SOM & KMeans Comparison

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Methods Tags Batch no. | Algo | Initial Centroids/ Neurons | Feature Vector | Keywords per project | Results | | |
| Total Iterations | Clusters | Outliers |
| 5 | KMeans | Farthest : [0, 3] |  |  | 2 | C0: [0, 1, 6, 7, 8, 10, 12]  C1: [2, 3, 4, 5, 9, 11] | No |
| Random |  |  | 4 | C0: [2, 3, 4, 5, 9, 11, 12] C1: [0, 1, 6, 7, 8, 10] | 12 |
| SOM | Original tags Farthest : [0, 2] |  |  | 39 | C0: [0, 1, 6, 7, 8, 10, 12]  C1: [2, 3, 4, 5, 9, 11] | No |
| Base tags  Farthest : [0, 2] |  |  | 39 | C0: [0, 1, 6, 7, 8, 10, 12]  C1: [2, 3, 4, 5, 9, 11] | No |
| 4 | KMeans | Farthest : [0, 2] |  |  | 2 | C0: [0, 1, 6, 7, 8, 9, 10, 11, 12]  C1: [2, 3, 4, 5] | 9, 11 |
| Random |  |  | 2 | C0: [2, 3, 6, 8 10, 12]  C1: [0, 1, 4, 5, 7, 9, 11] | 2, 3, 0, 1, 7, 9, 11 |
| SOM | Farthest : [0, 2] |  |  | 39 | C0: [0, 1, 6, 7, 8]  C1: [2, 3, 4, 5, 9, 10, 11, 12] | 10, 12 |
| Random |  |  |  | Doesn’t give appropriate results |  |
| 10 | KMeans with original tags |  |  |  |  |  |  |
| KMeans with base words |  |  |  |  |  |  |
|  | Random |  |  | 3 | C0: [5, 7, 8, 10, 14]  C1: [0, 1, 2, 3, 4, 6, 9, 11, 12, 13, 15]  C2: [16] | Too many |
| SOM – Semantic Match | Farthest: [0, 4, 6] | 75 | all | 39 | C0: [0, 1, 2, 3, 9, 12, 14]  C1: [4, 5, 7, 8, 10, 11]  C2: [6, 13, 15, 16] | No |
| SOM – Semantic Match | Farthest: [0, 4, 6] | 70 | 10  12 keywords shud be used | 39 | C0: [5, 7, 8, 10, 11]  C1: [4]  C2: [0, 1, 2, 3, 6, 9, 12, 13, 14, 15, 16] | C0+C1 |
| SOM with original tags | Farthest: [0, 4] | 79 | 10 | 39 | C0: [0, 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]  C1: [4] |  |
| SOM with original tags | Farthest: [0, 4] | 84 | all | 39 | C0: [0, 1, 2, 3, 6, 9, 12, 13, 14, 15, 16]  C1: [4, 5, 7, 8, 10, 11] |  |
| 7 | KMeans | Farthest : [0, 3] |  |  | 2 | C0: [0, 1, 6, 7, 8, 10, 12]  C1: [2, 3, 4, 5, 9, 11] | No |
| Random |  |  | 4 | C0: [2, 3, 4, 5, 9, 11, 12] C1: [0, 1, 6, 7, 8, 10] | 12 |
| SOM | Original tags Farthest : [0, 7, 13, 14] |  |  | 39 | C0: [0, 1, 2, 3, 4, 5, 6]  C1: [7,8, 9, 10, 11, 12]  C2: [13, 15]  C3: [14, 16] | 14, 16 were better to be the part of C2 |
| Base tags  Farthest : [0, 7, 13] |  |  | 39 | C0: [0, 1, 2, 3, 4, 5, 6]  C1: [7, 8, 9, 10, 11, 12]  C2: [13, 14, 15, 16] | No |
| 6 | KMeans |  |  |  |  |  |  |
|  |  |  |  |  |  |
| SOM | Original tags Farthest : [0, 7] |  |  | 39 | C0: [0, 1, 2, 3, 4, 5, 6]  C1: [7, 8, 9, 10, 11, 12] | No |
| Base tags  Farthest : [0, 7] |  |  | 39 | C0: [0, 1, 2, 3, 4, 5, 6]  C1: [7, 8, 9, 10, 11, 12] | No |
| 8 | KMeans |  |  |  |  |  |  |
|  |  |  |  |  |  |
| SOM | Original tags Farthest : [0, 1, 4] |  |  | 39 | C0: [0, 3]  C1: [1,2, 6]  C3: [4, 5, 7] | No |
| Base tags  Farthest : [0, 1, 4] |  |  | 39 | C0: [0, 3]  C1: [1,2, 6]  C3: [4, 5, 7] | No |
| 9 | KMeans |  |  |  |  |  |  |
|  |  |  |  |  |  |
| SOM | Original tags Farthest : [0, 4, 6, 13] |  |  | 39 | C0: [0, 1, 2, 3, 9, 12, 14]  C1: [4, 5, 7, 8, 10, 11]  C2: [6, 16]  C3: [13, 15] | C2, C3 better to be merged |
| Base tags  Farthest [0, 4, 6] |  |  | 39 | C0: [0, 1, 3, 9, 14]  C1: [2, 4, 5, 7, 8, 10, 11, 12]  C2: [6, 13, 15, 16] | 2, 12 |
| Data provided by Shamsa | KMeans |  |  |  |  |  |  |
|  |  |  |  |  |  |
| SOM | Original tags Farthest : [0, 1, 4, 5, 7] |  |  | 39 | C0: [0]  C1: [1, 2, 3, 6]  C2: [4]  C3: [5]  C4: [7] | None |
| Base tags  Farthest [ |  |  | 39 |  |  |
| project\_domain \_keywords \_3\_10-keywords.csv | KMeans |  |  |  |  |  |  |
|  |  |  |  |  |  |
| SOM | Original tags Farthest : [0, 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] | 257 |  | 39 | C0: [0]  C1: [1]  C2: [2, 3, 4, 5, 6]  C3: [7]  C4: [8]  C5: [9]  C6: [10]  C7: [11]  C8: [12]  C9: [13]  C10: [14]  C11: [15]  C12: [16]  C13: [17]  C14: [18]  C15: [19]  C16: [20]  C17: [21, 22, 27, 31]  C18: [23]  C19: [24]  C20: [25]  C21: [26]  C22: [28]  C23: [29]  C24: [30]  C25: [32]  C26: [33]  C27: [34] |  |
| Base tags  Farthest [ | 183 |  | 39 | C0: [2, 3, 4, 5, 6, 31]  C1: [1]  C2: [0, 9, 11, 12, 13, 16, 17, 18, 19, 21, 22, 23, 24, 26, 27, 28, 29, 30, 32]  C3: [7]  C4: [8]  C5: [10]  C6: [14]  C7: [15]  C8: [20]  C9: [25]  C10: [33]  C11: [34] |  |
| project\_domain\_keyword (ordered by frequency of words).csv  Number of keywords: 10 | KMeans | Original tags Farthest : [0, 2, 7] | 204 |  |  | C0: [0, 1, 9, 10, 11, 12, 14, 16, 17, 18, 19, 20, 21, 23, 25, 26, 27, 28, 30, 32, 33]  C1: [2, 3, 4, 5, 6, 15, 22, 24, 31]  C2: [7, 8, 13, 29, 34] |  |
|  |  |  |  |  |  |
| SOM | Original tags Farthest : [0, 2, 7] | 204 |  | 39 | C0: [0, 1]  C1: [7, 8, 11, 12, 13, 19, 21, 23, 25, 29, 34]  C2: [2, 3, 4, 5, 6, 9, 10, 14, 15, 16, 17, 18, 20, 22, 24, 26, 27, 28, 30, 31, 32, 33] |  |
| Base tags  Farthest |  |  | 39 |  |  |
| wikigames\_project\_domain\_keywords | KMeans with original tags |  |  |  |  |  |  |
| KMeans with base words |  |  |  |  |  |  |
|  | Random |  |  | 3 | C0: [5, 7, 8, 10, 14]  C1: [0, 1, 2, 3, 4, 6, 9, 11, 12, 13, 15]  C2: [16] | Too many |
| SOM – Semantic Match | Farthest: [0, 4, 6] | 75 | all | 39 | C0: [0, 1, 2, 3, 9, 12, 14]  C1: [4, 5, 7, 8, 10, 11]  C2: [6, 13, 15, 16] | No |
| SOM – Semantic Match | Farthest: [0, 4, 6] | 70 | 10  12 keywords shud be used | 39 | C0: [5, 7, 8, 10, 11]  C1: [4]  C2: [0, 1, 2, 3, 6, 9, 12, 13, 14, 15, 16] | C0+C1 |
| SOM with original tags | Farthest: [0, 4] | 36 | 10 | 39 | C0: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18] |  |
| SOM – Exact Matching | Farthest: [5, 9, 15] | 541 | all | 39 | C0: [0, 1, 2, 3, 4, 5, 6, 7]  C1: [8, 9, 10, 11, 12, 13]  C2: [14, 15, 16, 17, 18] | No |

* Reason of outliers in batch 4: compound tags