

Association Rule Mining Technique for Psychometric Personality Testing and Behaviour Prediction

Syed Khalid Perwez¹, Hamza Mohd. Zubair^{*2}, Muhammad Rukunuddin Ghalib^{#3}, Kausar Ahmed P^{#4}, Mohammed Iftekhhar^{†5}

¹VIT Business School, ^{*} Dept. of Cognitive Science, [#] School of Computing Science and Engineering, [†]Dept. of Biotechnology

^{1,3,4}VIT University, Vellore, India

²Indian Institute of Technology, Gandhinagar

⁵Dayananda Sagar Institutions, Bangalore

¹khalid_mba@rediffmail.com

²hamzazubair2009@hotmail.com

³ghalib.it@gmail.com

⁴kausarahmed@vit.ac.in

⁵iftubip@gmail.com

Abstract— At the heart of personality psychology lies one single fundamental motive and that is to be able to anticipate how an individual will think, behave and feel at any future instant. Quite unfortunately this field has not been very successful in achieving this. Though this field has given us great insights about the working of the mind, cognitive processes and emotions, it has failed to accomplish its central objective i.e., to predict human behaviour. We propose in this paper a novel technique of predicting human behaviour without the need of any abstraction about the mind or its internal workings. We propose the use of simple and straightforward statistics for this purpose. Applying simple association rule mining on behaviours of thousands of people, association rules having high confidence values can be identified. And based on these rules, strong conclusions can be made in anticipating the behaviour of an individual. An analytical study was conducted on answers provided by 1414 candidates to a 163-question personality survey. The survey was based on the famous questionnaire prepared by Raymond Cattell. This survey was chosen to first try and prove the ambiguity in the current psychological concepts. Following that simple association rule mining was applied on the data to obtain associations between variables. The strongest association obtained with 97.2% confidence was an inter-class association rather than an intra-class association as would be expected from traditional psychology point-of-view.

Keyword- Personality Psychology; Personality Theories; Raymond Cattell's personality factor; Data Mining; Association Rule Mining

I. INTRODUCTION

As our understanding of this universe grows amidst the swelling breadths and depths of our knowledgebase, we are increasingly made aware of the fact that nothing in this universe is compartmentalized. The universe has proved that it cannot be classified or divided into parts rather there is a gradient everywhere and an unclassifiable diversity is present everywhere we look. Earlier we used to divide the heavenly bodies into stars, the sun and the moon. Now we have discovered a variety of heavenly objects of varying sizes and shapes. Life on earth was either plants or animals; the animals were classified into insects, mammals, birds and fishes. One of the first attempts at classifying the living beings by Carl Linnaeus gives a hint of this fact. And now as time passes our classification systems have begun to increase in complexity, now there are amphibians, flightless birds, etc. Today, even the division into flora and fauna is considered artificial. Now this earth is not even classified into the living and the non-living, because we have such things as viruses that blur the border between the living and the non-living.

Man has been on the search for an answer to what personality is from the start of time, but apart from a few useful insights we have not been able to provide a conclusive answer to what a personality is and how it works. No doubt we have spent more than two thousand years already in the hunt for a convincing answer. Yet in this paper we argue that the real motive behind the search for a satisfying answer to what personality actually is, is not greater understanding of this world through a better understanding of how our minds work. The objective is not just to be able to comprehend the complete nature of a mind by understanding various individual personalities; we argue that most likely the real motive behind the urge to make complete sense of human psych stems from the desire to predict. Humans crave for prediction and they try to predict whenever they are able to perceive any sort of patterns. This craving for prediction enforces us to look for the underlying thread of abstract ideas that can explain everything we perceive about an individual's thoughts,

behaviours and feelings.

Yet for the most part in the field of psychology, and in general; people, their personalities and their behaviours are divided into types. The concept of trait theories also suffers from this abstract conceptualization of this diversity in personality into dimensions

Personality psychology is a branch of psychology that studies personality and its individual differences. Its areas of focus include:

- Construction of a coherent picture of the individual and his or her major psychological processes
- Investigation of individual differences
- investigating human nature and human's similarities

"Personality" is defined as a dynamic and organized set of characteristics possessed by a person who uniquely influences his or her cognitions, emotions, motivations, and behaviors in various situations. The word "personality" originates from the Latin *persona*, which means mask. In the theatre of the ancient Latin-speaking world, the mask was not used as a plot device to *disguise* the identity of a character, but instead was a convention employed to represent or *typify* that character. Allport argues [17] that the psychology of personality can make the following contributions,

- predict a person's behavior on the basis of his/her individual characteristics
- discover common traits
- develop general laws as to how an individual's uniqueness comes about
- discover the individual person's own point of view of who she or he is
- codify knowledge as to the nature of human nature
- turn interpersonal impressions into more reliable knowledge
- adequately represent the individual in science, and provide that individual with respect

No doubt predicting what another person is going to think and feel, and how the other person is going to behave has been on our minds since for the first time we started to think as humans. And since that time the field of personality psychology has grown so much that we have different models and a variety of theories that contribute in their own way in understanding the human psych. Yet one thing that we haven't been able to achieve even today is to be able to foretell with a certain amount of confidence as to how a human being is going to behave or think in future[3].

II. RELATED WORK

A. Personality Theories

The study of personality is based on the essential insight that all people are similar in some ways, yet different in others [4][5]. There have been many different definitions of personality proposed. However, many contemporary psychologists agree on the following definition:

Personality is that pattern of characteristic thoughts, feelings, and behaviours that distinguishes one person from another and that persists over time and situations [4].

1. Trait Theories:-

According to the *Diagnostic and Statistical Manual* of the American Psychiatric Association, personality traits are "enduring patterns of perceiving, relating to, and thinking about the environment and oneself that are exhibited in a wide range of social and personal contexts." Theorists generally assume that a) traits are relatively stable over time, b) traits differ among individuals, and c) traits influence behaviour. They consistently are used in order to help define people as a whole. Traits are relatively constant; they do not usually change. Traits are also bipolar; they vary along a continuum between one extreme and the other (e.g. friendly vs. unfriendly) [5][22].

The most common models of traits incorporate three to five broad dimensions or factors. All trait theories incorporate at least two dimensions, extraversion and neuroticism, which historically featured in Hippocrates' humoral theory [6]. Gordon Allport delineated different kinds of traits, which he also called *dispositions*. *Central traits* are basic to an individual's personality, while *secondary traits* are more peripheral. *Common traits* are those recognized within a culture and thus may vary from culture to culture. *Cardinal traits* are those by which an individual may be strongly recognized. In his book, *Personality: A Psychological Interpretation*, Gordon Allport (1937) both established personality psychology as a legitimate intellectual discipline and introduced the first of the modern trait theories [7][20].

Raymond Cattell's research [1][2] propagated a two-tiered personality structure with sixteen "primary factors" (16 Personality Factors) and five "secondary factors." In Cattell's lengthy career, he had written 50 books, 500 journals, and 30 different types of standardized tests. For Cattell, personality itself was defined in terms of behavioral prediction. He defined personality as *that which permits a prediction of what a person will*

do in a given situation.

2. Type Theories:-

Personality type refers to the psychological classification of different types of people. Personality types are distinguished from personality traits, which come in different levels or degrees. For example, according to type theories, there are two types of people, introverts and extroverts. According to trait theories, introversion and extroversion are part of a continuous dimension, with many people in the middle. The idea of psychological types originated in the theoretical work of Carl Jung [8] and William Marston, whose work is reviewed in Dr. Travis Bradberry's *Self-Awareness*. Jung's seminal 1921 book on the subject is available in English as *Psychological Types*. [19]

3. Psychoanalytical Theories:-

Psychoanalytic theories explain human behaviour in terms of the interaction of various components of personality. Sigmund Freud was the founder of this school of thought. Freud drew on the physics of his day (thermodynamics) to coin the term psychodynamics. Based on the idea of converting heat into mechanical energy, he proposed psychic energy could be converted into behaviour. Freud's theory places central importance on dynamic, unconscious psychological conflicts [9].

4. Behaviourist Theories:-

Behaviourists explain personality in terms of the effects external stimuli have on behavior. The approaches used to analyse the behavioural aspect of personality are known as behavioural theories or learning-conditioning theories. These approaches were a radical shift away from Freudian philosophy. One of the major tenets of this concentration of personality psychology is a strong emphasis on scientific thinking and experimentation. This school of thought was developed by B. F. Skinner who put forth a model which emphasized the mutual interaction of the person or "the organism" with its environment. Ivan Pavlov is another notable influence. He is well known for his classical conditioning experiments involving dogs. These physiological studies led him to discover the foundation of behaviourism as well as classical conditioning [10][21].

5. Social Cognitive Theories:-

In cognitive theory, behaviour is explained as guided by cognitions (e.g. Expectations) about the world, especially those about other people. Cognitive theories are theories of personality that emphasize cognitive processes, such as thinking and judging. [24].

6. Humanistic Theories:-

Humanistic psychology emphasizes that people have free will and that this plays an active role in determining how they behave. Accordingly, humanistic psychology focuses on subjective experiences of persons as opposed to forced, definitive factors that determine behaviour. Abraham Maslow and Carl Rogers were proponents of this view, which is based on the "phenomenal field" theory of Combs and Snygg [11]. Rogers and Maslow were among a group of psychologists that worked together for a decade to produce the *Journal of Humanistic Psychology*. This journal was primarily focused on viewing individuals as a whole, rather than focusing solely on separate traits and processes within the individual.

7. Biopsychological Theories:-

Biology plays a very important role in the development of personality. The study of the biological level in personality psychology focuses primarily on identifying the role of genetic determinants and how they mold individual personalities [12][23][25].

B. Raymond Cattell's 16 Personality Factor Questionnaire

The Sixteen Personality Factor Questionnaire (or 16PF) given in Table 1 in supplementary data, is a multiple-choice personality questionnaire which was developed over several decades of research by Raymond B. Cattell, Maurice Tatsuoka and Herbert Eber. Beginning in the 1940s, Cattell used the new techniques of factor analysis (based on the correlation coefficient) in an attempt to try to discover and measure the fundamental traits of human personality [1]. The questionnaire measures the 16 primary traits shown in Table 2 in supplementary data, which have become popularized by other authors in recent years. From early in his research, Cattell found that the structure of personality was multi-level and hierarchical, with a structure of interdependent primary and secondary level traits [2]. The sixteen primary factors were a result of factor-analysing hundreds of measures of everyday behaviours to find the fundamental traits behind them. Then, they discovered the five global (or second-order) factors by factor-analysing the sixteen primary traits themselves, to find the basic, organizing forces among the sixteen basic traits. Thus, the 16 PF test gives scores on both the five second-order global traits which provide an overview of personality at a broader, conceptual level, as well as on the more-numerous and precise primary traits, which give a picture of the richness and complexity of each unique personality. A listing of these traits can be found in

the article on the 16 Personality Factor Model. Given below is a precise model of the sixteen factors and how they are used [3].

III. METHODOLOGY

Data for this research was collected from 'http://personality-testing.info/_rawdata/', which stores data on personality surveys for the purposes of academic research and personal use. The questionnaire is based on the famous model of personality created by Raymond Cattell. This particular questionnaire used 163 questions divided in 16 categories. The answers to the questions are in a gradient form, ranging from 'Strongly Agree', 'Agree', 'Neither Agree nor Disagree', 'Disagree', 'Strongly Disagree'. The data from web was obtained in the 'csv' format. The file had data stored in numerical form where Strongly Agree is equal to 5 and Strongly Disagree is equal to 1

Preliminary Analysis

The questions used for the survey that generated this dataset are given in table 2 additional sheet. The questions were provided with the codebook available with the dataset. The questions highlighted in pink are negative intent questions, where their scores mean opposite of the category they represent. The questions were classified into 16 categories as seen above; these categories are derived from the system developed by Raymond Cattell. The data also had four more variables apart from questions for storing variables like age, gender, accuracy and time taken to fill the form. Of the total 1414 candidates 878 were female, 535 were male and one was recorded as 'other'. The ages of the candidates under study varied from 14 to 80. The mean age of all the candidates was found to be 25.4. Most of the candidates were in the ages between 16 and 26 -a total of 925 people out of 1414. The highest number i.e. 209 people were aged 17.

For the purpose of this research the numerical ranges were converted into binary data of True and False denoted by 1 and 0 respectively. Answers ranging from 1 to 3 were converted to 0 and answers 4 and 5 were turned into 1. This conversion reduced the gradient into binary information which could be dealt with easily. It was also assumed at this point that people who possess the trait in question are normally confident about that and answer the question with high ranks, i.e. 'Agree' or 'Strongly Agree'. The moderate answer therefore was converted into a 0 i.e. False.

Another important thing to be considered, while analysing the data set was that not all the questions had the same tone in asking i.e. in each of the sixteen categories there were a few, mostly at the end, whose intent was negative. So for each of these questions the polarity was reversed, i.e. the answers 'Strongly Disagree' and 'Disagree' were converted to '1' and the rest to zero. The size of the data set at this point was 1414 × 163, 1414 rows and 163 columns. Each column denotes the variable that we were dealing with. A sum of all the variables was generated. These values would be used in the future to calculate the confidence of associations. Given below is a graph showing the number of times 'True' occurs for each of the 163 questions (figure 1). The categories are indicated by their respective colours.

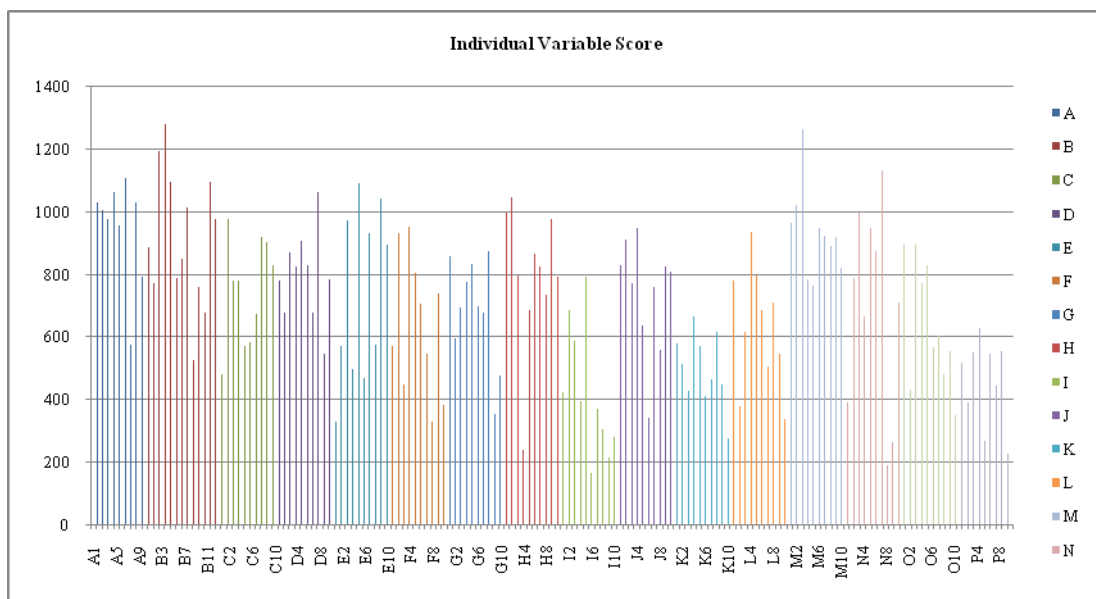


Figure 1: Score for 163 Questions

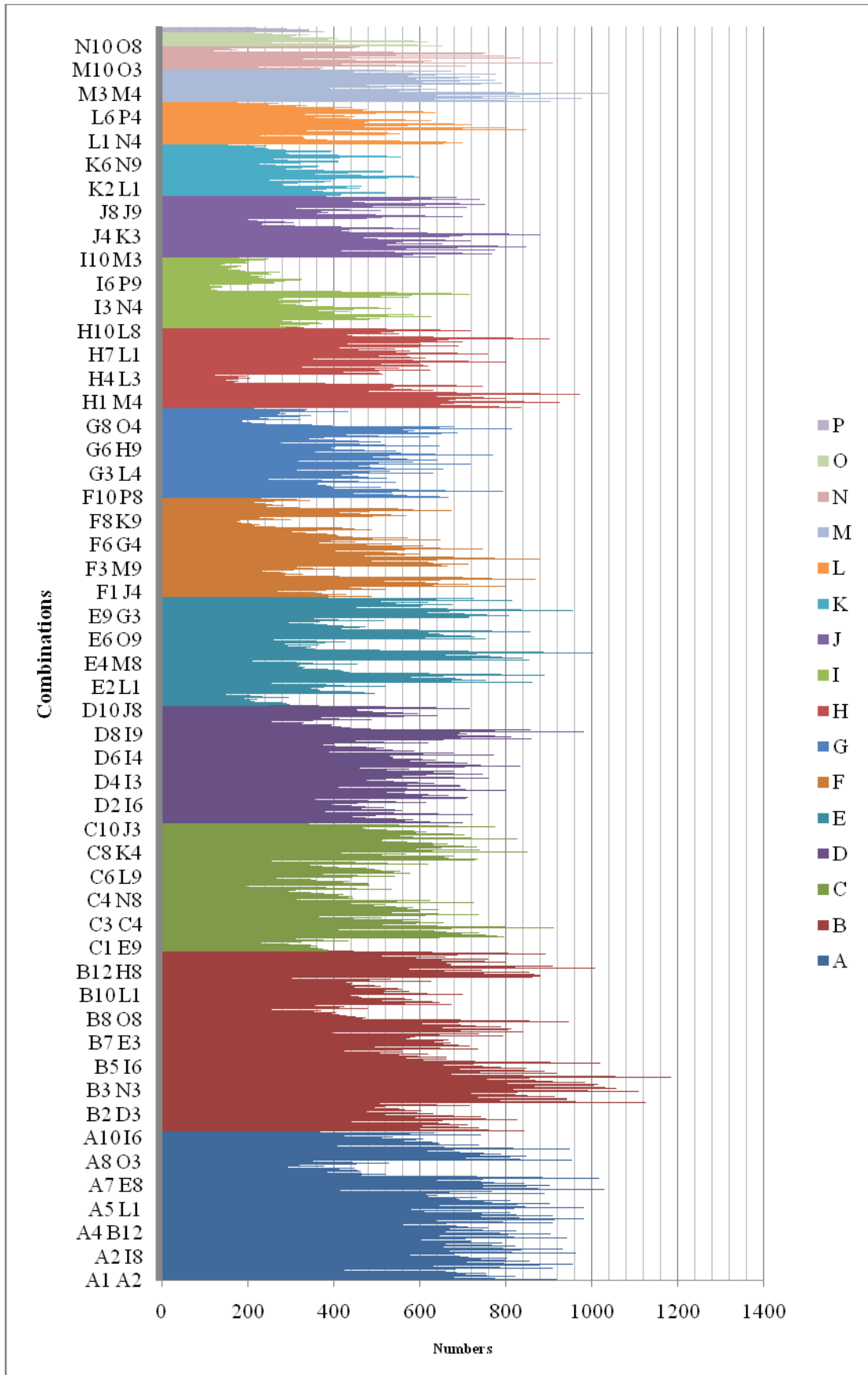


Figure 3: Associations (13,203) of 163 Variables

For the sake of ease and simplicity in using the above matrix in the algorithms that follow, its dual was created, i.e. the lower triangle of the matrix was duplicated into the upper triangle to obtain a symmetric matrix. The code for creating a symmetric matrix is given in the appendix. A section of the duplicate matrix is shown in figure 4.

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	E1	E2	E3	E4
A1	NA	825	806	771	862	784	852	453	805	534	639	536	856	318	736	550	619	727	363	556	481	774	677	351	737	532	601	425	424	507	678	672	603	532	490	645	805	673	629	515	762	401	563	277	458	727	404
A2	825	NA	823	808	878	732	894	470	833	635	676	556	891	357	824	579	635	770	362	569	509	736	700	362	758	613	619	453	447	538	710	630	643	616	526	669	640	636	652	531	786	413	575	232	501	779	442
A3	806	823	NA	777	836	767	858	475	806	629	663	555	892	332	802	564	639	746	379	544	505	735	632	326	716	562	573	410	388	468	655	636	531	581	494	644	508	670	619	501	767	393	527	263	433	732	391
A4	771	808	777	NA	806	765	844	480	832	533	659	548	851	305	782	563	610	745	357	530	432	786	674	346	703	563	585	400	437	508	659	653	588	563	480	631	588	648	538	468	748	363	540	244	435	718	395
A5	862	878	836	806	NA	861	913	467	850	633	694	598	910	363	858	583	642	773	381	585	539	631	728	390	794	646	641	464	461	551	743	720	663	633	531	632	645	718	660	544	616	424	606	304	496	617	434
A6	784	732	767	735	861	NA	828	432	772	595	637	541	836	301	788	540	595	716	377	537	432	786	675	377	731	644	640	463	484	553	746	736	688	637	541	634	652	736	684	550	648	430	609	287	469	605	
A7	852	854	859	844	913	828	NA	430	886	610	720	604	949	1029	877	615	679	827	330	606	530	871	767	377	731	644	640	463	484	553	746	736	688	637	541	634	652	736	684	550	648	430	609	287	469		
A8	459	470	475	480	467	432	430	NA	493	378	384	337	437	522	465	335	362	426	237	333	301	466	405	169	411	332	319	236	229	286	381	372	352	339	309	386	345	371	361	306	462	249	314	182	258	416	
A9	805	833	808	832	850	772	886	493	NA	645	617	557	884	354	821	569	629	754	378	581	517	835	740	360	739	605	596	430	470	543	707	700	651	587	487	646	611	668	624	532	810	412	582	265	465		
A10	594	635	629	593	633	595	670	678	645	NA	528	439	690	738	627	450	504	581	318	466	420	645	581	278	570	466	338	345	352	404	540	542	504	459	375	478	475	505	497	418	643	322	451	186	310		
B1	639	676	663	659	684	637	720	384	671	528	NA	598	806	843	747	589	532	711	397	541	432	716	658	308	638	557	508	353	382	435	581	588	555	522	454	532	583	598	593	450	725	370	524	200	330		
B2	536	556	558	548	588	541	604	337	557	439	598	NA	703	742	634	530	500	614	377	510	423	679	589	273	556	489	440	328	339	379	515	505	489	482	435	518	518	545	522	412	625	341	469	196	404		
B3	856	891	832	851	910	836	943	497	884	690	806	703	NA	1128	982	713	785	916	481	677	603	962	862	415	839	694	654	492	494	585	786	761	713	704	610	760	740	809	745	525	943	478	675	266	470		
B4	918	957	932	905	983	901	1023	524	954	738	843	1126	NA	1056	745	816	961	499	736	632	1031	919	454	924	744	723	537	551	630	866	840	784	739	635	810	788	758	738	1005	515	739	300	526				
B5	793	842	802	782	858	788	877	465	821	627	747	694	982	1056	NA	675	710	837	458	678	571	916	809	357	811	677	635	460	487	549	741	727	696	663	599	712	686	715	810	464	647	268	476				
B6	550	579	584	583	593	540	615	335	589	500	589	713	745	675	NA	501	609	365	483	415	663	588	262	539	491	442	314	327	369	505	500	487	503	439	556	568	568	517	662	347	479	197	328				
B7	619	635	639	610	642	595	679	362	623	504	532	500	785	816	501	NA	716	354	513	445	635	642	394	508	486	375	374	403	572	548	529	494	402	521	520	578	512	408	656	330	499	177	309				
B8	727	770	746	745	773	716	827	426	754	581	714	916	916	961	637	609	716	NA	416	611	512	840	750	368	740	623	588	427	452	514	696	674	634	610	524	659	653	707	654	518	805	407	597				
B9	363	362	373	357	381	377	390	237	378	318	397	711	481	499	458	385	354	416	NA	381	380	474	420	155	282	284	209	237	260	367	352	368	339	314	355	350	376	364	326	452	262	347					
B10	556	569	544	530	585	589	606	331	581	466	541	510	677	736	678	683	513	611	353	NA	419	675	614	292	582	501	459	367	419	448	564	567	580	503	417	504	517	552	521	455	647	317					
B11	491	509	505	492	539	492	603	632	571	415	445	512	380	419	NA	419	419	419	419	419	419	675	614	292	582	501	459	367	419	448	564	567	580	503	417	504	517	552	521	455	647	317					
B12	774	796	795	786	831	788	871	466	835	645	761	679	962	1031	919	663	691	840	474	675	576	NA	861	390	775	654	610	445	495	554	723	747	685	638	536	702	688	632	633	564	882	454	685				
B13	677	700	692	678	728	685	767	465	821	627	747	694	982	1056	NA	675	710	837	458	678	571	916	809	357	811	677	635	460	487	549	741	727	696	663	599	712	686	715	810	464	647	268	476				
C1	351	362	326	346	390	367	377	189	360	278	206	273	415	454	393	262	297	368	195	292	257	390	342	NA	387	328	334	266	282	350	387	380	342	301	240	322	312	332	311	257	376	194	314				
C2	737	758	716	703	794	736	791	411	739	570	638	556	839	324	801	539	634	740	383	582	518	775	675	387	NA	637	646	466	480	567	797	719	687	605	495	666	624	689	627	533	770	406	628				
C3	532	613	582	583	646	617	644	332	605	466	557	489	694	744	677	491	508	629	352	501	433	654	576	328	637	NA	515	379	413	473	593	582	579	531	444	541	539	590	557	462	653	361					
C4	601	619	573	585	641	592	640	319	596	438	508	440	654	723	635	442	486	588	284	459	385	610	543	334	646	515	NA	416	432	478	599	584	543	457	379	528	508	521	509	402	602	312	486				
C5	425	453	410	400	464	424	483	236	430	345	323	492	537	460	314	375	427	209	367	292	445	407	266	466	379	416	NA	334	356	447	429	423	329	276	369	345	376	373	303	435	227	358					
C6	424	447	388	437	461	445	484	239	470	352	323	494	551	487	327	374	452	237	419	307	495	447	282	480	413	432	334	NA	432	482	485	447	354	284	386	389	397	371									

A1	NA	A2	A3	A4	A5	A6	A7	A8	A9	A10	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
A1	NA	0.841837	0.822449	0.786735	0.879592	0.8	0.869388	0.468367	0.821429	0.606122	0.652041	0.546939	0.873469	0.936735	0.812245	0.561224	0.631633	0.741837	0.370408	0.567347	0.490816	0.789796
A2	0.801749	NA	0.799806	0.785238	0.853256	0.769679	0.868805	0.456754	0.809524	0.611704	0.656948	0.54033	0.865889	0.930029	0.800777	0.562682	0.617104	0.748299	0.351798	0.552964	0.494655	0.773567
A3	0.801193	0.818091	NA	0.772366	0.831014	0.762425	0.852883	0.472167	0.801193	0.625249	0.659046	0.55169	0.88668	0.926441	0.797217	0.560636	0.635189	0.741551	0.37674	0.540755	0.501988	0.790258
A4	0.78915	0.827021	0.795292	NA	0.824974	0.752303	0.863869	0.4913	0.851586	0.60696	0.674514	0.560901	0.871034	0.926305	0.800409	0.576254	0.62436	0.762538	0.365404	0.542477	0.503582	0.804504
A5	0.810913	0.825964	0.786453	0.758231	NA	0.809972	0.85889	0.439323	0.799624	0.595484	0.643462	0.553151	0.856068	0.924741	0.80715	0.548448	0.603951	0.727187	0.35842	0.550329	0.507056	0.78175
A6	0.820084	0.828452	0.802301	0.768828	0.900628	NA	0.866109	0.451883	0.807531	0.622385	0.666318	0.5659	0.874477	0.942469	0.824268	0.564854	0.622385	0.748954	0.394351	0.595188	0.514644	0.803347
A7	0.769648	0.807588	0.775068	0.762421	0.824752	0.747967	NA	0.442638	0.800361	0.605239	0.650407	0.545619	0.857272	0.929539	0.792231	0.555556	0.613369	0.747064	0.352304	0.547425	0.478771	0.786811
A8	0.799652	0.818815	0.827526	0.836237	0.813589	0.752613	0.853659	NA	0.858885	0.658537	0.66899	0.587108	0.865854	0.909408	0.810105	0.583624	0.630662	0.74216	0.412892	0.580139	0.52439	0.811847
A9	0.782313	0.809524	0.783285	0.808552	0.826045	0.750243	0.86103	0.479106	NA	0.626822	0.652089	0.541302	0.859086	0.927114	0.797862	0.552964	0.611273	0.73275	0.367347	0.564626	0.50243	0.811467
A10	0.750948	0.802781	0.795196	0.749684	0.800253	0.752212	0.847029	0.477876	0.815424	NA	0.667509	0.554994	0.872314	0.932996	0.792668	0.5689	0.637168	0.734513	0.402023	0.589128	0.530973	0.815424
B1	0.720406	0.762212	0.747463	0.742954	0.771139	0.718151	0.811725	0.43292	0.756483	0.595265	NA	0.674183	0.908681	0.950395	0.842165	0.664036	0.667418	0.801578	0.447576	0.609921	0.554679	0.857948
B2	0.695201	0.721141	0.719844	0.710765	0.762646	0.701686	0.783398	0.437095	0.722438	0.56839	0.775616	NA	0.911803	0.962387	0.90013	0.687419	0.648508	0.796368	0.488975	0.661479	0.55642	0.880674
B3	0.717519	0.746857	0.747695	0.713328	0.762783	0.700754	0.795474	0.416597	0.740989	0.578374	0.675608	0.589271	NA	0.943839	0.823135	0.597653	0.658005	0.767812	0.403185	0.567477	0.505448	0.80637
B4	0.716069	0.74649	0.726989	0.705928	0.766771	0.702808	0.802262	0.407176	0.74415	0.575663	0.657566	0.578783	0.878315	NA	0.823713	0.581123	0.636505	0.74961	0.389236	0.574103	0.49298	0.804212
B5	0.725615	0.751139	0.731085	0.712853	0.782133	0.718323	0.799453	0.423883	0.748405	0.571559	0.680948	0.632634	0.895169	0.962625	NA	0.615314	0.64722	0.76299	0.417502	0.618049	0.52051	0.836828
B6	0.698856	0.735705	0.716645	0.715375	0.740788	0.68615	0.781449	0.425667	0.722999	0.571792	0.748412	0.673443	0.905972	0.946633	0.857687	NA	0.636595	0.773825	0.463787	0.613723	0.527319	0.84244
B7	0.728235	0.747059	0.751765	0.717647	0.755294	0.7	0.798824	0.425882	0.74	0.592941	0.696471	0.588235	0.925229	0.96	0.835294	0.589412	NA	0.842335	0.416471	0.603529	0.523529	0.812941
B8	0.715551	0.757874	0.734252	0.733268	0.760827	0.704724	0.813976	0.419291	0.742126	0.57185	0.699803	0.604331	0.912173	0.945866	0.823819	0.599409	0.704724	NA	0.409449	0.601378	0.503937	0.826772
B9	0.688805	0.686907	0.719165	0.677419	0.72296	0.71537	0.740038	0.449715	0.717268	0.603416	0.753321	0.71537	0.912713	0.946869	0.86907	0.6926	0.671727	0.789374	NA	0.669829	0.721063	0.899431
B10	0.732543	0.749671	0.716733	0.698287	0.770751	0.749671	0.798419	0.438735	0.765481	0.613966	0.71278	0.671937	0.891963	0.969697	0.893281	0.636364	0.675889	0.805007	0.465086	NA	0.552042	0.889328
B11	0.70944	0.750737	0.744838	0.725664	0.794985	0.725664	0.817111	0.443953	0.762537	0.619469	0.725664	0.632743	0.889381	0.932153	0.842183	0.612094	0.656342	0.755162	0.560472	0.617994	NA	0.849558
B12	0.705561	0.725615	0.724704	0.7165	0.757521	0.700091	0.793984	0.424795	0.761167	0.587967	0.69371	0.618961	0.876937	0.939836	0.836828	0.604376	0.6299	0.765725	0.432088	0.615314	0.525068	NA
B13	0.691522	0.715015	0.706844	0.688458	0.743616	0.679265	0.783453	0.413687	0.755873	0.593463	0.672114	0.601634	0.88049	0.938713	0.826353	0.600613	0.637385	0.766088	0.429009	0.627171	0.504597	0.879469
C1	0.73125	0.754167	0.679167	0.720833	0.8125	0.764583	0.785417	0.39375	0.75	0.579167	0.641667	0.56875	0.864583	0.945833	0.81875	0.545833	0.61875	0.766667	0.40625	0.608333	0.535417	0.8125
C2	0.753579	0.775051	0.732106	0.718814	0.811861	0.752556	0.808793	0.420245	0.755624	0.582822	0.652352	0.568507	0.857873	0.944785	0.819018	0.551125	0.648262	0.756646	0.391616	0.595092	0.529652	0.792344
C3	0.758974	0.785897	0.746154	0.747436	0.828205	0.791026	0.825641	0.425641	0.775641	0.597436	0.714103	0.626923	0.889744	0.953846	0.867949	0.629487	0.651282	0.80641	0.451282	0.642308	0.555128	0.838462
C4	0.770513	0.79359	0.734615	0.75	0.821795	0.758974	0.820513	0.408974	0.764103	0.561538	0.651282	0.564103	0.838462	0.926923	0.814103	0.566667	0.623077	0.753846	0.364103	0.588462	0.49359	0.82051
C5	0.746924	0.796134	0.720562	0.702988	0.815466	0.745167	0.813708	0.414763	0.755712	0.606327	0.620387	0.578207	0.864675	0.943761	0.808436	0.551845	0.659051	0.750439	0.367311	0.644991	0.513181	0.782074
C6	0.729776	0.769363	0.667814	0.752151	0.79346	0.765921	0.833046	0.41136	0.80895	0.605852	0.674699	0.583477	0.850258	0.948365	0.83821	0.562823	0.643718	0.777969	0.407917	0.72117	0.528399	0.851979
C7	0.753343	0.799406	0.695394	0.754829	0.818722	0.787519	0.821694	0.424963	0.806835	0.600297	0.66136	0.56315	0.839525	0.936107	0.81575	0.548291	0.598811	0.763744	0.38633	0.665076	0.542348	0.82318
C8	0.736957	0.771739	0.711957	0.716304	0.807609	0.729348	0.81087	0.41413	0.768478	0.586957	0.631522	0.559783	0.845438	0.941304	0.805435	0.548913	0.621739	0.756522	0.398913	0.613043	0.56413	0.792391
C9	0.745011	0.764967	0.7051	0.730599	0.798226	0.733925	0.815965	0.412417	0.776053	0.600887	0.651885	0.559867	0.843681	0.931264	0.805987	0.554324	0.607539	0.747228	0.390244	0.628603	0.545455	0.82816
C10	0.727382	0.75633	0.712907	0.709288	0.799759	0.747889	0.80579	0.424608	0.785283	0.607961	0.669481	0.589867	0.860072	0.945718	0.839566	0.587455	0.638118	0.764777	0.443908	0.675513	0.569361	0.826297
C11	0.759949	0.790757	0.745828	0.722721	0.81258	0.759949	0.817715	0.435173	0.75353	0.589217	0.67009	0.618742	0.903723	0.948652	0.851091	0.6457	0.634146	0.783055	0.435173	0.6457	0.557125	0.81899
D2	0.724852	0.778107	0.730769	0.710059	0.785503	0.726331	0.800296	0.448225	0.720414	0.554734	0.671598	0.643491	0.902367	0.939349	0.841716	0.649408	0.594675	0.775148	0.464497	0.616864	0.573964	0.792899
D3	0.741379	0.768966	0.74023	0.725287	0.795402	0.714943	0.797701	0.443678	0.742529	0.549425	0.68046	0.595402	0.873563	0.931034	0.818391	0.63908	0.598851	0.757471	0.408046	0.57931	0.511494	0.806897
D4	0.73156	0.773881	0.711004	0.711004	0.779927	0.714631	0.788392	0.41717	0.738815	0.574365	0.704958	0.62636	0.8948	0.943168	0.829504	0.68682	0.628779	0.789601	0.423216	0.625151	0.52237	0.831923
D5	0.742826	0.770419	0.739514	0.715232	0.792494	0.731788	0.812362	0.409492	0.738411	0.557395	0.660044	0.601545	0.892936	0.94702	0.835541	0.626932	0.637969	0.780353	0.417219	0.609272	0.534216	0.807947
D6	0.757831	0.785542	0.745783	0.720482	0.795181	0.740964	0.824096	0.43494	0.751807	0.598795	0.714458	0.628916	0.893976	0.943373	0.861446	0.642169	0.616867	0.787952	0.436145	0.627711	0.53494	0.83494
D7	0.759587	0.783186	0.738938	0.719764	0.80236	0.731563	0.811209	0.451327	0.784661	0.616519	0.663717	0.60767	0.877581	0.941003	0.846608	0.615044	0.60177	0.764012	0.480826	0.671091	0.60767	0.831858
D8	0.715493	0.738028	0.720188	0.702347	0.766197	0.703286	0.796244	0.433803	0.760563	0.603756	0.680751	0.596854	0.885446	0.943662	0.826291	0.621596	0.615962	0.755869	0.424413	0.607512	0.525822	0.828169

Figure 5: Confidence Value Matrix

2) R [16] Snippet for 2D Matrix Calculation

```

binary=read.csv("binary.csv") #Reads the file that contains all the scores in binary
comb=combn(163,2) #Creates a matrix (163 x 2) of combinations
resultmatrix=diag(NA,nrow=163) #Creates a null matrix (163 x 163)
for(k in 1:ncol(comb)) #For loop runs through combination matrix
{
    result = binary[,comb[1,k]+1] & binary[,comb[2,k
```


The confidence values for associations with A1, C1, C9, P8 and O6 selectively are shown in the figure 6 to figure 10

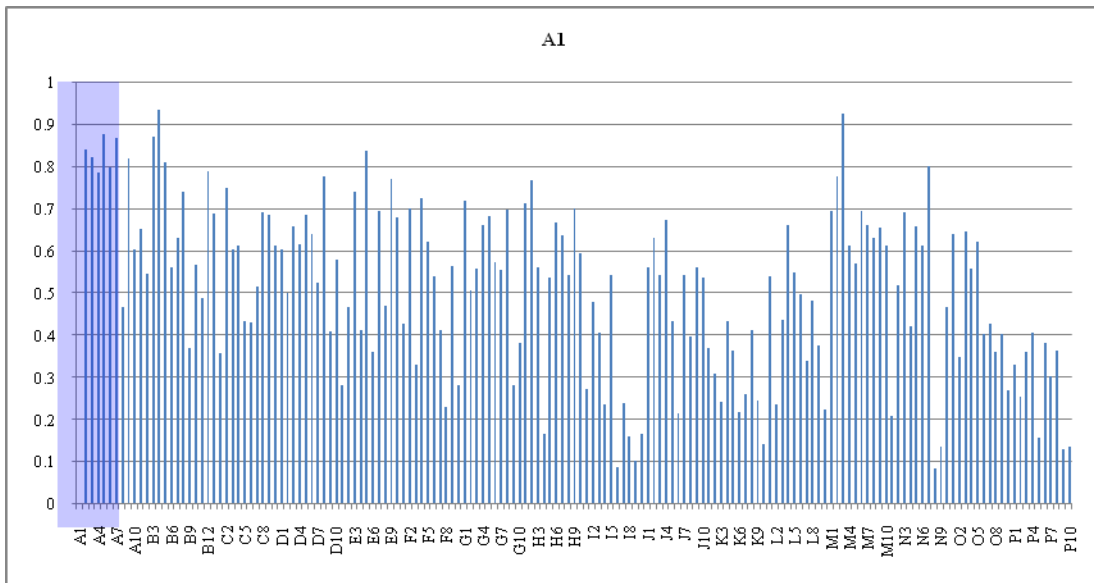


Figure 6: Confidence Values for Association with A1

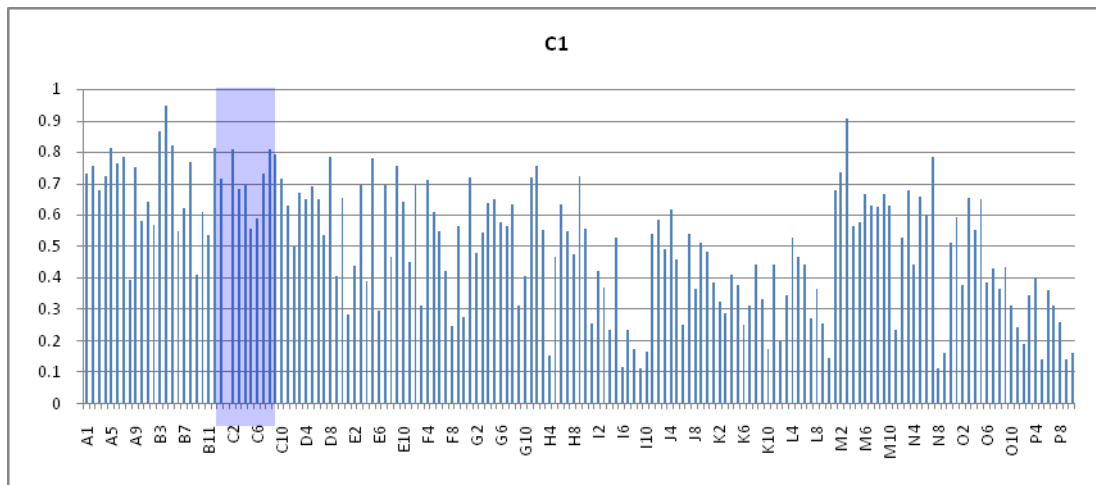


Figure 7: Confidence Values for Association with C1

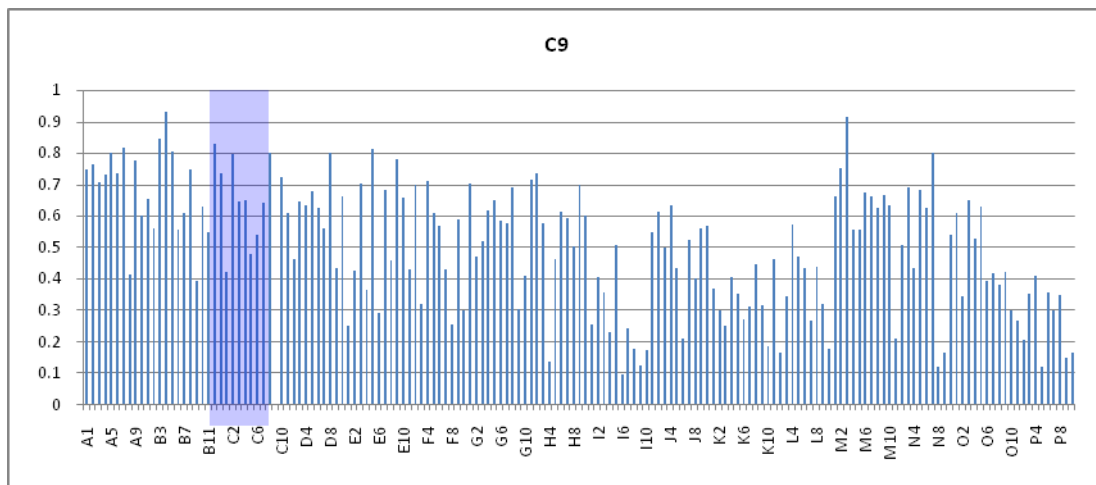


Figure 8: Confidence Values for Association with C9

So remaining figures from 12 to 17 are given as supplementary figures.

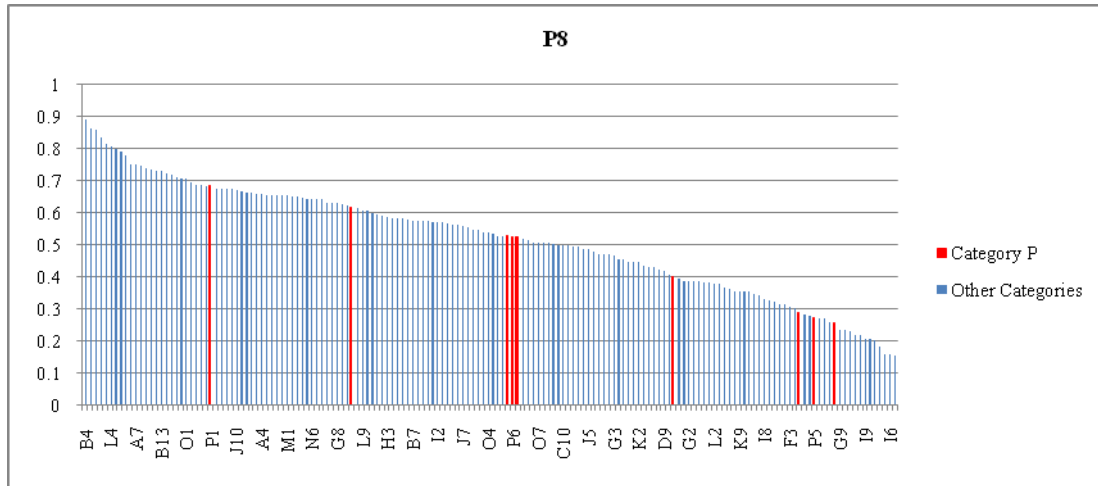


Figure 9: Confidence Values for Association with P8 Sorted

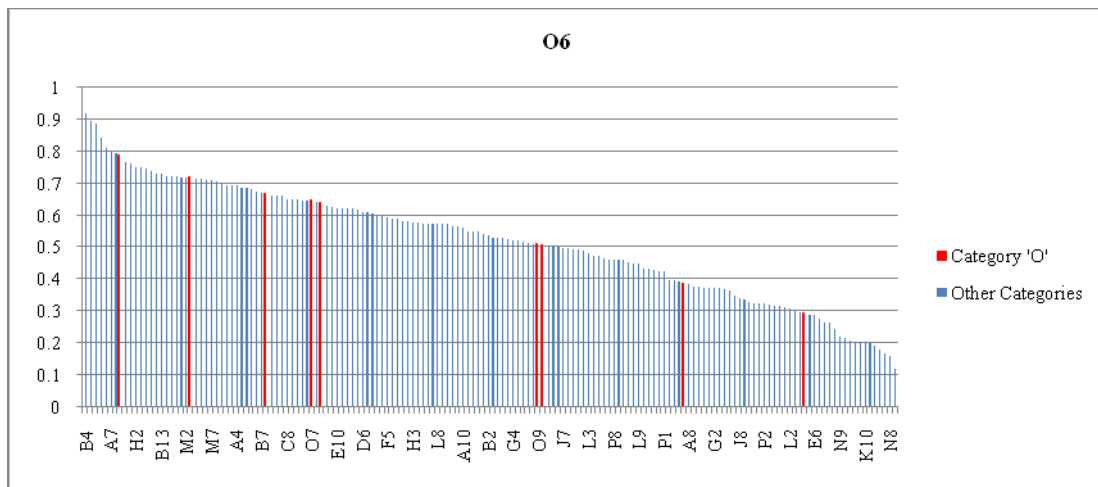


Figure 10: Confidence Values for Association with O6 Sorted

Similarly, remaining figures from 20 to 22 are given as supplementary figures.

V. CONCLUSION AND FUTURE WORK

We proposed a new technique to predict human behaviour, which is the use of simple statistics. As we demonstrate through our paper, words or phrases can be studied using a survey to discover their correlations in populations. In this paper we demonstrate the correlations between phrases of the Raymond Cattell’s 16 personality factors.

This process led to the discovery of quite a few unexpected associations. The list of 26,569 is reduced to 26,406 by subtracting 163 associations of the variables by themselves (i.e. A1 with A1, etc.). Out of these 26,406 associations there were many which were obvious, for e.g., H3 with H1, ‘Read a lot’ and ‘like to read’, B4 given B2, ‘Use my brain’ given ‘Learn quickly’ and many more of the like. Although such relations do form a part of this analysis and help us proving the effectiveness and applicability of this method, since all such relations which are obvious and expected formed the major bulk of the top five hundred strongest correlations. Of all the 26,406 associations that we obtained, 299 of them exceeded 90% in their confidence values. Also there were 1326 association rules that had confidence values of 80% and above. That makes up 5% of all the associations. Yet in what follows, we focus on some unexpected and interesting results that we obtained (Figure 23). The percentage values given in the arrows indicate the confidence values and hence are directional.

Also as one would expect ‘get angry easily’ would have strong association with ‘get irritated easily’, it seems, that is not the case. There is a stronger relationship with a variable of a different category. The Figure 11 is such type of strong relations we were able to form.

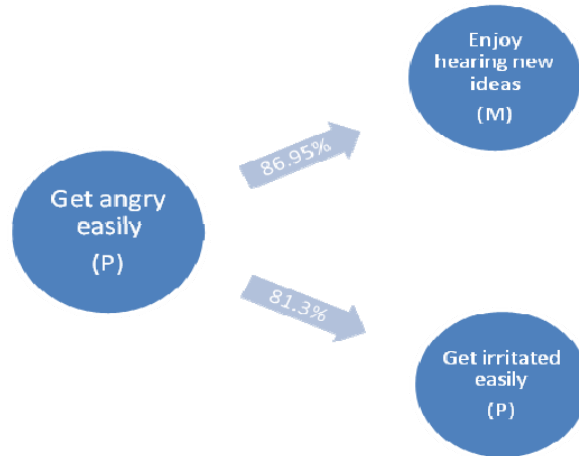


Figure 11: Strong relations discovered

Interesting or not, the usefulness of this approach has been proved quite extensively. Yet this approach in this context suffers certain drawbacks which are listed below.

- i. Associations have been established using a method that was proven less reliable in this paper itself.
- ii. The phrases of behaviour and thought process have been taken from a well-developed system whose foundations we claim to be flawed. And therefore all techniques employing this system are bound to be misleading. Also, like every other psychometric test or questionnaire that involves objective testing of personality, the questionnaire.

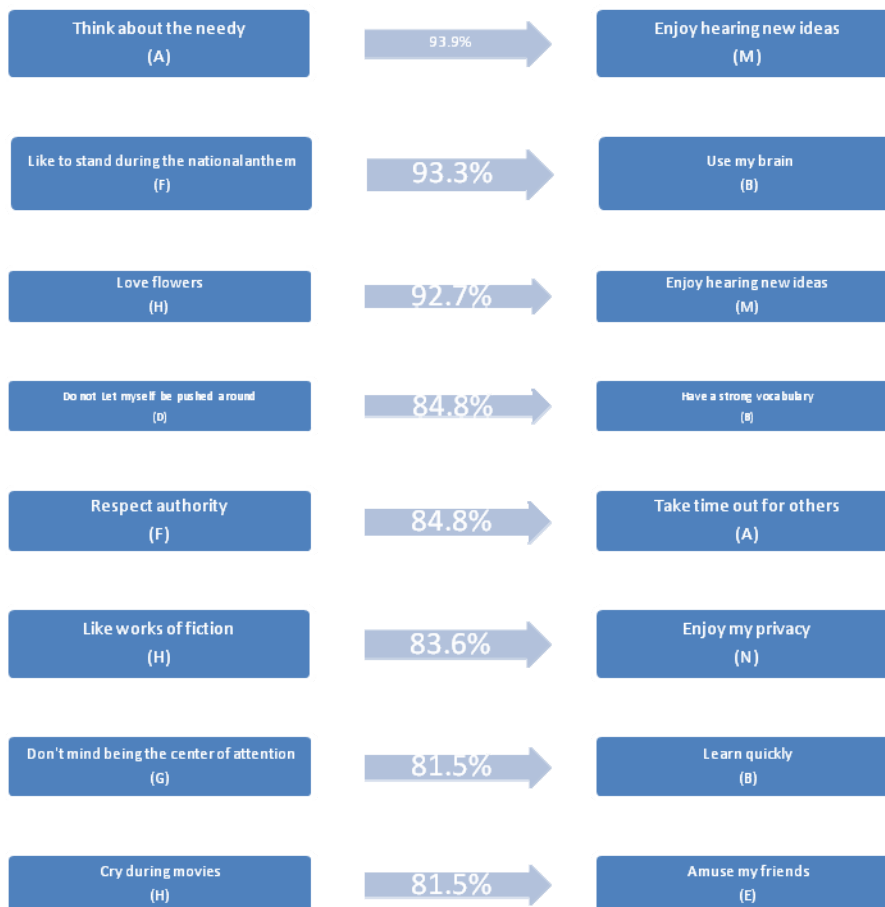


Figure 12: Important Associations discovered

used in this paper is also bound to have misrepresentation of self-image, i.e. wrong self-perception. This can be proved by looking at the correlations, where an astounding 61 of the top 100 association contain some instance of the question index 'B4' that represent 'using of brain'. Smart or not every person assumes continuous usage of brain for active thinking. Similar were the instances involving 'Enjoy hearing new ideas' as well as 'Tend to analyse things', all associated with 'reasoning' or 'openness to change'. So to be able to apply this technique in the real world the first step would be to do away with the drawbacks, by using data obtained through surveys that are not based on any system rather are simple questionnaires concerning individual preferences. Yet a better approach would be to collect behavioural information of an individual from his/her friends, family and the network of colleagues. This, in today's world of connectedness through social networking sites, would be a trivial task. The data would contain simple words or phrases that demonstrate the behaviour of an individual. For e.g., 'gets angry fast', 'enthusiastic', 'approachable', 'friendly', etc. Once such a data is obtained simple association rule mining would produce associations that are not only useful but reliable as well. A system of behaviour prediction developed on the foundations of simple statistics has great potential in times to come. Once such a database of thousands of words and phrases is developed, which we call, 'reflection database', it can be used to fine grain over certain important characteristics. For e.g., different databases can be developed for different age groups, and likewise for different regions or countries, cultures, religion, ethnicities, financial background, gender and all such factors which have been proven to contribute to personality.

Also, such databases could be developed for different purposes and the scope of this method could be diversified. Consider for e.g. how behaviour is related to facial features, or how behaviour is associated with where a person is brought up, i.e. rural vs. urban. Moreover certain psychological puzzles concerning career choices can be resolved like 'are introverts better at science or vice versa?' Also how behaviour in childhood is associated with behaviour in adulthood as well as how behaviour can be used to enhance recruitment process in the corporate sector.

"What an increase in mental power mankind would achieve if the basic mental conditions which determine the different aptitudes for being a poet, a scientist, or a man of practical ability could be fully ascertained beyond any doubt by means of psychological analysis! If this were possible, we could recognize the tree, not from its fruit, but from its very first budding leaves, and could transplant it immediately to a place suited to its nature." – Franz Brentano.

ACKNOWLEDGMENT

We express our profound gratitude to Prof. Ramanathan K., Prof. Anand A. Programme Manager (B.Tech - Bioinformatics), Division Leader (Bioinformatics) SBST, VIT University, for providing resources to complete this research work. We would like to extend our thanks to our Chancellor Dr. G. Viswanathan too.

REFERENCES

- [1] Cattell, R.B., Cattell, A.K., & Cattell, H.E.P. (1993). *16PF Fifth Edition Questionnaire*. Champaign, IL: Institute for Personality and Ability Testing.
- [2] Cattell, R.B. (1957). *Personality and motivation structure and measurement*. New York: World Book.
- [3] Conn, S.R., & Rieke, M.L. (1994). *The 16PF Fifth Edition technical manual*. Champaign, IL: Institute for Personality and Ability Testing, Inc.
- [4] Phares, E.J.; Chaplin, W.F. (1997). *Introduction to personality* (Fourth ed.). New York: Longman. pp. 8–9. ISBN 0-673-99456-2.
- [5] Feist, Jess Feist, Gregory J. (2009). *Theories of personality* (7th ed. ed.). Boston: McGraw Hill Higher Education. ISBN 978-0073382708.
- [6] Aluja, Anton; García, Óscar; García, Luís F. (2004). "Replicability of the three, four and five Zuckerman's personality super-factors: Exploratory and confirmatory factor analysis of the EPQ-RS, ZKPQ and NEO-PI-R". *Personality and Individual Differences* **36** (5):1093. Doi:10.1016/S0191-8869(03)00203-4.
- [7] McAdams, Dan P. (2009). *The person: a new introduction to personality psychology* (5th ed.). Hoboken, N.J.: Wiley. p. 113. ISBN 978-0-470-12913-5.
- [8] Sharp, Daryl (1987). *Personality types: Jung's model of typology*. Toronto, Canada: Inner City Books. p. 128. ISBN 0919123309.
- [9] Kahn, Michael (2002). *Basic Freud: psychoanalytic thought for the twenty first century* (1. paperback ed. ed.). New York: Basic Books. ISBN 9780465037162.
- [10] Cheney, W. David Pierce, Carl D. (2008). *Behavior analysis and learning* (4th ed. ed.). New York, NY: Psychology Press. ISBN 9780805862607.
- [11] Combs, Arthur W., and Snygg, Donald. : *A New Frame of Reference for Psychology*. New York, Harper and Brothers Article on Snygg and Combs' Phenomenological Field Theory.
- [12] Plomin, R., DeFries, J.C., McClearn, G.E., & Rutter, M. (1997). *Behavioral genetics* (3rd Ed.). New York: Freeman.
- [13] Piatetsky-Shapiro, Gregory (1991), *Discovery, analysis, and presentation of strong rules*, in Piatetsky-Shapiro, Gregory; and Frawley, William J.; eds., *Knowledge Discovery in Databases*, AAAI/MIT Press, Cambridge, MA.
- [14] Agrawal, R.; Imieliński, T.; Swami, A. (1993). "Mining association rules between sets of items in large databases". *Proceedings of the 1993 ACM SIGMOD international conference on Management of data -SIGMOD '93*. p. 207. Doi:10.1145/170035.170072. ISBN 0897915925.
- [15] Hipp, J.; Güntzer, U.; Nakhaeizadeh, G. (2000). "Algorithms for association rule mining --- a general survey and comparison". *ACM SIGKDD Explorations Newsletter* **2**: 58. Doi: 10.1145/360402.360421.
- [16] R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-research.org/>.

- [17] Allport, G. W. (1937). *Personality: A psychological interpretation*. New York, NY: Henry Holt & Company.
- [18] Brentano, F. (1995). *Psychology from an empirical standpoint*. (Trans. AC Rancurello, DB Terrell, & LL McAlister). (Original work published 1874).
- [19] Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior?. *Perspectives on Psychological Science*, 2(4), 396-403.
- [20] Ajzen, I. (1987). Attitudes, traits, and actions: Dispositional prediction of behavior in personality and social psychology. *Advances in experimental social psychology*, 20(1), 63.
- [21] DeFleur, M. L., & Westie, F. R. (1958). Verbal attitudes and overt acts: An experiment on the salience of attitudes. *American Sociological Review*, 23(6), 667-673.
- [22] McGuire, W. J. (1969). The nature of attitudes and attitude change. *The handbook of social psychology*, 3, 136-314.
- [23] Bijou, S. W., Peterson, R. F., & Ault, M. H. (1968). A method to integrate descriptive and experimental field studies at the level of data and empirical concepts. *Journal of applied behavior analysis*, 1(2), 175.
- [24] Vernon, P. E., & Vernon, P. E. (1964). *Personality assessment: A critical survey*. London: Methuen.
- [25] Sherif, C. W. (1998). Bias in psychology. *Feminism & Psychology*, 8(1), 58-75.

Author's Profile:

Dr.Syed Khalid Perwez

He did his UG in Psychology, PG in Industrial and Clinical Psychology from Ranchi University, and MBA (HR and Marketing) from Utkal University, Bhubaneswar, and PhD from Ranchi University in the field of Psychology. He has 8 years of industry experience and currently working as Associate Professor in VITBS, VIT University, Vellore since 3 years. He has in his name several peer reviewed publications in national and international conferences and journals. His research interest is stress, personality and management.

Hamza Mohd. Zubair

He has B.Tech in Bioinformatics from VIT University and presently pursuing his Master in Cognitive Science from IIT, Gandhinagar. His research interest falls in the line of data mining techniques in Psychological testing and Behaviour Predictions. He is attributed with publications in national and international conferences.

Muhammad Rukunuddin Ghalib

He has B.Tech(IT) and M.E.(CC) from Anna University, Chennai and right now on the verge of completing his PhD in CSE from Anna University, Chennai. Presently he is an Assistant Professor (Senior) in SCSE, VIT University, Vellore from last 6 years. His research interest includes Data Mining in Bioinformatics specially in Microarray Data Analysis , Neural networks and fuzzy logics algorithm design, Protein Structure prediction and Medical Image Processing. He has several peer-reviewed national and international conferences and journals publications.

Kauser Ahmed P

He has Bachelor and Master degree in Computer Science, additional M.Tech(IT) from VIT University, Vellore and right now pursuing his PhD in CSE from VIT University, Vellore. Presently he is an Assistant Professor in SCSE, VIT University, Vellore from last 6 years. His research interest includes Bioinformatics Algorithms, Neural networks and fuzzy systems and Information Security. He has several peer-reviewed national and international conferences and journals publications.

Mohammed Iftekhar

He has Bachelor degree in Biotechnology from Bangalore University, Master in Bioinformatics from Bhartiya University, Coimbatore. Currently pursuing his PhD in Bioinformatics from Tiruvalluvar University, Vellore. He has been a visiting professor in Dayanand Sagar Institute, Bangalore. He is also the Chairperson-Board of Directors of a consultancy firm, Biocuration Pvt. Ltd. in Bangalore.