Journal of
Data Acquisition and Processing

ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.98549727

A STUDY ON SENTIMENT ANALYSIS USING FLEISS KAPPA STATISTIC FOR ESTIMATING A MEASURE OF AGREEMENT IN ENGINEERING DISCIPLINES

A. Vijay Bharath[1], A. Shanthini[2] A.Subbarayan[3]

[1] Department of Data Science and Business Systems, SRM Institute of Science and Technology, Tamilnadu, India.

[2] Department of Data Science and Business Systems, SRM Institute of Science and Technology, Tamilnadu, India.

[3] Department of Mathematics, Dr. MGR Educational and Research Institute, Tamilnadu, India

ABSTRACT

The classification of subjects and the methods of Measurement is determined by the application field. Typically, the rater's are classified according to their skills and expertise. Depending on the nature of the study the number of raters and the number of subjects to be studied are determined. As a part of this study, the authors attempted to evaluate the nominal scale agreement between a fixed number of raters by using Fleiss Kappa statistic.. In this study the rates are the students who are studying different subjects in engineering disciplines in a higher education institute. The basic data are obtained from the students feedback returns submitted by the students at the end of the semester. A detailed analysis is carried out for four engineering disciplines and Fleiss Kappa Statistic and related computations are made. The study has clearly revealed that the estimated value of the measure of agreement is above 0.6000 in respect of all the four engineering disciplines. There is some difference between the estimated value of agreement measurement between engineering disciplines. The reason for the difference may be due to the personal opinion of the students pursuing their engineering discipline and the same form the solid basis for Sentiment Analysis.

Keywords: *Sentiment Analysis, Nominal Scale, Fleiss Kappa Statistic, Coefficient of Agreement.*

1. INTRODUCTION

1.1 Basics of Sentiment Analysis

In Sentiment Analysis the researcher attempts to study the computational aspects involving opinions, attitudes, sentiments, emotions etc., this primarily involves products, films, entities, occasions, issues, subjects and their respective features. Sentiment mining, survey mining, text mining, opinion extraction subjectivity investigation, emotion examination and so forth. Sentiment analysis is the assessment of the certain information that is extracted. Recent research into the analysis of sentiment has provided a variety of methods for extracting and analyzing emotions. In the present study the authors have made an attempt for measuring the nominal scale agreement of raters which will form the basis for Sentiment Analysis.

1.2 Objectives of the Study

The objectives of the study are:-

(i) To discuss in detail agreement analysis aspects for the data relating to categories of subjects offered under four engineering disciplines in a Institute of Higher Education system.

(ii) To compute the Fleiss Kappa Statistic ($\hat{k}_{\mathrm{mc}}$) for measuring the nominal size agreement between a fixed pair of reviewers to rate different types of items.

(iii) To compare the estimated value of Fleiss Kappa Statistic ($\hat{k}_{\mathrm{mc}}$) in respect of the courses under four engineering disciplines.

In section 2 the aspects relating to Agreement Analysis are presented. An updated literature review in respect of agreement analysis is also given in this section. The methodological aspects of the study and related results are discussed in Section 3. Section 4 deals with empirical analysis based on a dataset relating to agreement aspects. Findings and conclusions based on the study are given in Section 5.

2. Agreement Analysis and Related Aspects

2.1 Need for agreement analysis with reference related to Sentiment Analysis.

In the present research we focus mainly to study the computational aspects involving opinions of students relating to agreement. Sentiment Analysis is the assessment of certain information extracted. Recent research into the analysis of sentiment has provided a variety of methods for extracting and analyzing opinions involving agreement aspects.

- Agreement between two or more measuring devices is of prime importance to Statisticians,
- Psychologists, Clinicians, Epidemiologists, and many other Scientists. There has been a
- substantial amount of research in the area of agreement, particularly when subjects are assessed
- on a categorical scale. The level of complexity of this study varies from simple and practical
- implementation to very complex routine implementation by scientists from different fields.

2.2 Agreement Analysis Based on Continuous Assessment (CA) in Engineering Disciplines.

CA is a vital part of teaching and learning. It is planned process of gathering, identifying and interpreting information of the learners. The purpose of CA are primarily related to enhancement of student learning and improvement of faculties teaching skills. At the end of the course the students are asked to give their feedback about the subjects undertaken in their engineering disciplines.

In this study we have taken the primary data on a sample based from the Student Feedback Returns (SFR). Based on the student's assessment we have three categories viz. **Excellent (E)**, **Good (G)**, and **Average (A)** in respect of each sampled student for four subjects under four different engineering disciplines. It is significant to note that the term agreement refers to the student's assessment in respect of the subjects studied. The details of the subjects and the engineering disciplines are presented in the subsequent sections.

2.3 A Review Measures of Agreement Analysis in Research Studies.

Scott (1955) has made pioneering contributions for the study of reliability of content analysis for the nominal scale based on pi statistic. Cohen (1960) derived a coefficient of agreement for nominal scales and this is popularly known as Cohen's Kappa Statistic. This measure assumes two raters rating a set of items. Fleiss, Cohen and Everitt (1969) studied in detail the properties of Kappa and weighted Kappa.

Fleiss (1971) introduced the statistic Kappa for measuring the nominal scale agreement between a fixed numbers of raters and generalized the same. The generalized measure can be used for measuring agreement among a constant number of raters. It is important to note that there is no connection between the raters in making the judgment in respect of the various subjects. Landis and Koch (1977) have further studied the associated aspects relating to chance corrected measure of overall agreement of Fleiss (1971).

Agresti (1992) presented a survey of statistical modelling patterns of observer agreement and disagreement for categorical responses (nominal and ordinal scales). Shoukri (2011) discussed in detail the computational procedure for computing Fleiss Kappa Statistic ($\hat{k}_{\mathrm{mc}}$). Shiv Gautam (2014) proposed a new method for evaluating agreement among multiple raters.

3. An Approach for the Study of Agreement Analysis

3.1 Fleiss Kappa as a Basic Measure of Agreement

Fleiss Kappa is a statistical measure for assessing the reliability of agreement between a fixed numbers of raters when assigning categorical ratings to a number of items. This measure computes the degree of agreement in classification over that which could be expected by chance. Scott's pi (1955) and Cohen's Kappa (1960) are used for assessing the agreement between two raters. It is vital to note that Fleiss Kappa analysis is a statistical technique that calculates categorical rankings for a fixed number of items provided a random selection of raters is maintained for each item.

3.2 Fleiss Kappa and its related concepts

Fleiss generalized Kappa to the case where each of the subjects is rated on a nominal scale was rated by the same number of raters, but where the raters for one subject are not necessarily the same as those rating another.

3.2.1 Formula for the Computational of Multiple Raters for the Multiple Categories

Let k_{ij} be the number of raters who assign the ith subject to the jth category (i = 1, 2, 3,k, j = 1, 2, 3,c)

Define

$$p_j = \frac{1}{nk} \sum_{i=1}^{k} kij \qquad \ldots (1)$$

Here p_j is the proportion of all assignments to the jth category.

The chance corrected measure of overall agreement is given by

$$\hat{}k_{mc} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{c} k_i j^2 - kn\{1+(n-1)\sum_{j=1}^{c} p_j^2\}}{kn(n-1)(1-\sum_{j=1}^{c} p_j^2)} \qquad \ldots (2)$$

(the subscript ‘mc’ denotes for multiple categories)

For computational purpose, we write equation 1 as follows.

$$\hat{}k_{mc} = \frac{p_o - p_e}{1 - p_e} \qquad \text{.. . (3)}$$

where,

$$p_o = \frac{\sum_{i=1}^{k} \sum_{j=1}^{c} k_i j^2 - nk}{kn(n-1)} \ and \ p_e = \sum_{j=1}^{c} p_j^2 \ where \ p_j = \frac{1}{nk} \sum_{j=1}^{c} k_i j$$

The factor p$_o$ - p$_e$ denotes the degree of agreement actually achieved and the factor 1 - p$_e$ gives the degree of agreement that is attainable above chance. If the raters are in complete agreement, then $\hat{k}_{mc}$ = 1.

Asymptotic expression of variance of $\hat{k}_{mc}$ has been derived by Davis and Fleiss (1982) (when $\hat{k}_{mc}$ = 0).

Woolson (1987) derived the estimated variance of $\hat{k}_{mc}$ and it is given by

$$Var(\hat{}k_{mc}) = \frac{2}{kn(n-1)} \left(\frac{p_e - (2n-3)p_e^2 + (2n-1)\sum_{j=1}^{c} p_j^3}{1 - p_e} \right) \qquad \text{. . . . . (4)}$$

The approximate 95% confidence interval for $\hat{k}_{mc}$ is computed based on:

$$\hat{}k_{mc} \pm 1.96 \sqrt{Var(\hat{}k_{mc})} \qquad \text{. . . . . (5)}$$

4. Dataset for the Analysis of Multiple Raters in respect of Multiple categories

4.1 Engineering Disciplines and the subjects studied by students

In this research the researcher has used random sampling method for the selection of students

who are identified as raters. It is important to note that the students submit their feedback at

the end of the course in respect of the subjects studied and categorized them **as Excellent (E),**

Good (G) and Average (A).

4.1.1 Engineering Disciplines Considered for the Study

The disciplines considered for the study are Computer Science and Engineering (CSE), Electronics and Communication Engineering (ECE), Mechanical Engineering (ME) and Biotechnology Engineering (BTE). We have taken a sample of 30 feedback returns for each discipline. It is to be noted that the subjects studied under different engineering disciplines are not common. The engineering disciplines and the subjects studied by the students are given in the following Table 1.

Table 1. Engineering Disciplines and the list of Subjects

Engineering Discipline	Subjects Studied	No. of feedback returns
Computer Science and Engineering (CSE)	S1: Biology S2: Data structure and Algorithm S3: Object Oriented design and Programming S4: Computer Organization and Architecture	30
Electronics and Communication Engineering (ECE)	S1: Environmental Science S2: Electronic Devices S3: Digital Electronics Principles S4: Signal and Systems	30
Mechanical Engineering (ME)	S1: Environmental Science S2:Transform and Boundary value Problems S3: Thermodynamics S4: Fluid mechanics	30
Biotechnology Engineering (BTE)	Human Physiology and Health Biochemistry Cell Biology Microbiology	30

4.2 Empirical Analysis: Computer Science and Engineering

The categories identified by the raters (students) in respect of the subjects are given in Table 2.

Table 2. Data on 30 Raters (students) in respect of 4 subjects

Student Feedback	**S1**	**S2**	**S3**	**S4**
1	G	E	E	G
2	A	A	A	A
3	G	G	G	G
4	E	E	E	E
5	G	G	G	G
6	E	E	E	E
7	A	E	E	A
8	G	E	E	G
9	G	G	G	G
10	G	G	A	G
11	E	E	E	A
12	A	E	G	G
13	E	E	E	E
14	G	G	A	A
15	E	E	E	E
16	A	A	A	A
17	G	E	G	G
18	E	E	E	E
19	E	E	E	E
20	E	E	E	E
21	E	E	E	E
22	E	E	E	A

23	G	A	A	A
24	A	E	G	A
25	G	G	G	A
26	E	E	E	E
27	E	E	E	E
28	G	G	G	G
29	E	E	E	E
30	A	A	A	A

Table 3 shows the values for k_{ij} in respect of four subjects

Table 3 Values for k_{ij}

Student feedback I	J			$k^{c}_{j=1}ij^2$
	E	G	P	
1	2	2		8
2			4	16
3		4		16
4	4			16
5		4		16
6	4			16
7	2		2	8
8	2	2		8
9		4		16
10		3	1	10
11	3		1	10
12	1	2	1	6
13	4			16

14		2	2	8
15	4			16
16			4	16
17	1	3		10
18	4			16
19	4			16
20	4			16
21	4			16
22	3		1	10
23		1	3	10
24	1	1	2	6
25		3	1	10
26	4			16
27	4			16
28		4		16
29	4			16
30			4	16
Total	59	35	26	392

For the above data p_1 = 59/ (30) (4) = 0.4916, p_2 = 35/ (30) (4) = 0.2916, p_3 = 26/ (30) (4) = 0.2166

$p_e = \sum_{i=1}^{3} pi^2 = (0.4916)^2 + (0.2916)^2 + (0.2166)^2 = 0.3735$

$$p_o = \frac{\sum_{i=1}^{k}\sum_{j=1}^{c} k_ij^2 - nk}{kn(n-1)}$$

= **(**392 – 30 x 4) / (30 x 4 x 3) = 0.7555

$$\wedge k_{mc} = \frac{p_o - p_e}{1 - p_e}$$

= (0.7555 – 0.3735) / (1 – 0.3735) = 0.6097

$\sum_{j=1}^{3} pj^3 = (0.4916)^3 + (0.2916)^3 + (0.2166)^3 = 0.1529$

Substituting in equation (4) we get Variance of $^{\wedge}K_{mc} = 0.0083$

Approximate confidence interval for $^{\wedge}K_{mc}$

The approximate expression for confidence interval based on equation 5 is computed as:

$$p = 0.6097 \pm 1.96\surd 0.0083$$
$$= (0.4311, 0.7883)$$

4.3 Empirical Analysis: Electronics and communication Engineering

The categories identified by the student in respect of the subjects are given in Table 4.

Table 4. Data on 30 Raters (students) in respect of 4 subjects

Student Feedback	**S1**	**S2**	**S3**	**S4**
1	G	G	G	G
2	G	G	G	G
3	E	E	E	E
4	E	E	E	E
5	G	A	A	G
6	E	E	E	E
7	E	E	A	E
8	E	E	E	E
9	G	G	G	G
10	E	E	E	E
11	G	G	G	G
12	G	G	A	E
13	E	E	E	E
14	E	E	E	E
15	E	E	G	E
16	E	E	E	E
17	E	G	E	E
18	E	E	E	E
19	E	E	E	E

20	G	G	G	G
21	E	E	E	E
22	E	E	G	E
23	G	G	G	G
24	E	E	E	E
25	G	G	A	G
26	G	G	A	G
27	E	E	E	E
28	E	E	G	E
29	E	E	E	E
30	E	E	E	E

Table 5 shows the values for k_{ij} in respect of four subjects

Table 5 Values for k_{ij}

Student feedback i	j			$k_{j=1}^{c} i j^2$
	E	G	P	
1		4		16
2		4		16
3	4			16
4	4			16
5		2	2	8
6	4			16
7	3		1	10
8	4			16
9		4		16
10	4			16

11		4		16
12	1	2	1	6
13	4			16
14	4			16
15	3	1		10
16	4			16
17	3	1		10
18	4			16
19	4			16
20		4		16
21	4			16
22	3	1		10
23		4		16
24	4			16
25		3	1	10
26		3	1	10
27	4			16
28	3	1		10
29	4			16
30	4			16
Total	76	38	6	420

For the above data p_1 = 76/ (30) (4) = 0.6333, p_2 = 38/ (30) (4) = 0.3166, p_3 = 6/ (30) (4) = 0.0500

$p_e = \sum_{i=1}^{3} pi^2 = (0.6333)^2 + (0.3166)^2 + (.0500)^2 = 0.5037$

p_0 = (420 – 30 x 4) / (30 x 4 x 3) = 0.8333

$^{\wedge}K_{mc}$ = (0.8333 – 0.5037) / (1 – 0.5037) = 0.6641

$$\sum_{j=1}^{3} pj^3 = (0.6333)^3 + (0.3166)^3 + (0.0500)^3 = 0.2857$$

Substituting in equation (4) we get Variance of ^K$_{mc}$ = 0.0214

Approximate confidence interval for ^K$_{mc}$

The approximate expression for confidence interval based on equation 5 is computed as:

p = 0.6641 ± 1.96√0.0214
= (0.3776, 0.9506)

4.4 Empirical Analysis: Mechanical Engineering

The categories identified by the student in respect of the subjects are given in Table 6.

Table 6. Data on 30 Raters (students) in respect of 4 subjects

Student Feedback	S1	S2	S3	S4
1	G	G	G	G
2	E	E	E	E
3	G	G	G	G
4	E	E	E	E
5	E	E	E	E
6	A	A	A	A
7	E	E	E	E
8	G	G	G	G
9	G	G	E	G
10	G	G	E	G
11	A	A	A	A
12	E	G	E	E
13	A	A	A	A
14	E	E	E	E
15	G	E	E	E
16	A	E	G	G
17	E	E	E	E
18	E	G	E	E

19	E	E	E	E
20	A	A	E	E
21	G	E	E	E
22	E	E	E	E
23	G	G	G	G
24	A	A	A	A
25	E	E	E	E
26	E	E	E	E
27	E	E	E	E
28	E	E	E	E
29	G	G	G	G
30	G	E	E	E

Table 7 shows the values for k_{ij} in respect of four subjects

Table 7 Values for k_{ij}

Student feedback i	j			$k^{c}_{j=1}ij^{2}$
	E	G	P	
1		4		16
2	4			16
3		4		16
4	4			16
5	4			16
6			4	16
7	4			16
8		4		16
9	1	3		10

10	1	3		10
11			4	16
12	3	1		10
13			4	16
14	4			16
15	3	1		10
16	1	2	1	6
17	4			16
18	3	1		10
19	4			16
20	2		2	8
21	3	1		10
22	4			16
23		4		16
24			4	16
25	4			16
26	4			16
27	4			16
28	4			16
29		4		16
30	3	1		10
Total	68	33	19	420

For the above data p_1 = 68/ (30) (4) = 0.5666, p_2 = 33/ (30) (4) = 0.2750, p_3 = 19/ (30) (4) = 0.1583

$p_e = \sum_{i=1}^{3} pi^2 = (0.5666)^2 + (0.2750)^2 + (0.1583)^2 = 0.4216$

p_0 = (420 – 30 x 4) / (30 x 4 x 3) = 0.8333

$^\wedge K_{mc}$ = (0.8333 – 0.5037) / (1 – 0.5037) = 0.7117

$$\sum_{j=1}^{3} pj^3 = (0.5666)^3 + (0.2750)^3 + (0.1583)^3 = 0.2064$$

Substituting in equation (4) we get Variance of $^\wedge K_{mc}$ = 0.0128

Approximate confidence interval for $^\wedge K_{mc}$

The approximate expression for confidence interval based on equation 5 is computed as:

p = 0.7117 ± 1.96√0.0128
= (0.4901, 0.9333)

4.5 Empirical Analysis: Biotechnology Engineering

The categories identified by the student in respect of the subjects are given in Table 8.

Table 8. Data on 30 Raters (students) in respect of 4 subjects

Student Feedback	S1	S2	S3	S4
1	G	G	G	G
2	E	E	E	E
3	G	G	G	G
4	E	E	E	E
5	G	G	E	E
6	G	G	G	G
7	E	E	E	E
8	E	E	E	E
9	E	E	E	E
10	G	G	G	G
11	G	G	G	G
12	E	E	E	E
13	G	G	G	G
14	E	E	E	E
15	E	G	E	E
16	E	E	E	E
17	E	E	E	E

18	E	E	E	E
19	G	G	E	G
20	E	E	E	E
21	G	G	G	G
22	G	G	G	G
23	E	E	E	E
24	E	E	E	E
25	E	E	E	E
26	E	E	E	E
27	G	A	A	A
28	E	E	E	E
29	G	G	G	G
30	E	E	G	E

Table 9 shows the values for k_{ij} in respect of four subjects

Table 9 Values for k_{ij}

Student feedback i	j			$k^{c}_{j=1} i j^{2}$
	E	G	P	
1		4		16
2	4			16
3		4		16
4	4			16
5	2	2		4
6		4		16
7	4			16
8	4			16

9	4			16
10		4		16
11		4		16
12	4			16
13		4		16
14	4			16
15	3	1		10
16	4			16
17	4			16
18	4			16
19	1	3		10
20	4			16
21		4		16
22		4		16
23	4			16
24	4			16
25	4			16
26	4			16
27		1	3	10
28	4			16
29		4		16
30	3	1		10
Total	73	44	3	444

For the above data p_1 = 73/ (30) (4) = 0.6083, p_2 = 44/ (30) (4) = 0.3666, p_3 = 3/ (30) (4) = 0.0250

$p_e = \sum_{i=1}^{3} pi^2 = (0.6083)^2 + (0.3666)^2 + (0.0250)^2 = 0.5049$

$p_0 = (444 - 30 \times 4) / (30 \times 4 \times 3) = 0.9000$

$^{\wedge}K_{mc} = (0.9000 - 0.5049) / (1 - 0.5049) = 0.7980$

$$\sum_{j=1}^{3} pj^3 = (0.6083)^3 + (0.3666)^3 + (0.0250)^3 = 0.2742$$

Substituting in equation (4) we get Variance of $^{\wedge}K_{mc} = 0.0198$

Approximate confidence interval for $^{\wedge}K_{mc}$

The approximate expression for confidence interval based on equation 5 is computed as:

$p = 0.7980 \pm 1.96\sqrt{0.0198}$

$= (0.5223, 1.0737)$

5. Findings and Conclusions

The study has revealed the following results in respect of the four Engineering Disciplines.

A Comparison of the results obtained are given in the following Table 10.

Table 10 Estimation of $^{\wedge}K_{mc}$

S.No	Department	$^{\wedge}K_{mc}$	$V(^{\wedge}K_{mc)}$	Confidence Interval for $^{\wedge}K_{mc}$
1	CSE	0.6097	0.0083	(0.4311, 0.7883)
2	ECE	0.6641	0.0214	(0.3766, 0.9506)
3	ME	0.7117	0.0128	(0.4901, 0.9333)
4	BTE	0.7980	0.0198	(0.5223, 1.0737)

It is interesting to note that the estimation of the measurements of agreement among the sampled feedback returns of students is above 0.6000. The estimates clearly indicates that they are significant and helps the educational planners to make improvements in the subjects taught for further increase in the measurement of agreement. We observe that there is some difference in the estimate of agreement between engineering disciplines. This may be due to personal opinions of students pursuing their engineering discipline which is the solid base for Sentiment Analysis.

REFERENCES

[1] Scott, W (1955). Reliability of Content Analysis: The Case of Nominal Scale Coding Public Opinion
Quarterly, Vol 19(3), pp 321-325.

[2] Cohen, Jacob (1960). A Coefficient of Agreement for Nominal Scales. Education and Psychological
Measurements, Vol 20(1), pp 37-46.

[3] Fleiss J.L, Cohen, J and Everitt BS (1969). Large Sample Standard Errors of Kappa and Weighted Kappa. Psychological Bulletin, Vol 72, pp 323-327.

[4] Fleiss, J.L (1971). Measuring Nominal Scale Agreement Among Many raters. Psychological Bulletin. Vol 76, pp 378-382.

[5] Landis, J.R. and Koch C.G (1977). A One-way Components of Variance Model for Categorical Data. Biometrics Vol 33, pp 159-174.

[6] Woolson, R, F, (1987). Statistical Methods for the Analysis of Biomedical Data. New York NY: Wiley & Sons.

[7] Agresti, A. (1992). Modelling Patterns of Agreement and Disagreement. Statistical Methods in research. Vol 1, pp 201-218.

[8] Mousumi Banerjee, Michelle Capozzoli, Laura McSweeney, and Debajyoti Sinha, D (1999). Beyond Kappa: A Review of Interrater Agreement Measures. Canadian Journal of Statistics, Vol 27, pp 3-23. https://doi.org/10.2307/3315487.

[9] Shoukri, M.M (2011). Measures of Interobserver Agreement and reliability, second Edition, CRC Press. Taylor & Francis Groups. London, New York.

[10] Shiva Gautam (2014) A- Kappa A Measure of Agreement Among Multiple Raters. Journal of Data Science, Vol 12(4), pp 697-716.