**CHAPTER 4**

**EXPERIMENTAL RESULTS AND ANALYSIS**

This chapter describes the experimental results obtained from a document illustration. It also presents the analysis to account for the tests performed.

**4.1 Illustration**

The algorithm proposed in Chapter 3 was implemented and tested using the benchmark document collection. This chapter puts forward the results of this experimentation. To clearly illustrate the effectiveness of the proposed algorithm for Construction of Collaborative Research Interest Group, a case study is presented to describe a typical scenario, where

* There are 5 researchers viz. i, j, k, n & m. Therefore, Nu = 5
* There are 5 entries in each of the researcher’s blog site.

The following table 4.1 shows the blog entries of each of the Researcher i, j, k, n & m.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  **Researcher**  **Entry** |  **i** |  **j** |  **k** |  **n** |  **m** |
|  **1.** | w1, w16, w3, w2, w17, w9, w24, w25 | w14,w8, w6, w7, w17, w21, w25 | w11, w7, w2, w9, w19, w21, w25 | w13, w13, w10, w14,w21, w22 | w10,w15w2, w21, w23, w24 |
|  **2.** | w4, w2,w3, w14, w11, w18, w21, w23 | w1, w16, w11, w7, w18, w17, w6, w23 | w14,w10,w4,w9,w19, w20 | w11,w13,w6, w5, w20, w21 w22, w25 | w14,w16w9, w8, w18, w23 w24 |
|  **3.** | w1,w2,w6,w13, w20 | w7, w3, w18,w8, w17, w24 | w9, w19, w11,w10,w17, w23 | w13, w14,w18, w12, w20, w22 | w15,w19w1, w16, w20, w23 w24 |
|  **4.** | w1,w2, w4,w8,w15, w10 | w6, w6,w7,w17, w22 | w12, w9, w19,w16, w24 | w17, w13,w2, w20, w21, w22 | w11,w17, w6, w15, w24, w25 |
|  **5.** | w1,w2,w5,w3, w19 | w7,w18,w15, w2,w18 ,w6,w17, w1 | w19, w9, w17, w10, w10 | w18,w7,w13,w13, w20, w23 w24 | w3, w13, w22, w23 w24, w25 |

 **Table 4.1:** Sample blog entries of 5 researchers

**4.1.1 Interest Mining Module**

This module is concerned with the construction of Collaborative Interest Group by uncovering shared interest relationships between people, based on their blog document entries. The key point of constructing this Collaborative Interest Group is the calculations of interest similarity relations and application of the K-means clustering technique to cluster researchers with similar interests into the same group.

**4.1.1.1 Interest Vector calculations**

We have the interest vector corresponding to each of the researcher i, j, k, n & m represented as Vi, Vj, Vk, Vn, Vm. The calculation for these vectors using equation 3.2 is shown below:

**For Researcher i:** The Interest Vector is: Vi = (Si1, Si2, Si3, Si4, Si5)where ;

 Si1=ef (w1) x log [5 / uf (w1)]

 Si2=ef (w2) x log [5 / uf (w2)]

 Si3=ef (w3) x log [5 / uf (w3)]

 Si4=ef (w4) x log [5 / uf (w4)]

 Si5=ef (w5) x log [5 / uf (w5)]

Now, from table 4.1, we find the values for ef’s and uf’s for the corresponding words:

ef (w1)=4 ; uf (w1)=3 => Si1 = 4 \* log (5/3) = 0.8874

ef (w2)=5 ; uf (w2)=4 => Si2 = 5 \*log (5/4) = 0.4846

ef (w3)=3 ; uf (w3)=2 => Si3 = 3\*log (5/2) = 1.1938

ef (w4)=2 ; uf (w4)=1 => Si4  = 2\*log (5/1) = 1.3979

ef (w5)=1 ; uf (w5)=1 => Si5 = 1\*log (5/1) = 0.6989

Thus,

Vi= (0.8874, 0.4846, 1.1938, 1.3979, 0.6989)

**For Researcher j:** The Interest Vector is: Vj = (Sj6, Sj7, Sj8, Sj17, Sj18)where;

 Sj6=ef (w6) x log [5 / uf (w6)]

 Sj7=ef (w7) x log [5 / uf (w7)]

 Sj8=ef (w8) x log [5 / uf (w8)]

 Sj17=ef (w17) x log [5 / uf (w17)]

 Sj18=ef (w18) x log [5 / uf (w18)]

Now, from table 4.1, we find the values for ef’s and uf’s for the corresponding words

ef (w6)=4; uf (w6)=3 => S6 = 4\*log (5/ 3) = 0.8874

ef (w7)=5; uf (w7)=2 => S7 = 5\*log (5/2) =1.9897

ef (w8)=2; uf (w8)=2 => S8 = 2\*log (5/2) = 0.7959

ef (w17)=5; uf (w17)=4 => S17 = 5\*log (5/4) = 0.4845

ef (w18)=3; uf (w18)=3 => S18 = 3\*log (5/3) = 0.6655

Thus,

Vj = (0.8874, 1.9897, 0.7959, 0.4845, 0.6655)

**For Researcher k:** The interest vector is:Vk = (Sk9, Sk10, Sk11, Sk12, Sk19)where;

 Sk9=ef (w9) x log [5 / uf (w9)]

 Sk10=ef (w10) x log [5 / uf (w10)]

 Sk11=ef (w11) x log [5 / uf (w11)]

 Sk12=ef (w12) x log [5 / uf (w12)]

 Sk19=ef (w19) x log [5 / uf (w19)]

Now, from table 4.1, we find the values for ef’s and uf’s for the corresponding words

ef (w9)=5; uf (w9)=2 => S9 = 5\*log (5/2) = 1.9897

ef (w10)=3; uf (w10)= 3 => S10 = 3\*log (5/3) = 0.6655

ef (w11)=2; uf (w11)=4 => S11 = 2\*log (5/4) = 0.1938

ef (w12)=1; uf (w12)=1 => S12 = 1\*log (5/1) = 0.6988

ef (w19)=5 uf (w19)=2 => S19 = 5\*log (5/2) = 1.9897

Thus,

Vk = (1.9897, 0.6655, 0.1938, 0.6988, 1.9897)

**For Researcher n:** The Interest Vector is:Vn = (Sn13, Sn14, Sn20, Sn21, Sn22)where;

 Sn13=ef (w13) x log [5 / uf (w13)]

 Sn14=ef (w14) x log [5 / uf (w14)]

 Sn20=ef (w20) x log [5 / uf (w20)]

 Sn21=ef (w21) x log [5 / uf (w21)]

 Sn22=ef (w22) x log [5 / uf (w22)]

Now, from table 4.1, we find the values for ef’s and uf’s for the corresponding words

ef (w13)=5; uf (w13)=2 => S13 = 5\*log (5/2) = 1.9897

ef (w14)=2; uf (w14)=4 => S14 = 2\*log (5/4) = 0.1938

ef (w20)=4; uf (w20)=3 => S20 = 4\*log (5/3) = 0.8874

ef (w21)=3; uf (w21)=4 => S21 = 3\*log (5/4) = 0.2907

ef (w22)=4; uf (w22)=2 => S22 = 4\*log (5/2) = 0.8874

Thus,

Vn = (1.9897, 0.1938, 0.8874, 0.2907, 0.8874)

**For Researcher m:** The Interest Vector is: Vm = (Sm15, Sm16, Sm23, Sm24, Sm25)where;

 Sm13=ef (w15) x log [5 / uf (w15)]

 Sm14=ef (w16) x log [5 / uf (w16)]

 Sm20=ef(w23) x log [5 / uf (w23)]

 Sm24=ef(w24) x log [5 / uf (w24)]

 Sm25=ef(w25) x log [5 / uf (w25)]

Now, from table 4.1, we find the values for ef’s and uf’s for the corresponding words

ef (w15)=3; uf (w15)=2 => S15 = 3\*log (5/2) = 1.1938

ef (w16)=2; uf (w16)=3 => S16 = 2\*log (5/3) = 0.4436

ef (w23)=4; uf (w23)=4 => S23 = 4\*log (5/4) = 0.3876

ef (w24)=5; uf (w24)=4 => S24 = 5\*log (5/4) = 0.4845

ef (w25)=2; uf (w25)=4 => S25 = 2\*log (5/4) = 0.1938

Thus,

Vm = (1.1938, 0.4436, 0.3876, 0.4845, 0.1938)

**4.1.1.2 Interest Similarity Score calculations**

Using the formula defined in equation 3.3, we calculate the values of Similarity Score between each of the 2 researchers:

Rij = 0.7063; Rik = 0.7110; Rin = 0.7502; Rim = 0.8064; Rjk = 0.6688; Rjn = 0.6132

Rjm = 0.7424; Rkn = 0.8786; Rkm = 0.8140; Rnm = 0.9169

As all the elements of both the vectors taken at a time to calculate the similarity score are positive, thus the range of similarity score is between 0 to 1.

This indicates that:

* The value of 1 means that the 2 researchers have exactly similar interests and;
* The value of 0 means that the 2 researchers do not have any similar interests at all.

Therefore, we can say that:

* The researchers n & m have almost similar interests (as Rnm= 0.9169, approx 1 )
* The researchers k & n have similar interests to a very great extent (as Rkn = 0.8786)
* The researchers “k & m” and “i & m” have quite a lot similar interests (as R km = 0.8140 and Rim = 0.8064)
* The researchers “j & k” and “j & n” have quite less similar interests (as R jk = 0.6688 and Rjn = 0.6132)

**4.1.1.3 Collaborative Interest Group Construction**

We construct the collaborative interest group by using the technique of K-means clustering algorithm with the help of two basic steps. We first construct the researcher groups by finding the membership of each of the researcher using the formula defined in equation 3.4. This step would give us the total number of clusters required, denoted by K. And then we assign points to the closest centroid by taking the proximity measure as the distance between two researchers using the formula defined in equation 3.5.

1. **Construction of Researcher Groups**
2. Membership for group i

 Step 1: Calculate the threshold for this group i.e. Ti

 Ti = $ \frac{1}{5}$ [Rii + Rij + Rik +Rin + Rim]

 = $ \frac{1}{5}$ [1 + 0.7063 + 0.7110 + 0.7502 + 0.8064]

 = 0. 79478

 Step 2: Deciding the members for group i

 As we can see,

 Rii >Ti and Rim >Ti , therefore Researcher i and Researcher m belong to group i.

1. Membership for group j

 Step 1: Calculate the threshold for this group i.e. Tj

 Tj = $ \frac{1}{5}$ [Rji + Rjj + Rjk +Rjn + Rjm]

 = $ \frac{1}{5}$ [0.7063 + 1 + 0.6688 + 0.6132 + 0.7424]

 = 0.74614

 Step 2: Deciding the members for group j

 As we can see,

 Only Rjj >Tj, therefore Researcherj belongs to group j.

1. Membership for group k

 Step 1: Calculate the threshold for this group i.e. Tk

 Tk = $ \frac{1}{5}$ [Rki + Rkj + Rkk +Rkn + Rkm]

 = $ \frac{1}{5}$ [0.7110 + 0.6688 + 1 + 0.8786 + 0.8140]

 = 0.81448

 Step 2: Deciding the members for group k

 As we can see,

 Rkk >Tk and Rkn >Tk, therefore Researcherk and Researchern belong to group k.

1. Membership for group n

 Step 1: Calculate the threshold for this group i.e. Tn

 Tn = $ \frac{1}{5}$ [Rni + Rnj + Rnk +Rnn + Rnm]

 = $ \frac{1}{5}$ [0.7502 + 0.6132 + 0.8786 + 1 + 0.9169]

 = 0.83178

 Step 2: Deciding the members for group n

 As we can see,

 Rnk >Tn, Rnn >Tn and Rnm >Tn, therefore Researcherk, Researchern and

 Researcherm belong to group n.

1. Membership for group m

 Step 1: Calculate the threshold for this group i.e. Tm

 Tm = $ \frac{1}{5}$ [Rmi + Rmj + Rmk +Rmn + Rmm]

 = $ \frac{1}{5}$ [0.8064 + 0.7424 + 0.8140 + 0.9169 + 1]

 = 0.85594

 Step 2: Deciding the members for group m

 As we can see,

 Rmn >Tm and Rmm >Tm, therefore Researchern and Researcherm belong to group m.

 Following are the Researcher Groups so formed and their respective members:

 Group i Group j Group k Group n Group m

Researcher k, n and m

Researcher n and m

Researcher i and m

Researcher

 j

Researcher k and n

1. **Construction of Clusters**
2. Total number of clusters

Now as we know total number of clusters i.e. K is equivalent to the minimum number of groups required to cover all the data points. Therefore, K=3. In other words, we can say that there are total three number of clusters required with the centroid as i, j, and n respectively.

 n

 j

 i

 1st cluster 2nd cluster 3rd cluster

1. Assigning points to the closest Centroid

In this step we assign points (researchers m and k) to the closest centroid by taking the proximity measure as the distance between two researchers. Therefore using the formula defined in equation 3.5, we calculate the distance of these two researchers with each of the above researchers:

dki = 0.289

 dkj = 0.3312

 dkn = 0.1214

 Since dkn is minimum, therefore researcher k belongs to the 3rd cluster with centroid as n.

dmi = 0.1936

dmj = 0.2576

dmn = 0.0831

Similarly, Since dmn is minimum, therefore researcher m also belongs to the 3rd cluster with centroid as n.

So, after the first iteration we have the following clusters:

1st cluster 2nd cluster 3rd cluster

 n, k,m

 i

 j

Now, the 2nd iteration begins. We will recompute the centroid of the 3rd cluster.

Distance between each of the two researchers is as follows:

dij = 0.2937; din = 0.2498; djn = 0.3868; dkm = 0.186; dki = 0.289; dkj = 0.3312; dkn = 0.1214; dmi =0.1936; dmj = 0.2576; dmn = 0.0831

Assuming n to be the centroid:

S1= dnm + dnk = 0.1214 + 0.0831= 0.2045

Assuming m to be the centroid:

S2= dmk + dmn = 0.186 + 0.0831= 0.2691

Assuming k to be the centroid:

S3= dkm + dkn = 0.186 + 0.1214= 0.3074

Since S1 is minimum , therefore n remains the centroid.

**4.1.2 Accessing Expertise in Collaborative Interest Group**

Expert from the 1st and 2nd cluster are i and j respectively since there are no more researchers in that cluster.

Expert from the 3rd cluster

Using the formula defined in equation 3.6, we find the Researcher n’s level of expertise as:-

 en = $ \frac{1}{3}$ [acn1+ acn2 + acn3]

 = $\frac{1}{3}$ [acnn+ acnm + acnk]

As we know,



In our experiment, Let T1 = average of Rnn, Rnm and Rnk

= $\frac{1}{3}$ [Rnn+ Rnm + Rnk]

 = $\frac{1}{3}$ [ 1 + 0.9169 + 0.8786]

 = 0.9318

Now, since

Rnn =1 > 0.9318, therefore acnn = 1

Rnm= 0.9169 < 0.9318, therefore acnm = 0

Rnk= 0.8786 < 0.9318, therefore acnk = 0

Therefore, en = $\frac{1}{3}$ [1+ 0 +0 ] = 0.333

Similarly;

* Researcher k’s level of expertise = $\frac{1}{3}$ [ackk+ ackn+ ackm]

 = $ \frac{1}{3}$ [1+0+0]

 = 0.3333

* Researcher m’s level of expertise = $ \frac{1}{3}$ [acmk+ acmn+ acmm]

 = $\frac{1}{3}$ [0+0+1]

 = 0.3333

Since the expertise value of all the three Researchers viz. n, k and m belonging to the 3rd cluster is same, therefore we conclude by saying that all the three researchers are expert in this particular field and that no one is better than the other.

Alternatively, we can also consider the expert of a particular field as the centroid of that corresponding cluster. This is because centroid is a point which has got the maximum similarity in respect to the cluster in which it belongs to. In other words, we can say that the researcher corresponding to that particular centroid will possess a high level of knowledge in a particular domain and will be a reliable source of relevant resources and information.

Thus, we can consider the Researcher n to be an expert of the 3rd cluster.

**4.1.3 Recommendation Module**

Let us consider the following review as an example. We have assumed that a person who has an expertise in cars has given the following review about a car say “Innova”.

This car is a complete blend of great power and style, with exciting features. It has very good fuel efficiency and engine is pretty impressive too. It has got beautiful interiors and the compact dimensions make it an excellent traffic warrior. Of course it has its faults – the colours are boring. The headlights are not very strong and rear seats are less comfortable. The ride is not too bad, but there is a little stiffness and it crashes over sharp bumps. The size of this car is never big and this makes its price quite reasonable and affordable. It demands hardly any maintenance and its performance and safety are also quite amazing. Not much of car service is required.

* + - 1. **Feature Extraction**
1. POS Tagging
2. Pre-Processing
3. POS Tagging

Each of the sentences along with their POS tag information is saved in the Review Database.

 <S>

 <NG><WC = ‘This’>*This*</W><WC = ‘car’ </W></NG>

 <V><WC = ‘is’>*is* </W></V>

 <ART><WC=’a’>*a*</W></ART>

 <AG><WC=’complete’> *complete*</W><WC=’blend’> </W> </AG>

 <P><WC=’of’>*of*</W></P>

<A><WC=’great’>*great*</W></A>

<N><WC = ‘power’>*power*</W></N>

<CONJ><WC=’and’>*and*</W></CONJ>

<N><WC=’style’>*style/<W>*</N><WC= ‘,’>,</W>

<P><WC=’with’>*with</W>*</P>

<A><WC=’exciting’> *exciting*</W></A>

<N><WC=’features’>*features*</W></N><WC=’.’>.</W>

</S>

<S>

<N><WC = ‘It’>*It*</W></N>

*<V* ><WC = ‘has’>*has*</W></V>

<ART><WC = ‘a’> *a*</W></ART>

<AG><WC =’‘very’> *very*</W><WC = ‘good’> *good*</W></AG>

<NG><WC = ‘fuel’> *fuel*</W><WC = ‘efficiency’> *efficiency*</W></NG>

<CONJ><WC = ‘and’> *and*</W></CONJ>

<N><WC = ‘engine’> *engine*</W></N>

<V><WC = ‘is’> *is*</W></V>

<AG><WC = ‘pretty’> *pretty*</W><WC = ‘impressive’> *impressive*</W></AG>

<P><WC = ‘too’> *too*</W></P><WC=’.’>.</W>

</S>

<S>

<N ><WC = ‘It’> *It*</W><WC = ‘car’ </W></NG>

<VG><WC = ‘has’> *has* </W><WC = ‘car’ </W><WC = ‘got’> *got*</W></AG>

<A ><WC = ‘beautiful’> *beautiful*</W></A>

<N><WC = ‘interiors’> *interiors*</W></N>

<CONJ><WC = ‘and’> *and*</W></CONJ>

<ART><WC = ‘the’> *the*</W></ART>

<A><WC = ‘compact’> *compact*</W></A>

<N ><WC = ‘dimensions’> *dimensions*</W></N>

<V><WC = ‘make’> *make*</W></V>

<N><WC = ‘it’> *it*</W></N>

<ART><WC = ‘an’> *an*</W></ART>

<A><WC = ‘excellent’> *excellent*</W></A>

<NG><WC=‘traffic’>*traffic*</W><WC=‘warrior’> *warrior*</W></NG><WC=’.’>.</W>

</S>

<S>

<P ><WC = ‘Of’> *Of*</W><WC = ‘course’> *course* </W></P>

<N ><WC = ‘it’> *it* </W></N>

<V ><WC = ‘has’> *has*</W</V>

<NG><WC = ‘its’> *its*</W><WC = ‘faults’> *faults*</W></NG><WC=’-’>-</W>

<ART ><WC = ‘the’> *the*</W></ART>

<N><WC = ‘colours’> *colours*</W></N>

<V ><WC = ‘are’> *are*</W></V>

<A><WC = ‘boring’> *boring*</W><WC = ‘,’ >,</W></A>

<ART><WC = ‘the’> *the*</W></ART>

<N><WC = ‘headlights’> *headlights* </W></N>

<V><WC = ‘are’> *are*</W></V>

<AG><WC=‘not’>*not*</W><WC=‘very’>*very*</W><WC= ‘strong’> *strong*</W></AG>

<CONJ><WC = ‘and’> *and*</W></CONJ>

<N><WC = ‘rear-seats’> *rear- seats*</W></N>

<V><WC = ‘are’> *are*</W></v>

<AG><WC=‘less’>*less*</W><WC=‘comfortable’> *comfortable*</W></AG><WC=’.’>.</W>

</S>

<S>

<ART ><WC = ‘The’> *The* </W></ART>

<N ><WC = ‘ride’> *ride*</W></N>

<V><WC = ‘is’> *is*</W></V>

<AG><WC = ‘not’> *not*</W><WC = ‘too’> *too*</W><WC = ‘bad’> *bad*</W></AG>

<WC = ‘,’>, </W>

<CONJ ><WC = ‘but’> *but*</W></CONJ>

<P><WC = ‘There’> *there*</W></P>

<V><WC = ‘is’> *is*</W></V>

<ART><WC = ‘a’> *a*</W></ART>

<A><WC = ‘little’> *little*</W></AG>

<N><WC = ‘stiffness’> *stiffness*</W></N>

<CONJ><WC = ‘and’> *and*</W></CONJ>

<N><WC = ‘it’> *it*</W></N>

<V><WC = ‘crashes’> *crashes*</W></V>

<P><WC = ‘over’> *over*</W></P>

<A><WC = ‘sharp’> *sharp*</W></A>

<N><WC = ‘bumps’> *bumps*</W></N><WC=’.’>.</W>

</S>

<S>

<ART><WC = ‘The’> *The* </W></ART>

<N><WC = ‘size’> *size*</W></N>

<P><WC = ‘of’> *of*</W></P>

<NG ><WC = ‘This’> *this*</W><WC = ‘car’>car </W></NG>

<V ><WC = ‘is’> *is*</W></V>

<NG ><WC = ‘never’> *never*</W><WC = ‘big’> *big*</W></NG>

<CONJ ><WC = ‘and’> *and*</W></CONJ>

<N ><WC = ‘this’> *this*</W></N>

<V ><WC = ‘makes’> *makes*</W></VG>

<NG><WC = ‘its’> *its*</W><WC = ‘price’> *price*</W></NG>

<AG ><WC = ‘quite’> *quite*</W><WC = ‘reasonable’> *reasonable*</W></AG>

<CONJ ><WC = ‘and’>*and*</W></CONJ>

<A><WC = ‘affordable’> *affordable*</W></A><WC=’.’>.</W>

</S>

<S>

<N><WC = ‘It’> *It*</W></N>

<V><WC = ‘demands’>*demands*</W></V>

<AG><WC = ‘hardly’> *hardly*</W><WC = ‘any’> *any*</W></AG>

<N ><WC = ‘maintenance’> *maintenance*</W></N>

<CONJ><WC = ‘and’> *and*</W></CONJ>

<N><WC = ‘its’> *its*</W></N>

<N><WC = ‘performance’> *performance*</W></N>

<CONJ ><WC = ‘and’>*and* </W><WC = ‘car’ </W></CONJ>

<N><WC = ‘safety’>*safety*</W></N>

<V><WC = ‘are’> *are*</W></V>

<CONJ><WC = ‘also’> *also*</W></CONJ>

<AG><WC=‘quite’>*quite*</W><WC=‘amazing’>*amazing*</W></AG> <WC=’.’>.</W>

</S>

<S>

<AG><WC=’Not’> *Not* </W> <WC=’much’> *much* </W></AG>

<P><WC=’of’> *of* </W> </P>

<NG><WC=’car’> *car*</W> <WC=’service’> *service*</W></NG>

<VG><WC=’is’> *is* </W> <WC=’required’> *required* </W></VG> <WC=’.’>.</W>

</S>

1. Pre-processing

 A Transaction File is created which consists of pre-processed noun/noun-phrases of the sentences in the review database.

Following is the list of such noun/noun phrases:-

This car

Power

Style

Features

It

Fuel efficiency

Engine

Interiors

Dimensions

Traffic warrior

Its faults

Colours

Headlights

Rear-seats

Ride

Stiffness

Bumps

Size

This

Its price

Its performance

Safety

Maintenance

Car service

* + - 1. **Opinion Direction Identification**
1. Opinion Word Extraction
2. Opinion Word Orientation
3. Opinion Word Extraction

Since, we have considered the opinion words as the combination of the adjectives along with their adverbs which is collectively being called as an Adjective-Group (AG);

Therefore, we extract such an Adjective Group from the sentences in the review database.

Following is the list of our extracted opinion words :-

OW1 = complete blend

OW2 = great

OW3 = exciting

OW4 = very good

OW5 = pretty impressive

OW6 = beautiful

OW7 = compact

OW8 = excellent

OW9 = boring

OW10 = not very strong

OW11 = less comfortable

OW12 = not too bad

OW13 = little

OW14 = sharp

OW15 = never big

OW16 = quite reasonable

OW17 = affordable

OW18 = hardly any

OW19 = quite amazing

OW20 = not much

1. Opinion Word Orientation

The strength of opinion word is calculated based on the adjective polarity and the adverb strength.

1. **Adjective Polarity**:

We use a set of seed adjectives whose orientations we know, & then grow this set by searching in the WordNet. We consider the following initial Adjective Seed-List:-

Adjective Seed- List =

 (great, exciting, good, impressive, beautiful, strong, comfortable, fast, big,

 excellent, reasonable, affordable, compact, amazing, blend, sharp,

 bad, boring, nasty, flow, wrong, poor, awful, scary, dull, inferior, dirty,

 sick, unfortunate, ridiculous, huge)

Among these few adjectives considered, few adjectives have a positive orientation while the few of these have a negative orientation.

Adjectives which have a positive orientation are:

 Great, exciting, good, impressive, beautiful, strong, comfortable, fast, big, excellent,

 reasonable, affordable, compact, amazing, blend

Adjectives which have a negative orientation are:

Sharp, bad, boring, nasty, flow, wrong, poor, awful, scary, dull, inferior, dirty, sick,

 unfortunate, ridiculous, huge

1. **Adverb Strength** :

We manually mark the strengths of a few frequently used adverbs with values ranging from -1 to +1 based on our intuitions. We consider the most frequently used adverbs (for our illustration) along with their strength as herein below:-

|  |  |
| --- | --- |
| ADVERB | STRENGTH |
| CompleteMostExtremelyTooVeryMoreMuchAnyQuitePrettyLittleLessNotNeverhardly | +10.90.70.60.40.30.20.1-0.2-0.3-0.4-0.6-0.8-0.9-1  |

1. **Opinion Word Strength** :

 The strength of each opinion word is given by the formula defined in equation 3.7 as:

S(OWi) = P(adji) . S(advi)

So, the following list gives the list of strength of each opinion word extracted in the previous step:

S ( OW1) = S (complete blend) = 1 x 1 = 1

S (OW2) = S (great) = 0.5 x 1 = 0.5

S (OW3) = S (exciting) = 0.5 x 1 = 0.5

S (OW4) = S (very good) = 0.4 x 1 = 0.4

S (OW5) = S (pretty impressive) = (-0.3) x1 = -0.3

S (OW6) = S (beautiful) = 0.5 x 1 = 0.5

S (OW7) = S (compact) = 0.5 x 1 = 0.5

S (OW8) = S (excellent) = 0.5 x 1 = 0.5

S (OW9) = S (boring) = 0.5 x (-1) = -0.5

S (OW10) = S (not very strong) = (-0.8) x 0.4 x 1 = -0.32

S (OW11) = S (less comfortable) = (-0.6) x 1 = -0.6

S (OW12) = S (not too bad) = (-0.8) x 0.6 x (-1) = -0.48

S (OW13) = S (little) = (-0.4) x 1 =-0.4

S (OW14) = S (sharp) = 0.5 x (-1) =-0.5

S (OW15) = S (never big) = (-0.9) x 1 = -0.9

S (OW16) = S (quite reasonable) = (-0.2) x 1 = -0.2

S (OW17) = S (affordable) = 0.5 x 1 = 0.5

S (OW18) = S (hardly any) = (-1) x 0.1 x 1 = -0.1

S (OW19) = S (quite amazing) = (-0.2) x 1 = -0.2

S (OW20 )= S (not much) = (-0.8) × 0.2 x 1 = -0.16

We observe here that the strength of each opinion word i.e.$ S(OW\_{i})$ lies in the range of -1 to +1 as the value of P(adji) is either -1 or +1 and the value of S(advi) ranges from -1 to +1.

Hence, $ S(OW\_{i})$ also ranges from -1 to +1.

**4.1.3.3 Review Orientation**

The overall strength of a Review R is given by the formula defined in equation 3.8 as:

S (R) = $\frac{1}{\left|OW\left(R\right)\right|}\*\sum\_{i=1}^{|OW(R)|}S(OW\_{i})$

 = ($4.88-4.18)/20$

 = 0.70 / 20

 = 0.035

Since the overall strength of the review is calculated as 0.035, which is very close to 0, therefore

we can conclude by saying that this review gives a Neutral Opinion about the car “Innova”.

i.e. As S(R)= 0.035 ~ 0 ;

 Hence, the review has a Neutral Orientation.