

Study of the evolution of antiemetic treatment through the application of Triadic Formal Concept Analysis

Lucas Ferreira¹, Cristiane Nobre¹, Luis Zárate¹, Mark Song¹

Pontifícia Universidade Católica de Minas Gerais
contatolucasgsf@gmail.com, {nobre, zarate, song}@pucminas.br

Abstract. Cancer treatment is always improving with new techniques and drugs, applied in a process with rigorous medical analysis. One way to analyze medical outcomes is to test a new drug in one group and placebos in another. In this article, we used a database with the results of a new drug for preventing chemotherapy-induced nausea and vomiting. After processing the database, we apply the Triadic Formal Concept approach to extract triadic rules, implications, and conditions that are used to identify correlations throughout the medical sessions. Our study shows that triadic analysis can help scientists perform clinical analyzes efficiently by looking at patient history data.

CCS Concepts: • **Computing methodologies** → **Association Rules**.

Keywords: Drug Therapy, Antiemetics, Data Mining, Triadic Concept Analysis, Triadic Rules

1. INTRODUÇÃO

O câncer consiste em doenças malignas que têm em comum o crescimento desordenado de células e podem invadir tecidos adjacentes e órgãos a distância. Atualmente dispomos de vários tipos de tratamento para essas doenças, entre elas a quimioterapia, que ao longo de várias sessões, bombardeia o tumor com substâncias químicas especiais através de infusões venosas, injeções ou comprimidos. Esse tratamento normalmente causa efeitos colaterais, figurando entre os mais comuns a queda de cabelo, náusea, vômito, e dor no corpo. Para a prevenção desses efeitos colaterais normalmente são administrados alguns medicamentos, entre eles os fármacos antieméticos, cujo principal objetivo é evitar a êmese.

O objetivo desse trabalho é a aplicação de um processo para extrair conhecimento de uma base de dados gerada por uma pesquisa realizada em 120 mulheres chinesas, que realizaram tratamento quimioterápico para câncer de mama, e que foram separadas aleatoriamente em 2 grupos de mesmo tamanho para comparar tratamentos antieméticos utilizando ou não olanzapina. Após um pré-processamento da base de dados, composta principalmente pelo tratamento de valores nulos, para obter os resultados das sessões após cada ciclo de quimioterapia, propomos a aplicação da teoria de conceitos [Ganter and Wille 1999], derivados da Análise Formal de Conceitos (AFC) para a análise dos dados.

A teoria da Análise Formal de Conceitos é um campo da matemática aplicada cujo objetivo é representar e extrair conhecimento através de bases de dados denominadas contextos [Singh et al. 2016]. Esta análise possibilita obter um conjunto de conceitos organizados hierarquicamente e, a partir destes, extrair regras que os relacionam.

Descrever os resultados dos tratamentos antieméticos pode ser significativo para entender como as diferentes substâncias utilizadas nesse processo afetam a qualidade de vida dos pacientes que realizam o tratamento quimioterápico. E os resultados apresentados neste trabalho mostram que a análise triádica pode ser eficiente para auxiliar os médicos a acompanharem o desempenho de determinados fármacos. Com os resultados pode-se verificar o comportamento do indivíduo antes do tratamento, durante as fases em que o fármaco foi aplicado e posteriormente ao término do uso do mesmo.

Este artigo está organizado da seguinte forma: A Seção 2 apresenta a fundamentação teórica uti-

lizada, a Seção 3 está o trabalho relacionado, a Seção 4 traz a metodologia, na Seção 5 estão os resultados do trabalho, e, por último, a Seção 6 traz a conclusão e os trabalhos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Análise Formal de Conceitos e Regras triádicas

A Análise Formal de Conceitos (AFC) é um ramo da matemática aplicada que está centrada nos seguinte elementos: Contexto Formal, Conceito Formal e Regras de Implicação [Ganter and Obiedkov 2016]. Um contexto formal consiste de dois conjuntos G e M , e uma relação de incidência I entre os mesmos, representada por $I \subseteq G \times M$. Os elementos do conjunto G indicam os objetos presentes na base, enquanto os elementos de M os atributos/características associadas. O contexto é representado pela notação $K := (G, M, I)$. Toda incidência gIm ou $(g, m) \in I$ indica que um objeto g possui certa característica m . Uma implicação $X \rightarrow Y$, válida para um contexto $K := (G, M, I)$, significa que cada objeto possuindo todos os atributos em X possui também todos os atributos em Y .

Para quaisquer subconjuntos $A \subseteq G$ e $B \subseteq M$, utiliza-se a operação de derivação ($'$) que produz [Wille 2002]:

- 1) $A' := \{m \in M \mid gIm \forall g \in A\}$
- 2) $B' := \{g \in G \mid gIm \forall m \in B\}$

A AFC lida apenas com dados bi-dimensionais. Em certas aplicações é preciso manipular dados tridimensionais, como o estudo apresentado neste trabalho. Neste caso, faz-se o uso da Análise Formal de Conceito Triádica (TCA). O Contexto Triádico é definido por uma quádrupla $K = (G, M, B, Y)$, em que $Y \in G \times M \times B$, em que G, M, B são objetos, atributos e condições, respectivamente. Portanto, $(g, m, b) \in Y$ pode ser lido como “o objeto g tem o atributo m sob a condição b ”. Um exemplo de contexto triádico está representado na Tabela I, que contém 2 ciclos de tratamento com 3 atributos cada (a1, a2, a3) [Zhuk et al. 2014].

	Ciclo 1			Ciclo 2		
Obj	a1	a2	a3	a1	a2	a3
1	x			x		x
2	x	x	x	x	x	x
3		x		x	x	
4			x			x

Tabela I: Exemplo de Contexto Triádico

Pode-se descrever dois tipos de regras de associação triádicas que são ser extraídas do contexto $K := (G, M, B, Y)$ [Biedermann 1997]: Biedermann Conditional Attribute Association Rule (BCAAR) e Biedermann Attributional Condition Association Rule (BACAR) [Zhuk et al. 2014].

A regra BCAAR é representada por: $R \rightarrow S \ C$ (sup.,conf.), onde $R, S \subseteq M$ e $C \subseteq B$. Ou seja, cada objeto possuindo todos os atributos em R possui também todos os atributos em S sob uma condição C , com um suporte (sup.) e uma confiança (conf.). A regra BACAR é representada por: $P \rightarrow Q \ N$ (sup.,conf.), onde $P, Q \subseteq B$ e $N \subseteq M$. Ou seja, cada objeto sob as condições em P também estará sob as condições em Q no atributo N , com um suporte (sup.) e uma confiança (conf.).

O suporte corresponde à proporção de objetos no subconjunto $g \in G$ que satisfazem a implicação $P \rightarrow Q$, em relação ao número total de objetos $|G|$ do contexto formal K (Equação 1) onde ($'$) corresponde ao operador de derivação.

$$Support(P \rightarrow Q) = \frac{|(P \cup \{Q\})'|}{|G|} \quad (1)$$

A confiança corresponde à razão dos objetos $g \in G$ que contêm P , que também contêm Q , em relação ao número total de objetos $|G|$ (Equação 2).

$$Confidence(P \rightarrow Q) = \frac{|(P \cup \{Q\})'|}{|P'|} = \frac{Support(P \rightarrow Q)}{Support(P)} \quad (2)$$

2.2 Quimioterapia e Medicamentos Antieméticos

As náuseas e os vômitos induzidos por quimioterapia são efeitos secundários comuns relacionados ao tratamento de câncer, sendo bastante temidos pelos pacientes. Esses sintomas têm um efeito mental que prejudica a qualidade de vida de doentes com neoplasia maligna e pode levar a reduções da dose e, em alguns casos, até mesmo à suspensão do tratamento.

Para evitar esses efeitos colaterais da quimioterapia, são utilizados fármacos chamados antieméticos. Recentemente, estão aparecendo evidências de benefícios com o uso do agente neuroléptico olanzapina. Um estudo recente mostrou que essa medicação, associada com outras, em pacientes com carcinoma de mama em esquema altamente emetogênico teve sucesso de controle total dos sintomas em 75% dos casos [Navari et al. 2016], sendo notavelmente mais eficaz para o controle de náuseas durante a fase tardia. É recomendada para pacientes recebendo agentes altamente emetogênicos que têm êmese refratária induzida pela quimioterapia, bem como para aqueles nos quais um alto nível de ansiedade é o principal gatilho para os vômitos induzidos pelo tratamento quimioterápico. Por outro lado, os eventos adversos mais comuns são sonolência, hipotensão postural e constipação.

As respostas ao tratamento antiemético podem ser classificadas entre [Ng et al. 2015]:

- (1) **Proteção Completa:** Não houve êmese, não foi utilizado medicamento de emergência e moderada ocorrência de náusea.
- (2) **Controle Total:** Não houve êmese, não foi utilizado medicamento de emergência e pouca ocorrência de náusea.
- (3) **Resposta Completa:** Não houve êmese e não foi utilizado medicamento de emergência e não ocorreu náusea.

3. TRABALHOS RELACIONADOS

O uso da Análise Formal de Conceitos foi utilizada em uma base de dados para prover um mecanismo de busca de prontuários para pacientes em tratamento de câncer de mama. Para isso, foi combinada a AFC com a lógica de descrição (DL – Description Logic) [Hu et al. 2004]. A lógica de descrição é uma família de representação de conhecimento, que se baseia em conceitos e propriedades, e é considerada crucial para o formalismo lógico para ontologias e Web Semântica, e ao serem combinadas com AFC, permitem a implementação de um mecanismo de busca. Nosso trabalho difere deste por não utilizar lógica de descrição, visto que, o intuito é extrair regras de associação triádicas presentes na base.

Em [Yeo et al. 2020] disponibiliza-se uma base de dados contendo registros de pacientes que foram avaliados durante 4 ciclos de quimioterapia, entretanto os autores não aplicaram AFC/TCA para obtenção de regras para a análise dos resultados. Além disso, o presente trabalho utiliza a ferramenta Lattice Miner, proposta em [Missaoui and Emamirad 2017], para a geração das regras triádicas.

4. METODOLOGIA

O objetivo desse trabalho é descrever as relações entre os atributos de um mesmo ciclo e também ao longo do tempo, utilizando regras de associação triádicas extraídas a partir de um valor mínimo de suporte e confiança. Com isso, é possível interpretar e entender os potenciais benefícios do novo fármaco antiemético no controle dos efeitos colaterais advindos do tratamento quimioterápico.

4.1 Materiais

Para esse trabalho foi utilizada a base de dados disponibilizada em [Yeo et al. 2020] que contém registros de 120 pacientes que foram avaliados durante 4 ciclos de quimioterapia, divididos em dois grupos de 60 pacientes, sendo que, no primeiro foi administrado o tratamento antiemético padrão, e no segundo foi adicionada a olanzapina.

É importante salientar que mesmo a base de dados possuindo um número reduzido de pacientes, ainda assim foi possível encontrar padrões entre as diferentes fases do tratamento definidas pela Tabela II, durante os 4 ciclos de quimioterapia, que são as condições desta Análise Triádica.

Atributo	Descrição
GR	Período geral. Intervalo de 0 a 120h após a sessão de quimioterapia.
AG	Período agudo. Intervalo de 0 a 24h após a sessão de quimioterapia.
TD	Período tardio. Intervalo de 24 a 120h após a sessão de quimioterapia.

Tabela II: Atributos

4.2 Pré-Processamento

Entre os tipos de resposta contidas na base de dados encontrada, foram utilizadas as respostas dos pacientes que indicaram Resposta Completa ao tratamento antiemético, pois se adequa melhor ao objetivo da nossa análise. Após essa seleção, foram realizadas na base as seguintes etapas de pré-processamento, descritos também em [Yeo et al. 2020]:

- (1) Conversão de um questionário de acompanhamento pós quimioterápico de 5 dias de cada sessão quimioterápica em uma base de dados.
- (2) Para eficácia de análise através de vários ciclos, foi utilizado o teste qui-quadrado para dados dicotômicos.

Após essa extração, ocorreu uma substituição nos valores que se encontravam nulos e tinham uma alta correlação com outros atributos, pois ao não ser constatado efeito no atributo GR (Geral), de forma relacionada, o mesmo acontece com os atributos AG (Agudo) e TD (Tardio).

Ainda sim, entre os 60 pacientes tratados com o fármaco olanzapina, ainda existiam valores nulos em que não seria possível preenchê-los, portanto, essas instâncias foram removidas. Sendo assim, após o pré-processamento, resultaram 56 objetos (pacientes) tratados com olanzapina, dando origem ao contexto triádico parcialmente apresentado na Tabela III. Cada linha representa um paciente, as colunas superiores um ciclo do tratamento, e as colunas inferiores (que se repetem) as condições associadas a cada ciclo.

Para obter os resultados desejados, utilizou-se a ferramenta Lattice Miner 2.0, considerada uma boa alternativa, para a extração das regras quando a base de dados possui um número reduzido de atributos e/ou objetos.

Para realizar a extração, toda a base foi convertida para o formato JSON (o padrão de entrada da ferramenta). A partir de então, alguns parâmetros foram informados como, por exemplo, a confiança e

Obj	Ciclo 1			Ciclo 2			Ciclo 3			Ciclo 4		
	GR	AG	TD	GR	AG	TD	GR	AG	TD	GR	AG	TD
1	x	x	x	x	x	x	x	x	x	x	x	x
2	x	x	x	x	x	x	x	x	x			
3	x	x	x	x	x	x	x	x	x	x	x	x
4					x		x	x	x		x	
5												
6				x	x	x	x	x	x		x	
7	x	x	x							x	x	x
8	x	x	x		x		x	x	x	x	x	x
9							x	x	x	x	x	x
10					x					x	x	x

Tabela III: Contexto Triádico Completo

suporte mínimos desejados, indicando após a leitura da entrada a classificação entre objetos, atributos e condições.

5. RESULTADOS

Tipo	Quantidade gerada	Quantidade após análise
BACARs	26	13
BCAARs	32	9

Tabela IV: Regras Geradas

No contexto longitudinal apresentado, foram encontradas 26 regras do tipo BCAARs, e 32 regras do tipo BACARs, ambas com suporte mínimo de 50% e confiança mínima de 90% (Tabela IV). É importante ressaltar que as regras triádicas são regras de associação geradas dentro do contexto.

Ressalta-se, também, que a ferramenta gera regras BACARs e BCAARs sem respeitar uma sequência de ciclos. Desta forma, realizou-se posteriormente uma triagem nas regras de implicações geradas a fim de se respeitar a sequência cronológica dos resultados, que em nosso contexto é caracterizada pelas sessões de quimioterapia.

Isso significa que regras geradas como: $(c3 \rightarrow c2) AG$ [sup. = 76,8% conf. = 93,5%], em que um ciclo posterior se relaciona com o anterior foram excluídas da análise final, gerando um subconjunto de regras como apontado pela Tabela IV. Como resultado tem-se, as seguintes regras:

(1) BACARs

- (a) $(c3 \rightarrow c4) AG$ [sup. = 76,8% conf. = 93,5%]
- (b) $(c2 \rightarrow c3) AG$ [sup. = 71,4% conf. = 90,9%]
- (c) $(c1 \rightarrow c3) AG$ [sup. = 62,5% conf. = 92,1%]
- (d) $(c1 \rightarrow c4) AG$ [sup. = 62,5% conf. = 92,1%]
- (e) $(c3,c2 \rightarrow c4) AG$ [sup. = 66,1% conf. = 92,5%]
- (f) $(c1,c2 \rightarrow c3) AG$ [sup. = 57,1% conf. = 97,0%]
- (g) $(c1,c2 \rightarrow c4) AG$ [sup. = 55,4% conf. = 93,9%]
- (h) $(c1,c2 \rightarrow c3) GR$ [sup. = 51,8% conf. = 96,7%]
- (i) $(c1,c2 \rightarrow c4) GR$ [sup. = 51,8% conf. = 96,7%]
- (j) $(c1,c2 \rightarrow c3) TD$ [sup. = 51,8% conf. = 96,7%]
- (k) $(c1,c2 \rightarrow c4) TD$ [sup. = 51,8% conf. = 96,7%]
- (l) $(c3,c1,c2 \rightarrow c4) GR$ [sup. = 50,0% conf. = 96,6%]
- (m) $(c3,c1,c2 \rightarrow c4) TD$ [sup. = 50,0% conf. = 96,6%]

(2) **BCAARs**

- (a) (AG → GR,TD) c3 [sup. = 75,0% conf. = 91,3%]
- (b) (AG → GR,TD) c1 [sup. = 64,3% conf. = 94,7%]
- (c) (GR → AG,TD) c2 [sup. = 69,6% conf. = 100,0%]
- (d) (TD → AG,TD) c4 [sup. = 73,2% conf. = 100,0%]
- (e) (TD → AG,GR) c3,c4 [sup. = 67,9% conf. = 100,0%]
- (f) (GR → AG,TD) c1,c2 [sup. = 53,6% conf. = 100,0%]
- (g) (GR → AG,TD) c2,c4 [sup. = 60,7% conf. = 100,0%]
- (h) (TD → AG,GR) c3,c2,c4 [sup. = 58,9% conf. = 100,0%]
- (i) (TD → AG,GR) c3,c1,c2,c4 [sup. = 50,0% conf. = 100,0%]

As regras BACARs descrevem relações entre determinados atributos em diferentes condições, ou seja, em nosso contexto, indica a relação do efeito do medicamento entre 2 ou mais ciclos em determinado período (geral, agudo ou tardio).

Tome como exemplo, a regra 1b: (c2 → c3) AG [sup. = 71,4% conf. = 90,9%]. É possível observar que as pessoas que reportaram a efetividade do medicamento nas primeiras 24h após o segundo ciclo quimioterápico também reportaram o efeito no terceiro ciclo do tratamento, ocorrendo em 71,4% dos pacientes. O efeito se repete também do terceiro para o quarto ciclo, mas agora, em 76,8% dos casos, como apontado pela Regra 1a: (c3 → c4) AG [sup. = 76,8% conf. = 93,5%].

Já as regras BCAARs descrevem relações entre condições em atributos distintos, portanto, no nosso caso, evidencia a relação do efeito do antiemético entre 2 ou mais períodos (geral, agudo ou tardio). Dentro do mesmo ciclo, consegue-se visualizar agora se o efeito nas primeiras 24h após a quimioterapia se estende durante os 5 dias observados.

De acordo com a Regra 2b: (AG → GR,TD) c1 [sup. = 64,3% conf. = 94,7%], durante o primeiro ciclo (c1), em 64,3% das vezes, se o fármaco teve efeito nas primeiras 24h (AG), ele persiste para as demais 96h observadas (ou seja, durante 120 horas - GR e TD). Efeito semelhante pode ser observado no terceiro ciclo, mas agora em 75% dos casos, de acordo com a Regra 2a: (AG → GR,TD) c3 [sup. = 75,0% conf. = 91,3%].

6. CONCLUSÃO E TRABALHOS FUTUROS

Os resultados obtidos neste trabalho colaboram com os estudos existentes para os tratamentos antieméticos, oferecendo uma maior compreensão sobre a eficiência de novos fármacos quando administrados para pacientes submetidos à quimioterapia emetogênica.

O estudo mostra a evolução dos mesmos ao longo das sessões quimioterápicas através de regras triádicas, auxiliando médicos e pesquisadores em diferentes aspectos, desde o teste de eficiência de novas propostas medicinais quanto à administração das mesmas perante a resposta do paciente. Ressalta-se que poucos são os trabalhos que analisam, através de regras triádicas, o efeito de medicamentos em pacientes.

Como trabalhos futuros, indica-se a busca por bases de dados de maior dimensionalidade para, além de validar as regras de associação apresentadas nos resultados desse trabalho, aprofundar ainda mais no que diz respeito à classificação dos objetos, tais como lógicas de descrição, ou algoritmos de machine learning. Além disso, procurar analisar as regras que descrevem comportamentos passados baseados em ciclos futuros, como os que foram removidos de nossa análise.

REFERENCES

- BIEDERMANN, K. How triadic diagrams represent conceptual structures. *Conceptual Structures: Fulfilling Peirces Dream Lecture Notes in Computer Science*, 1997.
- GANTER, B. AND OBIEDKOV, S. Conceptual exploration, 2016.
- GANTER, B. AND WILLE, R. Formal concept analysis, 1999.
- HU, B., DASMAHAPATRA, S., DUPPLAW, D., LEWIS, P., AND SHADBOLT, N. Managing patient record instances using dl-enabled formal concept analysis. *Engineering Knowledge in the Age of the Semantic Web Lecture Notes in Computer Science*, 2004.
- MISSAOUI, R. AND EMAMIRAD, K. Lattice miner-a formal concept analysis tool. *14th International Conference on Formal Concept Analysis*, 2017.
- NAVARI, R. M., QIN, R., RUDDY, K. J., LIU, H., POWELL, S. F., BAJAJ, M., DIETRICH, L., BIGGS, D., LAFKY, J. M., LOPRINZI, C. L., AND ET AL. Olanzapine for the prevention of chemotherapy-induced nausea and vomiting. *New England Journal of Medicine* 375 (2): 134–142, 2016.
- NG, T. L., HUTTON, B., AND CLEMONS, M. Chemotherapy-induced nausea and vomiting: Time for more emphasis on nausea? *The Oncologist* 20 (6): 576–583, 2015.
- SINGH, P. K., KUMAR, C. A., AND GANI, A. A comprehensive survey on formal concept analysis, its research trends and applications. *International Journal of Applied Mathematics and Computer Science* 26 (2): 495–516, 2016.
- WILLE, R. Why can concept lattices support knowledge discovery in databases? *Journal of Experimental & Theoretical Artificial Intelligence* 14 (2-3): 81–92, 2002.
- YEO, W., CHAN, V. T., LI, L., LAU, T. K., LAI, K. T., PANG, E., CHEUNG, M., AND MO, F. K. Dataset on chemotherapy-induced nausea and vomiting (cinv) and quality of life (qol) during multiple chemotherapy cycles among a chinese breast cancer patient population who were randomized to antiemetic regimens with or without olanzapine. *Data in Brief* vol. 30, pp. 105421, 2020.
- ZHUK, R., IGNATOV, D. I., AND KONSTANTINOVA, N. Concept learning from triadic data. *Procedia Computer Science* vol. 31, pp. 928–938, 2014.