Applying association rules to study Bipolar Disorder and Premenstrual Dysphoric Disorder comorbidity

Giovanna Castro*, Rogerio Salvini*, Fabrizzio A.A.M.N. Soares*, Andrew A. Nierenberg‡,
Gary S. Sachs‡, Beny Lafer§ and Rodrigo S. Dias§
*Instituto de Informatica, Universidade Federal de Goias, Goiania, GO, Brazil
email: {giovannacastro, rogeriosalvini, fabrizzio}@inf.ufg.br
‡Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
§Bipolar Disorder Research Program, Department of Psychiatry, University of Sao Paulo Medical School, Sao Paulo, SP, Brazil

***Abstract*—Bipolar Disorder (BD) is characterized by mood changes that manifest as depressive episodes alternating with episodes of euphoria, in varying degrees of intensity. Women with BD may experience worsening symptoms during events of their reproductive life, particularly those suffering from Premenstrual Dysphoric Disorder (PMDD). The presence of PMDD in the diagnoses of BD is considered a marker of severity for the disease. In this study, data from a cohort of 1099 women with BD were used for an exploratory analysis using association rules in order to find associations between PMDD and BD symptoms. Of the thousands of generated rules, those that have associations with PMDD were selected and categorized, with confidence levels between 70% and 100%.**

***Keywords*—Premenstrual Dysphoric Disorder, Bipolar Disorder, Association Rules, Apriori, Machine Learning**

I. Introduction

Bipolar Disorder (BD) is a mood disorder in which patients manifest episodes of depression and mania that can have negatives impacts on their lives. It has a chronic character that is capable of causing damage in personal and professional life, and leads to a poor quality of life and even to suicide [1], [2]. The BD occurs differently according the gender. In women, BD is characterized by an increased risk of mood instability during female-reproductive life such as the premenstrual, perinatal and perimenopausal periods. The onset of the disorder in women can be identified between 15 and 23 years old, with typical occurrences of depression preceded by episodes of mania [3]–[5].

Few studies have sought to understand the relation between BD and premenstrual dysphoric disorder (PMDD). PMDD affects from 3 to 9% of women [6], [7]. It was recently recognized in DSM-5 (Diagnostic and Statistical Manual of Mental Disorders - Fifth Edition) as an independent psychiatric disorder [8], while in ICD-10 (International Classification of Diseases and Related Health Problems - Tenth Revision) it is diagnosed as "Other mood disorders" [9].

Epidemiological studies performed by Slyepchenko [6] and Wittchen [7] showed that women with PMDD diagnosis are 8 times more likely to receive a diagnosis of BD, compared with women that did not have PMDD. These same women may experience more symptoms of BD during the perinatal period and while use contraceptive pills [6].

In the present study, we used a sample of 1,099 women with BD who provided detailed information about female-reproductive life as part of STEP-BD (Systematic Treatment Enhancement Program for Bipolar Disorder) [10]. The main objective is to perform an exploratory analysis of these data to find associations between the characteristics of the disease, the prevalence of comorbidity disorders and the symptoms of female-reproductive life in women diagnosed with PMDD.

To achieve this goal, we chose the association rules approach, an unsupervised machine learning technique for data mining. To the best of our knowledge, this type of approach has not been reported in the literature for this problem. In addition, the association rules have the advantage of being easily interpreted by physicians and professional experts. The easily analysis of the rules is given by the fact that rules are conditional events, that is, if the event A occurs then B also occurs.

The rest of this paper is organized as follows: Section II introduces some basic concepts of the association rules; Section III presents the methodology used in this work; Section IV presents and discusses the obtained results, and Section V concludes this study.

II. Association Rules

The discovery of association rules is one of the major techniques of data mining, and it is perhaps the most common form of local pattern discovery in unsupervised learning systems. It is a form of data mining that most closely resembles the process that most people think about when they try to understand the data - mining process, namely, "mining" for gold through a vast database. The gold in this case would be a rule that is interesting, that tells you something about your database that you did not already know and probably were not able to explicitly articulate [11].

More formally, the problem of discovering association rules proposed by *Agrawal et al* [12], [13] can be defined as: Let $I = \{a_1, ..., a_m\}$ the set of m items. Let $T = \{t_1, t_2, ..., t_n\}$ a set of transactions, where each transaction t_i is a set of items in which $t_i \subseteq I$. The set T is called *database*.

Association rules are implications rules that takes the form:

$$A \Rightarrow B,$$

which is interpreted as "if A then B", where:

- A is called the *antecedent* of the rule.
- B is called the *consequent* of the rule.

- $A \subset I$, $B \subset I$ and $A \cap B = \emptyset$.

The rule $A \Rightarrow B$ is valid in the database T with confidence c, if $c\%$ of transactions in T that contain A also contain B. The rule $A \Rightarrow B$ has *support* s in the database T, if $s\%$ of transactions in T contain $A \cup B$ [13]. *Support* is the frequency of the transaction patterns that occur in the database, and *confidence* is the probability of B be true when we already know that A is true.

The discovery of association rules can be divided two steps [12]:

1) Find *frequent itemsets*, which are sets of items where the *support* is greater or equal than a minimum established.
2) Generate the association rules based in the frequent itemset, selecting the rules with a minimum *confidence* degree.

So, the problem of data mining in relation to association rules is to produce rules that have minimum support and confidence.

The support of the rule $A \Rightarrow B$ is given by Equation 1:

$$Support(A \Rightarrow B) = \frac{\textit{Occurrences of A and B}}{\textit{Total transactions in T}} \quad (1)$$

and the confidence is given by Equation 2:

$$\textit{Confidence}(A \Rightarrow B) = \frac{\textit{Occurrences of A and B}}{\textit{Occurrences of A}} \quad (2)$$

A. Apriori Algorithm

The Apriori, proposed by *Agrawal et al.* [13], is an algorithm for frequent data mining and learning of association rules in the database.

Initially the algorithm searches the frequent itemsets, following these steps:

1) Calculates the support of each item in the database using Equation 1 and generate *1-itemset*. From these, it is selected those elements with support greater than the minimum and after this creates the set *1-FrequentItemset*.
2) The elements of the 1-FrequentItemset are combined to create the *2-itemset*. Calculates the support of each element of *2-itemset*. From these, there are selected the elements of *2-itemset* that have support greater than the minimum established and after this creates the set *2-FrequentItemset*.
3) Proceeds with the previous steps while the set of *k-items* found is not empty.

After the discovery of the frequent itemsets, the algorithm combines the elements of each frequent itemset and then generates the association rules. The rules that have a degree of confidence greater than or equal to the minimum established are selected. These rules are called *strong rules*.

III. Methods

We used the data from the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). The STEP-BD is the largest study of BD financed by Federal Government of the United States of America, characterized as long term outpatient study in which were enrolled 4.360 participants from 22 different regions of country over seven years (1998-2005). It was designated to find out which treatments, or combination of treatments, are most effectives for treating episodes of depression and mania, and for preventing recurrent episodes in people with bipolar disorder. It is an extensive research program that includes several different studies [10].

In this study we used the same data of the work done by Slyepchenko [6], which consists of a cohort of 1,099 women with BD, with 94 variables that involve social-demographic info, BD history, psychiatric comorbidities (presence or association of two or more illness in the same patient) and menstrual symptoms.

All variables in the database were used by Apriori for the generation of the association rules. We set the paramaeters of the algorithm as: confidence $\geq 70\%$, support $\geq 10\%$, minimun size rule 2 and maximum 3. The values of minimum confidence and minimum support were chosen because with a low support it is possible to establish events of little recurrence. Based in the minimum confidence, it is possible to obtain rules that present a considerable degree of confidence, in view of this index of 70% allows to show the veracity of the events.

For the purpose of this study, we selected only rules that have the variable *pmdd* with value "*yes*". The reason for this filter is to understand how the PMDD occurrence is associated with other factors such as BD symptoms, other comorbidity disorders, reproductive events and social-demographic data of patients.

The R language was used to run the Apriori algorithm with the *arules, arulesViz, devtools, Rsenal* packets.

IV. Results and Discussion

In total, Apriori found 621,062 association rules. From these rules, 36 were selected for the context of this work, accordin the previous section. Table 1 shows the meaning of the variables that appear in these rules. Tables 2 to 9 show the select rules grouped by their association types. It is also presented the percentage of support and confidence of each rule.

The association rule presented in Table II describes that PMDD is associated with white or Caucasian people. This rule appear in approximately 41% of the base and has confidence of 90%. However, although support and confidence have relatively high values, the cohort of women in STEP-BD mostly consisted of Caucasian women, which limits our ability to generalize these results cross-culturally.

Table III presents rules that associate PMDD with symptoms related to the first 5 years of the menstrual cycle, such as 10-day cycle duration, mood swings and depression, tension and irritability.

These associations with symptoms of the first menstrual cycles (showed in Table III) tend to remain in the current

TABLE I. MEANING OF VARIABLES

Variable	Description
pmdd	definition of pmdd according to DSM-5
race	ethnicity
mf.cycle10	cycle duration around 10 days during the first 5 menstrual years
mf.moodeprs	mood swing and depression during the first 5 menstrual years
mf.tensirrit	tension or irritability during the first 5 menstrual years
mp.irritdepmswing	irritability, depression and mood swing during the menstrual period
mp.physical	physical tiredness during the menstrual period
mp.irritangr	irritability during the menstrual period
mp.bloating	bloating during the menstrual period
mp.foodcrav	food cravings during the menstrual period
mp.length10	cycle duration around 10 days during the menstrual period
mp.cryeasly	cry easily during the menstrual period
mp.abdompan	abdominal pain during the menstrual period
mp.brestend	breast tenderness during the menstrual period
mp.overwhel	feeling overwhelmed during the menstrual period
mp.anx	anxiety during the menstrual period
mp.deprsion	depression during the menstrual period
mp.lackengy	lack of energy during the menstrual period
mp.sxwork	interference in work or school during the menstrual period
mp.moodswing	mood swing during the menstrual period
phobia	past phobia
ptsd	post-traumatic stress disorder
alcdepp	past alcohol dependence
alcabusp	past alcohol abuse
sxpreg	severe mood symptoms during pregnancy
sxppart	severe mood symptoms during postpartum
sxocp	severe mood symptoms during use contraceptive pills
rapidcycl	rapid cycling
antiman	use of antidepressant in the state of mania
bdtype	type of bd
polarity	polarity of the first episode of the BD
anxiety	anxiety disorder

TABLE II. PMDD AND SOCIAL-DEMOGRAPHIC ASSOCIATIONS

Rule	Conf.[%]	Sup.[%]
$pmdd \Rightarrow Caucasian$	90.0	41.0

TABLE III. PMDD AND FIRSTS MENSTRUAL PERIODS ASSOCIATIONS

Rules	Conf.[%]	Sup.[%]
$pmdd \Rightarrow mf.cycle10$	71.0	32.0
$pmdd \Rightarrow mf.moodeprs$	79.5	36.0
$pmdd \Rightarrow mf.tensirrit$	80.5	36.5

TABLE IV. PMDD AND CURRENT MENSTRUAL PERIOD ASSOCIATIONS

Rules	Conf.[%]	Sup.[%]
$pmdd \Rightarrow mp.length10$	79.0	36.0
$pmdd \Rightarrow mp.irritdepmswing$	100	45.5
$pmdd \Rightarrow mp.irritangr$	95.0	43.0
$pmdd \Rightarrow mp.physical$	97.0	44.0
$pmdd \Rightarrow mp.bloating$	84.5	38.5
$pmdd \Rightarrow mp.foodcrav$	81.0	37.0
$pmdd \Rightarrow mp.cryeasly$	79.0	35.5
$pmdd \Rightarrow mp.abdompan$	74.0	33.5
$pmdd \Rightarrow mp.brestend$	72.0	33.0

TABLE V. PMDD AND MENSTRUAL PERIOD MUTUAL ASSOCIATIONS

Rules	Conf.[%]	Sup.[%]
$pmdd \Rightarrow mp.overwhel$	75.0	34.0
$mp.overwhel \Rightarrow pmdd$	82.5	34.0
$pmdd \Rightarrow mp.anx$	82.0	37.5
$mp.anx \Rightarrow pmdd$	78.0	37.5
$pmdd \Rightarrow mp.deprsion$	85.0	39.0
$mp.deprsion \Rightarrow pmdd$	79.0	39.0
$pmdd \Rightarrow mp.lackengy$	79.5	36.0
$mp.lackengy \Rightarrow pmdd$	71.5	36.0
$pmdd \Rightarrow mp.sxwork$	100.0	45.5
$mp.sxwork \Rightarrow pmdd$	81.0	45.5
$pmdd \Rightarrow mp.moodswing$	94.0	42.5
$mp.moodswing \Rightarrow pmdd$	73.5	42.5

menstrual period, according the rules presented in Table IV. The first 3 rules in this Table show that PMDD is also associated with cycle duration around 10 days, irritability, depression and mood swing during the menstrual period. In additionally, other symptoms during the menstrual period appear associated with PMDD as physical tiredness, bloating, food cravings, cry easily, abdominal pain and breast tenderness.

TABLE VI. PMDD AND REPRODUCTIVE LIFE ASSOCIATIONS

Rules	Conf.[%]	Sup.[%]
$sxpreg \,\&\, pmdd \Rightarrow anxiety$	71.5	12.0
$sxppart \,\&\, pmdd \Rightarrow anxiety$	72.0	11.5
$sxocp \,\&\, pmdd \Rightarrow anxiety$	71.0	13.5

TABLE VII. PMDD AND OTHER PSYCHIATRIC COMORBIDITIES ASSOCIATIONS

Rules	Conf.[%]	Sup.[%]
$phobia \,\&\, pmdd \Rightarrow anxiety$	100.0	11.5
$ptsd \,\&\, pmdd \Rightarrow anxiety$	100.0	13.5
$alcdepp \,\&\, pmdd \Rightarrow anxiety$	76.0	10.0
$alcbusp \,\&\, pmdd \Rightarrow anxiety$	76.0	12.5

TABLE VIII. PMDD AND BD ASSOCIATIONS

Rules	Conf.[%]	Sup.[%]
$rapidcycl \,\&\, pmdd \Rightarrow anxiety$	75.5	13.5
$antiman \,\&\, pmdd \Rightarrow anxiety$	70.5	13.0
$bdtype = bpi \,\&\, pmdd \Rightarrow anxiety$	71.5	19.5
$polarity = depressive \,\&\, pmdd \Rightarrow anxiety$	70.0	23.0

Table V shows stronger relationships between PMDD and symptoms of the menstrual period, where the occurrence of PMDD implies a symptom and a symptom implies in PMDD. Symptoms involved in these relationships are: feeling overwhelmed, anxiety, depression, lack of energy, interference in work or school, depression and mood swings with premenstrual worsening during the menstrual period.

Table VI shows associations with the reproductive life of women. Severe mood swings during pregnancy, postpartum and use of contraceptive pills and PMDD appear to be related to anxiety.

Other psychiatric comorbidities such as phobia and post-traumatic stress disorder, as well as alcohol dependence and abuse, and PMDD, are related to anxiety, according to the rules presented in Table VII.

Specific characteristics of the BD: rapid cycling (rapid change between depression and mania), use of antidepressants in the manic state, BD type I and first episode of BD being depression, and PMDD, also imply some anxiety (Table VIII).

The 36 association rules selected have, on average, support of 31%. This value represents the frequency with which the associations appear in the database, considered high in most of the data sets. These same rules also have, on average, confidence of 81%, which also denotes that the generated rules are quite reliable.

The graph of Figure 1 summarizes the relationships among all the variables found in the 36 rules analyzed. These relationships are weighted according to the number of individual occurrences, purple have an moderate number and green have a low number.

From the graph it is possible to observe a lower association rate between PMDD and reproductive life symptoms (variables *sxpreg*, *sxppart*, *sxocp* and comorbidities such as alcohol abuse or dependence (*alcabusp* and *alcdepp*), post-traumatic

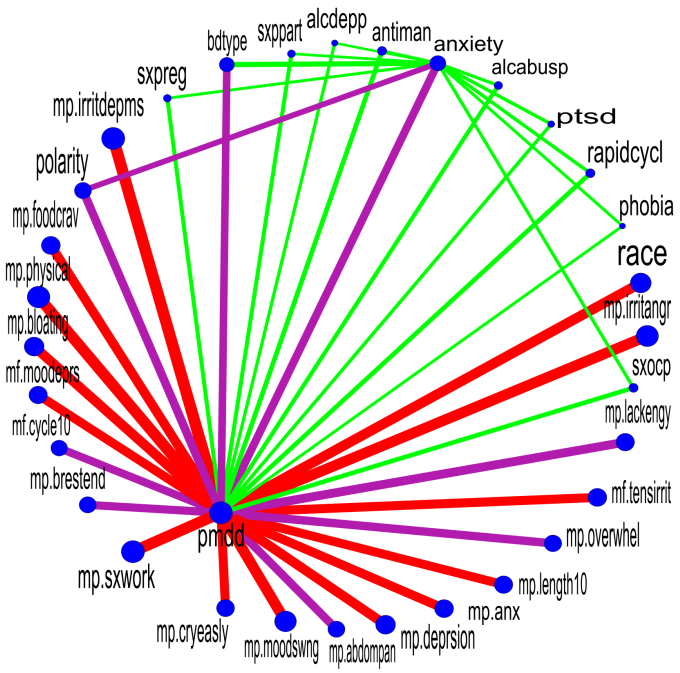

Fig. 1. Associations of variables in all rules

stress (*ptsd*) and phobia, for example (green lines on the graph). On the other hand, there are few occurrences of these characteristics in the database, observed in the size of these nodes.

The intermediate associations (purple lines on the graph) occur with BD characteristics, such as BD type and polarity that also involves anxiety. In addition to other relationships with the reproductive life of the woman as the beginning of the menstrual cycle (*mf.cycle10*) and lack of energy and breast tenderness in the menstrual period (variables *mp.lackengy* and *mp.brestend*).

The strongest PMDD associations that appear in the association rules (red lines on the graph) are related to the other variables of the menstrual period, such as bloating. The race also appear stronger related to PMDD. White or Caucasian people also appear to have a stronger association with PMDD.

V. Conclusion

Bipolar disorder (BD) is a mood disorder in which the patient has episodes of depression and mania. Women which also have the premenstrual dysphoric disorder (PMDD) may have BD symptoms aggravated during their events of Female-Reproductive life.

In this paper we used data from Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD) and the Apriori algorithm for discover associations rules between PMDD, reproductive symptoms, comorbidity disorders and social-demographic information of women with BD. As far as we know, machine learning techniques have not yet been applied to this problem.

The associations presented in the rules are consistent with the work done by Slyepchenko [6]. This study, using classic statistical methods, reports that women with PMDD are more likely to present: rapid cycling (rapid change between depression and mania), increased number of depressive and manic episodes, use of antidepressants in the manic state, depressive BD onset, history of alcohol abuse or dependence, history of post-traumatic stress disorder and mood swings during the female-reproductive life, compared to women that did not have PMDD.

As future work, we intend to apply other Machine Learning methods, especially of the supervised learning paradigm, in order to generate classifiers, and other forms of pattern recognition that could assist professionals working and studying this area.

References

[1] J. R. Geddes and D. J. Miklowitz, "Treatment of bipolar disorder," *The lancet*, vol. 381, no. 9878, pp. 1672–1682, 2013.

[2] L. L. Judd, H. S. Akiskal, P. J. Schettler, J. Endicott, A. C. Leon, D. A. Solomon, W. Coryell, J. D. Maser, and M. B. Keller, "Psychosocial Disability in the Course of Bipolar i and ii Disorders: A prospective, Comparative, Longitudinal Study," *Archives of general psychiatry*, vol. 62, no. 12, pp. 1322–1330, 2005.

[3] V. Bergink, N. Rasgon, and K. L. Wisner, "Postpartum Psychosis:Madness, Mania, and Melancholia in Motherhood," *The American Journal of Psychiatry*, vol. 173, no. 12, pp. 1179–1188, 1 Dec. 2016.

[4] P. C. Cirillo, R. B. F. Passos, M. C. N. Bevilaqua, J. R. R. A. López, and A. E. Nardi, "Bipolar disorder and Premenstrual Syndrome or Premenstrual Dysphoric Disorder comorbidity: a systematic review," *Revista Brasileira de Psiquiatria*, vol. 34, no. 4, pp. 467–479, 25 Apr. 2012.

[5] L.-Y. Hu, C.-C. Shen, J.-H. Hung, P.-M. Chen, C.-H. Wen, Y.-Y. Chiang, and Ti-Lu, "Risk of Psychiatric Disorders Following Symptomatic Menopausal Transition: A Nationwide Population-Based Retrospective Cohort Study," *Medicine*, vol. 95, no. 6, 2016.

[6] A. Slyepchenko, B. N. Frey, B. Lafer, A. A. Nierenberg, G. S. Sachs, and R. S. Dias, "Increased illness burden in women with bipolar and premenstrual dysphoric disorder: Data from 1,099 women," *Acta Psychiatrica Scandinavica*, 2017.

[7] H.-U. Wittchen, E. Becker, R. Lieb, and P. Krause, "Prevalence, incidence and stability of premenstrual dysphoric disorder in the community," *Psychological Medicine*, vol. 32, no. 01, pp. 119–32, 2002.

[8] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*, American Psychiatric Association, Ed., 2013.

[9] World Health Organization, *International Statistical Classification of Diseases and Related Health Problems*, World Health Organization, Ed., 2010, vol. 2.

[10] G. S. Sachs, M. E. Thase, M. W. Otto, M. Bauer, D. Miklowitz, S. R. Wisniewski, P. Lavori, B. Lebowitz, M. Rudorfer, E. Frank, A. A. Nierenber, M. Fava, C. Bowden, T. Ketter, L. Marangell, J. Calabrese, D. Kupfer, J. F. Rosenbaum, and M. W. Otto, "Rationale, design, and methods of the systematic treatment enhancement program for bipolar disorder (STEP-BD)," *Biological Psychiatry*, vol. 53, no. 11, pp. 1028–1042, 1 Jun. 2003.

[11] M. Kantardzic, *Data Mining: Concepts, Models, Methods, and Algoritms*. Wiley-IEEE Press, 2011.

[12] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," *ACM SIGMOD Record*, vol. 22, no. 2, pp. 207–216, 1993.

[13] R. Agrawal and R. Srikant, "Fast Algortihms for mining association rules in large databases," *International Conference on Very Large Data bases*, 1994.