Quantum Apriori Algorithm

Amanda G. Valério[1], Celso J. Villas-Bôas[1]

[1]*Physics Departament, Federal University of São Carlos (UFSCar), Brazil*

In classical computing, the Apriori algorithm [1, 2] is an association rules algorithm that can be classified into both data mining and artificial intelligence. Its main application is to generate recommendations by discovering patterns, correlations or associations in a database.

Let X and Y be sets of one or more items contained in the database, where $X \cap Y = \emptyset$, an association rule is an implication in the form $X \Rightarrow Y$, found in database transactions [5]. Association rules must be measured by numerical parameters that attest to the degree of correlation between the items in the set. Originally, Apriori uses the Support and Confidence metrics, which will be detailed below.

Support calculates the proportion of each item or set of items in the database transactions, as seen in equations 1 and 2.

$$\text{Support (X)} = \frac{\text{no. of transactions containing X}}{\text{total no. of transactions}}; \tag{1}$$

$$\text{Support (X, Y)} = \frac{\text{no. of transactions containing X and Y}}{\text{total no. of transactions}}. \tag{2}$$

The Confidence represents the proportion of times an item appears in combination with another $(X \cup Y)$, within all the occurrences of that first item (X), making it a measure of the "strength" of the association rule [5].

$$\text{Confidence(X, Y)} = \frac{\text{Support(X, Y)}}{\text{Support(X)}} \tag{3}$$

The Confidence rule has the limitation of not being able to correctly capture the correlation between the antecedent and consequent of association rules, a task that is better performed by the Lift measure [4]. This, in turn, is the proportion with which item X is present in transactions in combination with item Y, also taking into account the proportion of occurrences of item Y. This is represented in equation below.

$$\textit{Lift}\text{(X, Y)} = \frac{\text{Confidence (X,Y)}}{\text{Support (Y)}}$$

This association rule can take on 3 different meanings that show the correlations between antecedent and consequent, depending on its value [4]:

- if $\textit{Lift}\text{(X, Y)} > 1$: there is a positive correlation between the items;
- if $\textit{Lift}\text{(X, Y)} < 1$: there is a negative correlation between the items;
- if $\textit{Lift}\text{(X, Y)} = 1$: the correlation is independent.

Observing the properties of the Lift parameter, it can be seen that there is a similarity to the second-order correlation function in two-photon counting processes, which is well known in Quantum Optics. The second-order function is able to distinguish the nature of the light source by considering the coherence properties of light in the second order. It is not possible to distinguish quantum effects with the first-order function, but higher than second orders are also capable of the same feat, but this research will not address this topic so as not to deviate from the objective of generating a recommendation [3].

For a one-mode field, the second-order correlation function is written as [6]:

$$g^{(2)}(\tau) = \frac{\langle a^{\dagger}(t)a^{\dagger}(t+\tau)a(t+\tau)a(t)\rangle}{\langle a^{\dagger}(t)a(t)\rangle^2}, \tag{4}$$

where $a^\dagger$ and a, respectively, the creation and destruction operators, t the time variable and τ a positive variable indicating a later time. For a coherent state (the state of the light emitted by a good laser), we have that $g^{(2)}(0) = 1$. This also results in a Poisson distribution of the number of photons. In a number state (self-harmonic oscillator state) $\rho = |n\rangle \langle n|$, we have that $g^{(2)}(0) = 1 - \frac{1}{n}$, $n > 2$. Furthermore:

- if $g^{(2)}(\tau) < g^{(2)}(0)$, is called photon bunching;
- if $g^{(2)}(\tau) > g^{(2)}(0)$, is called photon antibunching.

Based on this observation, a quantum version of the Apriori algorithm is in development. It is expected that this novel algorithm will be efficient for solving real-world problems on a quantum computer, due to the translation of the problem into the quantum universe of the hardware.

References

[1] Agrawal, R., Srikant, R. & Others Fast algorithms for mining association rules. *Proc. 20th Int. Conf. Very Large Data Bases, VLDB.* **1215** pp. 487-499 (1994)

[2] Hegland, M. The apriori algorithm–a tutorial. *Mathematics And Computation In Imaging Science And Information Processing.* pp. 209-262 (2007)

[3] Scully, M. & Zubairy, M. Quantum optics. (American Association of Physics Teachers,1999)

[4] Hussein, N., Alashqur, A. & Sowan, B. Using the interestingness measure lift to generate association rules. *Journal Of Advanced Computer Science & Technology.* **4**, 156 (2015)

[5] Kotsiantis, S. & Kanellopoulos, D. Association rules mining: A recent overview. *GESTS International Transactions On Computer Science And Engineering.* **32**, 71-82 (2006)

[6] Walls, D. & Milburn, G. Quantum Optics. (Springer Berlin Heidelberg,2008), https://books.google.com.br/books?id=LiWsc3Nlf0kC