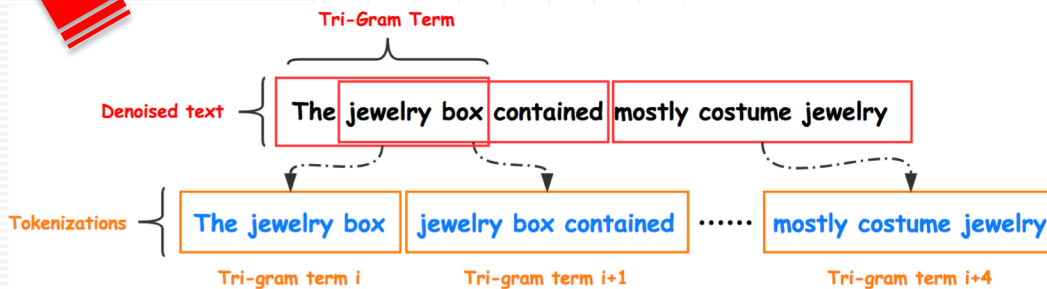
The female continued to grab my vest with her hands and I then deployed my ASP baton and instructed her to lay down on the ground. The suspect refused my commands and I then delivered an unknown amount of baton strikes to her left leg. After the suspect began kicking me, I continued my baton strikes to her legs and forearms instructing her to lay down on the ground. One strike on her left leg, one strike on her right leg, one strike on her left forearm, and one strike on her right forearm.

Narratives of the crime report

Terms	tfidf	Counts
Stolen	0.00345	3
Gun	0.00121	1
Break in	0.00048	1
Jewelry	0.00725	5
...

Representations of the Bag of Words





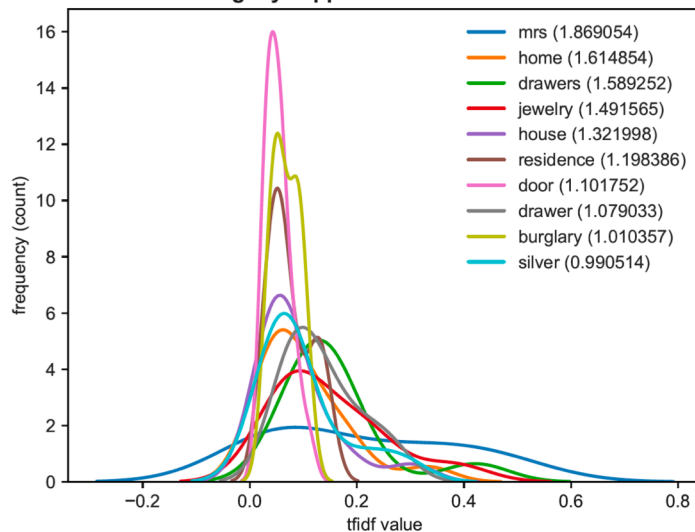
What is TF-IDF weighting

In information retrieval, TF-IDF, short for term frequency-inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus.

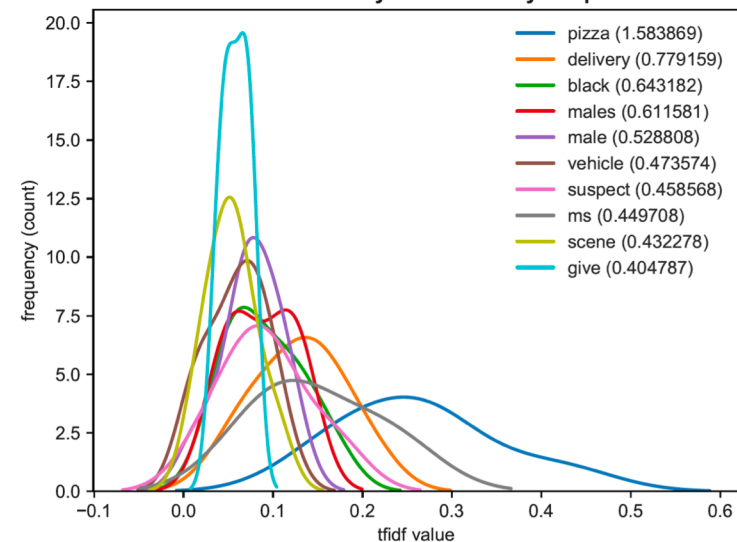
$$\text{TF-IDF}_{t,d} = \text{TF}_{t,d} \times \text{IDF}_t \quad (1)$$

$$\text{TF}_{t,d} = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \quad \text{IDF}_t = \log \frac{N}{|\{d \in D : t \in d\}|}$$

Burglary happened in Buckhead Area



Pedestrian Robbery Committed by Suspect A



Energy Function

$$E(v, h) = - \sum_{i,j} w_{ij} h_j \frac{v_i}{\sigma_j} - \sum_i \frac{(v_i - b_i)^2}{2\sigma^2} - \sum_j h_j c_j.$$

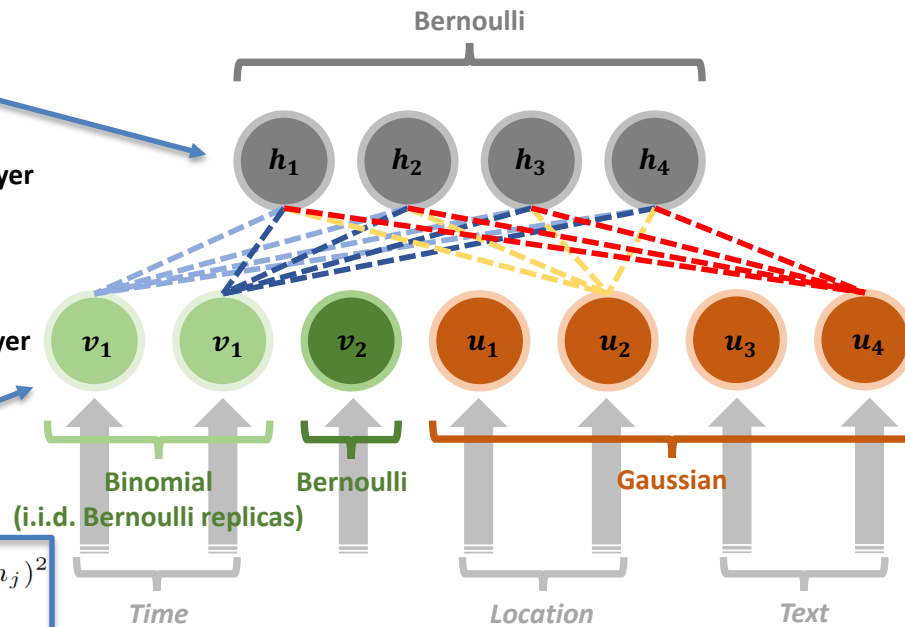
Independency

$$p(v|h) = \prod_{i=1}^m p(v_i|h), \quad p(h|v) = \prod_{j=1}^n p(h_j|v).$$

$$p(h_j = 1|v) = \sigma\left(\sum_{i=1}^m w_{ij} \frac{v_i}{\sigma_i} + c_j\right)$$

Hidden Layer

Visible Layer



$$p(v_i = v|h) = \frac{1}{\sigma_i \sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma_i^2} (v_i - b_i - \sigma_i \sum_{j=1}^n w_{ij} h_j)^2}$$

Likelihood

$$p(v) = \sum_h p(v, h) = \frac{\sum_h e^{-E(v, h)}}{\sum_{v, h} e^{-E(v, h)}}.$$

Contrastive Divergence

$$\begin{aligned} \text{CD}_k(\theta, v^{(0)}) = & - \sum_h p(h|v^{(0)}) \frac{\partial E(v^{(0)}, h)}{\partial \theta} \\ & + \sum_h p(h|v^{(k)}) \frac{\partial E(v^{(k)}, h)}{\partial \theta}. \end{aligned}$$

A. Fischer and C. Igel, “An Introduction to Restricted Boltzmann Machines,” Lecture Notes in Computer Science: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications , vol. 7441, pp. 14–36, 2012.

T. Tieleman, “Training restricted Boltzmann machines using approximations to the likelihood gradient,” Proceedings of the 25th international conference on Machine learning - ICML '08 , pp. 1064–1071, 2008.

G. E. Hinton, “Training products of experts by minimizing contrastive divergence,” Neural Computation , 2002.



EXPERIMENTS

CREATING THE NEXT®

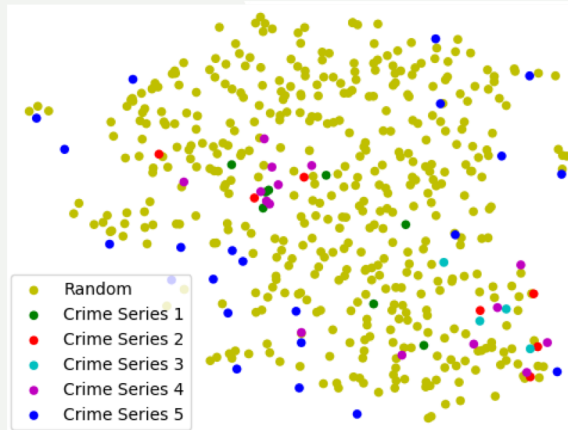
Table 1. Details of the test data

Id	Number	Category
Crime Series 1	8	<i>Robbery at Residence</i>
Crime Series 2	7	<i>Robbery at Gas Station</i>
Crime Series 3	4	<i>Pedestrian Robbery</i>
Crime Series 4	15	<i>Attempt Auto Theft</i>
Crime Series 5	22	<i>Burglary</i>
Random Cases	441	<i>Over 89 Categories</i>
Total	497	

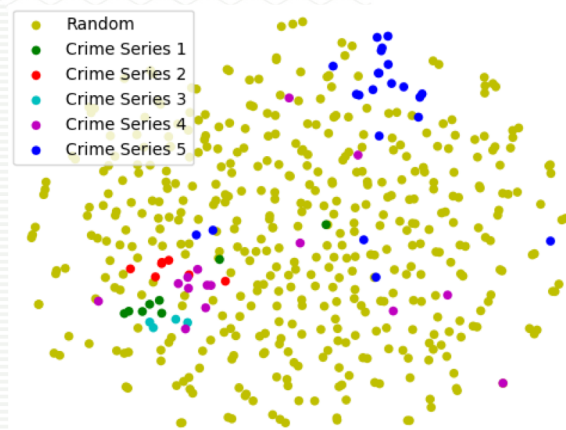
The dataset contains **five hand-labeled crime series** that were identified as committed by five individual arrestees, and **441 randomly selected irrelevant crime cases**.

Table 2. The comparison of the training time.

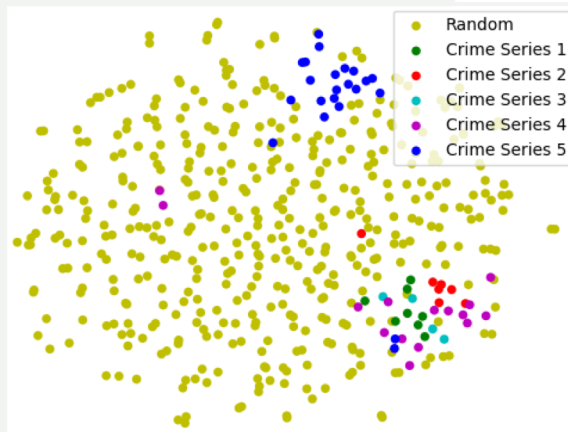
Methods	Training Time
GBRBMs with 1000 units	<i>~ 2 mins</i>
GBRBMs with 2000 units	<i>~ 3 mins</i>
GBRBMs with 5000 units	<i>~ 7 mins</i>
LDA with 1000 topics	<i>~ 5 mins</i>



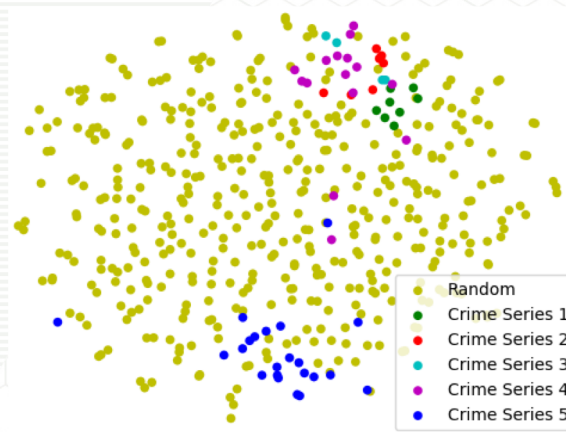
LDA (1000 topics)



GBRBM (1000 nodes)



GBRBM (2000 nodes)



GBRBM (5000 nodes)

- We have presented a novel approach for detecting crime series that are related, using embedding found by the **Gaussian-Bernoulli Restricted Boltzmann Machine** (GBRBM).
- The GBRBM tends to map related cases (that share certain correlation in the raw feature) into features that are in the **vicinity in the Euclidean space**.
- Our methods demonstrate very promising results on **real police data** and demonstrated that the feature embeddings can have advantages over the conventional text processing methods on detecting crime series in certain cases.
- Ongoing work is to develop an **online** crime series detection algorithm on a **larger** and **self-increasing** real dataset.

Thank You