

# Um sistema de suporte à decisões de diagnóstico de algumas doenças epidemiológicas

**Marcus Vinícius Benedito e Jacques Wainer**

<sup>1</sup>Instituto de Computação (IC), Universidade Estadual de Campinas (UNICAMP)  
Caixa postal 6176 – 13084-971 – Campinas – SP – Brasil

marcus.benedito@students.ic.unicamp.br, wainer@ic.unicamp.br

**Abstract.** *This paper presents decision support system that helps diagnoses thirty six infections diseases of interest to the Brazilian National Health Surveillance Agency (ANVISA). The system is based on abductive inference mechanism that uses parsimonious covering theory (PCT) with some modifications. Instead of using only the diseases associated with symptoms as in PCT, we propose to associate relevant factors that are not symptoms, determinant information sets, that without any other information determines that a patient is suspicious of having a disease, we used almost obligatory information of some symptoms in particular diseases and we eliminate the possibility of having multiple dysfunctions simultaneously.*

**Resumo.** *Este artigo apresenta um sistema de suporte à decisões que auxilia na diagnose de trinta e seis doenças epidemiológicas de interesse da Agência Nacional de Vigilância Sanitária Brasileira (ANVISA). O sistema é baseado no mecanismo de inferência abdutivo que usa a teoria das coberturas parcimoniosas (TCP) com algumas modificações. Ao invés de utilizar apenas as associações entre doenças e sintomas, como na TCP original, nós propomos associar também fatores relevantes que não são sintomas, conjuntos determinantes de informações, que sem depender de qualquer outra informação determina a suspeita de alguma doença, o conceito de sintomas quase obrigatórios e eliminamos a possibilidade de haver múltiplas disfunções simultâneas.*

## 1. Introdução

O diagnóstico médico de doenças epidemiológicas é de suma importância para retenção no avanço de doenças, incluindo possível proliferação em grande escala, e também para o tratamento de pacientes já infectados. Por estas razões, as agências de vigilância sanitária voltam grande parte de seus esforços para a identificação precoce destas e seus tratamentos.

Atualmente, um dos maiores problemas enfrentados por estas agências é o número limitado de especialistas para diagnosticar estas doenças. Em diversas regiões de muitos países, incluindo o Brasil, existem apenas alguns poucos médicos generalistas disponíveis para a população. A ANVISA define trinta e sete doenças como sendo de seu “interesse epidemiológico” [3], as quais requerem, obrigatoriamente, o envio de relatórios para a agência no caso de descobertas de novos casos de contaminação. As doenças de interesse epidemiológico são: AIDS, antraz ou carbúnculo, botulismo, cólera, coqueluche, dengue, difteria, doença de Chagas, esquistossomose mansônica, febre amarela, febre do Nilo

ocidental, febre maculosa brasileira, febre purpúrica brasileira, febre tifóide, hanseníase, hantaviroses, hepatites virais, gripe ou influenza, leishmaniose tegumentar americana, leishmaniose visceral, leptospirose, malária, meningites, parotidite infecciosa, peste, poliomelite, raiva, rubéola, sarampo, sífilis congênita, síndrome de rubéola congênita, tétano accidental, tétano neonatal, tracoma, tuberculose, varicela/herpes zoster e varíola.

O sistema descrito, neste artigo, fornece suporte à tomada de decisão de médicos generalistas e agentes de saúde no processo de geração de diagnóstico de trinta e seis doenças, dentre as trinta e sete destacadas. A única doença não tratada neste momento é a AIDS, pois não é diagnosticada diretamente pelos seus sintomas, mas sim por infecções oportunistas subseqüentes.

A TCP é utilizada desde sua criação [4, 6] para solucionar problemas de diagnóstico. Ela é baseada no mecanismo de inferência abdutivo (mecanismo da melhor explicação), ou seja, as hipóteses que melhor explicam as evidências são selecionadas para formar a diagnose. Por sua vez, todos os conjuntos parcimoniosos de doenças que poderiam ter sido a causa de um conjunto de manifestações observáveis são levantados como hipóteses.

Apesar de ser amplamente utilizada para diagnósticos em geral esta teoria possui algumas limitações. No diagnóstico médico de doenças epidemiológicas, por exemplo, devemos lidar com a imprecisão das descrições de manifestações fornecidas pelo paciente, tratar o fato de raramente acontecer a infecção por duas ou mais desordens simultaneamente, além de não utilizarmos a informação de que na presença de algumas doenças existem sintomas que são quase obrigatórios. Existem também fatores relevantes que não são sintomas, mas que poderiam auxiliar de forma bastante significativa na busca da solução do problema. Sabemos também que para gerar a suspeita de algumas doenças é necessário apenas um conjunto de informações específicas independentemente de quaisquer outras que possam existir. Tanto informações relevantes que não são sintomas quanto esta estrutura de conjuntos de informações determinantes de suspeita não são suportadas pela TCP em sua forma original.

Este artigo é organizado da seguinte forma. Na seção 2 mostramos a estrutura da TCP e suas limitações em detalhes. Na seção 3 será detalhado o modelo teórico proposto neste trabalho. Na seção 4 o sistema desenvolvido utilizando o modelo proposto é demonstrado através de exemplos de execução e suas explicações. Na seção 5 concluímos.

## 2. Teoria das coberturas parcimoniosas (TCP)

### 2.1. Formulação do problema de diagnóstico utilizando a TCP

A versão básica da TCP [6], que foi adotada em [7, 8], utiliza dois conjuntos finitos especiais para definir o escopo do problema de diagnóstico em questão. Estes são o conjunto D, que representa todas as possíveis **desordens**<sup>1</sup>, ou doenças,  $d_i$  que podem ser manifestadas, e o conjunto M, que representa todas as **manifestações**<sup>2</sup>, ou sintomas,  $m_j$  que podem ocorrer na presença de uma ou mais doenças.

Uma relação C, de D para M, associa cada doença aos sintomas que esta possa causar. Uma associação  $(d_i, m_j) \in C$  nos diz que  $d_i$  pode causar diretamente  $m_j$ . Isto não

<sup>1</sup>De agora em diante chamadas apenas de doenças.

<sup>2</sup>De agora em diante chamadas apenas de sintomas.

significa que  $d_i$  necessariamente causa  $m_j$ , mas apenas indica a *possibilidade* de fazê-lo. Os conjuntos D, M e a relação C estão na **base de conhecimento** (BC) de um problema de diagnóstico.

Para completar esta formulação precisamos de um **caso** K específico. Usamos  $M^+$ , um subconjunto de M, para denotar o conjunto de **observações**, ou seja, o conjunto de sintomas presentes em K.

**Definição 1** Um problema de diagnóstico P é representado pelo par  $(BC, K)$  onde:

- $BC = (D, M, C)$  : Base de conhecimento:
  - $D = \{d_1, d_2, \dots, d_m\}$ ,  $m \geq 1$  : é um conjunto finito, não-vazio, de objetos chamados *doenças*.
  - $M = \{m_1, m_2, \dots, m_n\}$ ,  $n \geq 1$  : é um conjunto finito, não-vazio, de objetos chamados *sintomas*.
  - $C \subseteq D \times M$  : é uma relação, de D para M, onde  $(d_i, m_j) \in C$  se  $d_i$  pode causar  $m_j$ . Esta é chamada de *associação causal*.
- $K = (M^+)$ , onde K é o caso.  $M^+ \subseteq M$  é o conjunto de observações.

## 2.2. Solução de um problema de diagnóstico utilizando a TCP

Para formalizar a geração de um diagnóstico, a TCP utiliza os conceitos de **cobertura** e **parcimônia** que são a base desta teoria.

**Definição 2** Para qualquer  $d_i \in D$ ,  $m_j \in M$ ,  $D_I \subseteq D$  e  $M_J \subseteq M$  em um problema de diagnóstico P, temos:

- $efeito(d_i) = \{m_j | \langle d_i, m_j \rangle \in C\}$  : conjunto de sintomas manifestados por  $d_i$ .
- $efeito(D_I) = \bigcup_{d_i \in D_I} efeito(d_i)$ .
- $causa(m_j) = \{d_i | \langle d_i, m_j \rangle \in C\}$  : conjunto de doenças que podem ser a causa de  $m_j$ .
- $causa(M_J) = \bigcup_{m_j \in M_J} causa(m_j)$ .

**Definição 3** O conjunto  $D_I \subseteq D$  é uma cobertura de  $M_J \subseteq M$  se  $M_J \subseteq efeito(D_I)$ .

Portanto, cobertura é um conjunto de doenças que pode causar *todos* os sintomas presentes em  $M_J \subseteq M$ .

Já o conceito de parcimônia possui definição imprecisa, que fica mais clara analisando os seus possíveis critérios destacados abaixo.

**Definição 4** Para cada  $D_I \subseteq D$ ,  $M_J$ ,  $M^+ \subseteq M$ , temos que uma cobertura  $D_I$  de  $M_J$  é:

- **Mínima** se a sua cardinalidade é a menor entre todas as coberturas de  $M_J$ .
- **Irredundante** se nenhum de seus subconjuntos próprios é também uma cobertura de  $M_J$ ; caso contrário ela é **redundante**.
- **Relevante** se é um subconjunto de  $causa(M^+)$ ; caso contrário ela é **irrelevante**.

**Definição 5** Um conjunto  $E \subseteq D$  é dito ser uma **explicação** de  $M^+$  sse E é cobertura de  $M^+$  e satisfaz um dado critério de parcimônia. Normalmente, usamos o critério de parcimônia **irredundante** como sugerido em [6].

**Definição 6** A solução de um problema de diagnóstico  $P = (BC, K)$ , designado  $Sol(P)$ , é o conjunto de todas as explicações para K.

### **2.3. Limitações da TCP**

Para o de diagnóstico médico de doenças epidemiológicas, a TCP possui algumas limitações que necessitam ser tratadas para tornar suas explicações mais consistentes.

Quando um paciente consulta um especialista ele normalmente descreve os seus sintomas. Neste momento a correta interpretação pelo profissional da saúde é muito importante para retratar a realidade à partir de informações, normalmente, imprecisas ou incorretas. Isto ocorre, pois os pacientes tem a tendência de confundir sintomas similares ou descendentes. Por exemplo, quando alguém diz que está com tosse este pode estar, na verdade, com tosse seca ou outro tipo de tosse. Este fato por si só poderia levantar hipóteses diferentes das reais causas. Neste sentido a TCP não fornece nenhum mecanismo de auxílio para lidar com informações imprecisas.

Ao contrário de diagnóstico de outras áreas, é raro ocorrerem co-infecção de duas ou mais doenças epidemiológicas simultaneamente, com exceção do caso específico de uma das doenças ser a AIDS. Apesar deste fato, a TCP gera explicações com mais de uma doença em quase todos os casos, aumentando significativamente o número de hipóteses possíveis.

Existem sintomas que se manifestam em quase todas as infecções de determinadas doenças, se não em todas. Estes se tornam uma informação valiosa no existência de dúvidas sobre as hipóteses. Sem informações estatísticas a TCP não faz qualquer distinção entre as doenças tratando-as da mesma maneira.

Existem fatores relevantes para o diagnóstico que não são sintomas, como, se um paciente tomou uma determinada vacina ou se o mesmo frequentou uma área onde encontra-se o foco de proliferação de algumas doenças, entre outros. Apesar destes fatores influenciarem de forma significativa na geração de um diagnóstico os mesmos não possuem qualquer tipo de suporte desta teoria. Podemos observar também que, em alguns casos, de certas doenças, existem conjuntos de informação, compostas por sintomas e fatores relevantes, que são suficientes para levantar a suspeita de uma doença, independentemente de qualquer outra informação disponível.

## **3. Sistema proposto**

Esta seção está divida em duas partes. A primeira trata da formulação do problema de diagnóstico. A segunda trata de como solucionamos o problema e geramos o resultado.

### **3.1. Adições e modificações de funcionalidades na TCP**

Com base nas limitações da TCP, criamos uma versão modificada desta teoria adicionando algumas funcionalidades ao sistema de diagnóstico.

Sabemos que um problema bastante comum em diagnósticos médicos é a imprecisão das informações fornecidas por pacientes. Os sintomas descritos são, geralmente, confundidos com outros similares, normalmente mais genéricos, e a intensidade de algumas manifestações também podem ser mal interpretadas. No exemplo da seção anterior tosse seca é uma especialização de tosse e cada um destes pode ser manifestado em doenças distintas. Por sua vez, quando analisamos apenas doenças que podem causar tosse existe a chance de excluirmos as reais causas do diagnóstico. Outro fator que, geralmente, gera confusão é a intensidade de um sintoma como no caso de febre e febre alta. A

intensidade e a especialização dos sintomas podem levantar novas hipóteses e excluir outras. Para solucionar esta limitação criamos as **árvores hierárquicas de sintomas**, onde os sintomas mais genéricos são “pais” dos especializados.

Dado um conjunto  $M^+$  de sintomas observados, ao invés de, apenas verificar quais doenças podem causar  $M^+$  como seria feito pela TCP, nós geramos variações de  $M^+$ , onde seus sintomas são substituídos por descendentes da árvore hierárquica de sintomas a qual pertencem.

Seja  $a_1$  uma árvore hierárquica de sintomas, onde os filhos de sua raiz  $m_j$  sejam  $m_j = \{m_k, m_l, m_o\}$ ;  $m_k = \{m_m, m_n\}$ ;  $m_l, m_o = \{\}$ . Suponha que  $M^+ = \{m_o, m_k, m_w\}$ . A TCP iria apenas utilizar estes três sintomas para gerar o diagnóstico. No modelo proposto nós iríamos não apenas utilizar este conjunto de sintomas, mas todos os seus descendentes, ou seja, verificaríamos quais doenças podem causar, simultaneamente, os sintomas  $\{m_o, m_k, m_w\}$  or  $\{m_o, m_m, m_w\}$  or  $\{m_o, m_n, m_w\}$ , porque  $m_m$  e  $m_n$  são descendentes de  $m_k$ .

Neste modelo proposto executamos a TCP para cada uma das variações logo após das mesmas terem sido geradas.

A segunda limitação se referencia ao fato de que, raramente ou nunca, doenças epidemiológicas co-infectam uma pessoa simultaneamente e, portanto, as explicações que possuem mais de uma doença deveriam ser excluídas do diagnóstico. Para lidar com este problema nós criamos e adotamos um novo critério de parcimônia chamado de **singular**.

**Definição 7** Para cada  $D_I \subseteq D$  e  $M_J \subseteq M$  em um problema de diagnóstico  $P$ , temos: Uma cobertura  $D_I$  de  $M_J$ , é dita ser **singular** se  $|D_I| = 1$ , ou seja, possui apenas um elemento.

Quando adotamos este critério, estamos lidando diretamente com a limitação descrita acima e indiretamente reduzindo o número de explicações possíveis. Note que, o critério de parcimônia singular também é mínimo, irredundante e relevante por sua natureza.

Outro fator importante ignorado pela TCP é o fato de alguns sintomas serem bastante comuns, ou quase obrigatórios, na presença de uma doença em particular. Para utilizar estas informações, as adicionamos na relação C, entre doenças e sintomas, e, portanto, para qualquer sintoma que uma doença possa causar agregamos a informação se este é **quase obrigatório** ou não em relação a mesma.

A TCP também não fornece suporte para informações relevantes que não são sintomas. Por exemplo, a informação de que um paciente tomou uma vacina contra determinada doença é tão importante quanto saber quais sintomas o mesmo está manifestando, no caso desta vacina servir contra as suspeitas levantadas. Para lidar com estes casos criamos estruturas chamadas de **fatores relevantes** que são compostas por **perguntas e respostas esperadas**. As respostas esperadas serão comparadas contra as respostas fornecidas pelos pacientes, auxiliando na eliminação de suspeitas ou até mesmo no fortalecimento de outras.

Como descrevemos, existem conjuntos de informações que levantam a suspeita da presença de algumas doenças sem mesmo termos que considerar quaisquer outras informações. Estes conjuntos serão chamados de **conjunto de informações determinantes de suspeita (CONINDESU)**. Cada doença pode possuir diversos CONINDESUs.

Um CONINDESU é formado por um **elemento E**, um **conjunto de elementos OU** e um **número mínimo de elementos OU**. Cada elemento, sendo ele um elemento E ou elemento OU é composto por sintomas e/ou fatores relevantes. A **ativação** de um CONINDESU é a confirmação que as informações necessárias foram satisfeitas e o paciente provavelmente possui a doença referente ao CONINDESU ativado. Sempre que dizemos que um CONINDESU está ativado significa que a sua respectiva doença é uma suspeita. Para ativarmos um CONINDESU, todos os sintomas e/ou fatores relevantes do elemento E devem, necessariamente, estar presentes e a quantidade de elementos OU presentes deve ser maior ou igual ao número mínimo de elementos OU.

### 3.1.1. Definição formal

**Definição 8** Um problema de diagnóstico  $P$  é representado pelo par  $(BC, K)$  onde:

- $BC = (D, M, C, A, F, DF, Co, DCo)$  : Base de conhecimento, composta por:
  - $D = \{d_1, d_2, \dots, d_m\}$ ,  $m \geq 1$  : é um conjunto finito, não-vazio, de objetos chamados *doenças*.
  - $M = \{m_1, m_2, \dots, m_n\}$   $n \geq 1$  : é um conjunto finito, não-vazio, de objetos chamados *sintomas*.
  - $C \subseteq D \times S \times \{\text{Sim}, \text{Não}\}$  : é uma relação que representa as associações  $(d_i, m_j, \{\text{Sim}, \text{Não}\})$ , onde  $d_i$  pode causar  $m_j$  e  $m_j$  é um sintoma quase obrigatório (Sim) ou não (Não) com relação a  $d_i$ .
  - $A = \{a_1, a_2, \dots, a_o\}$  : é um conjunto finito de árvores hierárquicas de sintomas.
    - \*  $a_z = (r_z)$ ,  $r_z \in M$  :  $r_z$  é a raiz da árvore  $a_z \in A$ .
    - \*  $\text{Filhos}(m, a_z) = \{m_1, m_2, \dots, m_p\}$ ;  $m, m_i \in M$ ;  $a_z \in A$  : é um conjunto finito que contém todos os filhos diretos de  $m$  na árvore  $a_z$ .
  - $F = \{f_1, f_2, \dots, f_q\}$  : é um conjunto finito de fatores relevantes.
    - \*  $f_i = (p_i, re_i)$ ;  $re_i = \{\text{Sim}, \text{Não}\}$  :  $p_i$  é a pergunta e  $re_i$  é a resposta esperada para o fator relevante  $f_i$ .
  - $DF \subseteq D \times F$  : é a relação, de  $D$  para  $F$ , onde  $(d_i, f_j) \in DF$  se  $d_i$  pode ser afetado por  $f_j$ .
  - $Co = \{co_1, co_2, \dots, co_r\}$  : é um conjunto finito de CONINDESUs.
    - \*  $co_i = (ee_i, EOU_i, n_i)$  :  $ee_i$  é o elemento E,  $n_i$  é o número mínimo de elementos OU e  $EOU_i$  é o conjunto de elementos OU de  $co_i$ .
    - \*  $EOU_i = \{eou_1, eou_2, \dots, eou_s\}$  : é um conjunto finito de elementos OU de  $co_i$ .
      - \*  $e = (M_e, F_e)$ ,  $M_e \subseteq M$ ,  $F_e \subseteq F$ ,  $e \in (ee \cup EOU)$ .
  - $DCo \subseteq D \times Co$  : é uma relação, de  $D$  para  $Co$ , onde  $(d_i, co_j) \in DCo$  quando a ativação de  $co_j$  levanta a suspeita de  $d_i$ .
- $K = (Mp^+, Mnp, Fp)$  : caso composto por:
  - $Mp^+ \subseteq M$  : é um conjunto finito, não-vazio, de sintomas observados no paciente.
  - $Mnp \subseteq M$  : é um conjunto de sintomas que, definitivamente, não estão presentes no paciente.
  - $Fp \subseteq F$  : é um conjunto finito de fatores relevantes respondidos pelo paciente.

### 3.2. Geração do diagnóstico

**Definição 9** Dizemos que um elemento  $e$  de um particular CONINDESU referente a uma doença  $d_i$  está **presente** quando  $e = (M_e, F_e)$  e  $M_e \subseteq Mp^+$  e para cada  $fe_j \in F_e$ ,  $(d_i, fe_j) \in DF$ ,  $F_e \subseteq Fp$ .

**Definição 10** Dizemos que um CONINSESU  $co_i = (ee_i, EOU_i, n_i) \in Co$  está **ativado** quando  $ee_i$  está **presente** e  $|SEOU_i| \geq n_i$ ,  $eou_j \in SEOU_i$  e  $eou_j$  está **presente** e  $SEOU_i \subseteq EOU_i$ .

**Definição 11** Dizemos que uma cobertura  $D_I \subseteq D$  é **inconsistente** sobre  $K = (Mp^+, Mnp, Fp)$  quando  $D_I$  é cobertura de  $Mp^+$  e  $\exists m_j \in Mnp$  e  $m_j \in efeito(D_I)$ . Caso contrário ela é **consistente**.

**Definição 12** Dizemos que uma doença  $d_i$  possui um **fator de impacto negativo** quando  $f_j = (p_j, re_j) \in Fp$  e  $(d_i, (p_j, re_k)) \in DF$  e  $re_j \neq re_k$ .

**Definição 13** Um elemento  $h \in D$  é chamado de **hipótese** de um caso particular  $K$  sse, com um critério de parcimônia singular,  $h$  é uma cobertura consistente de  $K$  e não possui qualquer fator de impacto negativo ou  $(h, co_i) \in DCo$  e  $co_i$  está ativado.

**Definição 14** Uma solução parcial de um problema de diagnóstico  $P = (BC, K)$ , designada  $SolPar(P)$ , é o conjunto de todas as hipóteses do caso  $K$ .

**Definição 15** Para qualquer  $m_i \in M$ ,  $a_z, a_x \in A$ ,  $K = (Ps^+, Mnp, Fp)$  em um problema de diagnóstico  $P$ , temos:

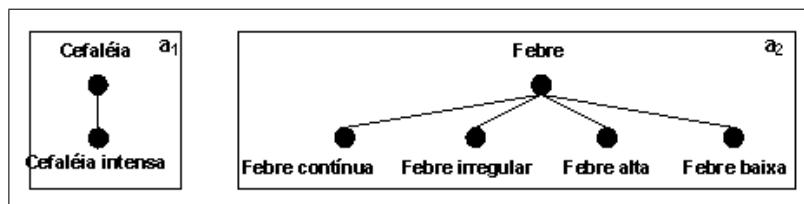
- $descendente(m_i, a_z) = \bigcup_{m_j \in Filhos(m_i, a_z)} descendente(m_j, a_z) \cup m_i$  : é um conjunto finito de todos os descendentes de  $m_i$  em  $a_z$ , incluindo  $m_i$ .
- $var(Mp^+) = descendente(mp_1, a_z) \times descendente(mp_2, a_x) \times \dots \times descendente(mp_n, a_y)$ ;  $mp_i \in Mp^+$ ;  $|Mp^+| = n$ : é o conjunto finito de todas as variações de  $Mp^+$ , considerando os descendentes de todos os seus elementos.
- $var(K) = \{k_1, k_2, \dots, k_u\}$ ;  $k_i = (Mp_i^+, Mnp, Fp)$ ;  $Mp_i^+ \in var(Mp^+)$ ;  $K = (Mp^+, Mnp, Fp)$  : é um conjunto finito de todas as variações possíveis do caso  $K$ , considerando as árvores hierárquicas de sintomas presentes em  $A$ .

**Definição 16** Finalmente, temos que o **diagnóstico final** do problema  $P$ , designado como  $Diag(P)$ , é o conjunto de todas as hipóteses das variações de  $K$ , ou seja,  $Diag(P) = \bigcup_{k_i \in var(K)} SolPar(BC, k_i)$ .

### 4. Demonstração do sistema

Nesta seção mostraremos um exemplo de execução do sistema construído utilizando o modelo proposto. Ilustraremos a sua interface e utilização de todas as alterações realizadas sobre a TCP.

Em um primeiro momento devemos escolher todos os sintomas presentes no paciente, através de uma lista que contém todos os sintomas do sistema. Logo após, é necessário solicitar a geração de um diagnóstico inicial pelo sistema e verificar quais doenças podem causar os sintomas informados, ou seus descendentes, simultaneamente.



**Figura 1. Árvores hierárquicas de sintomas presentes na base de conhecimento**

Através da Figura 1 podemos observar a disposição atual das árvores hierárquicas de sintomas na base de conhecimento do sistema.

Dado que os sintomas manifestados no paciente são “coriza”, “cefaléia” e “febre” nós temos as seguintes variações dos sintomas observados, considerando os seus descendentes: {“coriza”, “cefaléia”, “febre”}, {“coriza”, “cefaléia intensa”, “febre”}, {“coriza”, “cefaléia”, “febre baixa”}, {“coriza”, “cefaléia intensa”, “febre baixa”}, {“coriza”, “cefaléia”, “febre alta”}, {“coriza”, “cefaléia intensa”, “febre alta”}, {“coriza”, “cefaléia”, “febre irregular”}, {“coriza”, “cefaléia intensa”, “febre irregular”}, {“coriza”, “cefaléia”, “febre contínua”}, {“coriza”, “cefaléia intensa”, “febre contínua”}.

Usando a TCP, com o critério de parcimônia singular, o sistema seleciona apenas as doenças que podem causar, simultaneamente, uma das variações dos sintomas apresentados. O resultado é exibido em uma tela como detalhada na Figura 2.

**Figura 2. Perguntas do sistema**

No nosso exemplo, somente a doença “gripe” pode causar, simultaneamente, os sintomas {“coriza”, “cefaléia”, “febre”}, mas a “rubéola” pode causar os sintomas {“coriza”, “cefaléia”, “febre alta”} e, portanto, está presente como hipótese nesta fase do diagnóstico também.

Como podemos observar na Figura 2, esta pré-seleção retorna, através de perguntas, todos os fatores relevantes e também os sintomas quase obrigatórios de cada hipótese. Depois de respondidas as perguntas devemos clicar no botão “Refinar diagnóstico” para gerar o diagnóstico final. Fazendo isto, as seguintes ações são tomadas:

1. Verificamos a existência de qualquer CONINDESU ativado em todas as doenças do sistema. Se em alguma delas houver a ativação de qualquer um de seus CONINDESUs, devemos adicioná-la como hipótese no diagnóstico final.
2. Usamos a TCP para gerar novos conjuntos de hipóteses. A diferença é que nesta etapa devemos incluir os novos sintomas que foram adicionados, e também os

que, definitivamente, não estão presentes. Estes virão da verificação de presença dos sintomas quase obrigatórios.

3. Finalmente, removemos do diagnóstico final as hipóteses que foram geradas pela TCP, e não pela ativação de seus CONINDESUs, mas possuem fatores de impacto negativo.

O resultado é exibido na Figura 3. Esta nos mostra que as únicas hipóteses remanescentes para o caso K, com  $K = (\{\text{"coriza"}, \text{"cefaléia"}, \text{"febre alta"}, \text{"tosse"}\}, \{\text{"dor de garganta"}, \text{"mialgia"}\}, \{(\text{"O paciente está com doença aguda de até 5 dias?"}, \text{"Sim"}), (\text{"O paciente tomou a vacina contra a gripe no último ano?"}, \text{"Sim"})\})$ , é a “gripe”. Isto ocorre porque a ex-hipótese, “rubéola”, pode causar o sintoma “mialgia”, mas este, definitivamente, não está presente. “Gripe” também pode causar este sintoma e, portanto, possui um fator de impacto negativo, (“O paciente tomou vacina contra gripe no último ano?”, “Sim”), mas ele contém um CONINDESU ativado,  $co = (ee, \{eou_1, eou_2\}, 1)$ , onde  $ee = (\{\text{"febre"}\}, \{(\text{"O paciente está com doença aguda de até 5 dias?"}, \text{"Sim"})\})$ ,  $eou_1 = (\{\text{"tosse"}\}, \{\})$ ,  $eou_2 = (\{\text{"dor de garganta"}\}, \{\})$ .

Arquivo	Executar	Ajuda
Diagnóstico Final		
Voltar para hipóteses		
<p><b>Sintomas do Paciente</b></p> <p>Cefaléia Coriza Febre Tosse</p> <p>As seguintes doenças fazem parte de um possível diagnóstico: Influenza ou gripe</p> <p>Com a seguinte configuração do sistema:</p> <p>Sintomas presentes no paciente: Cefaléia Coriza Febre Tosse</p> <p>Sintomas não presentes no paciente: Dor de garganta Mialgia</p> <p>Fatores que influenciaram o diagnóstico: O paciente está com doença aguda de duração de até 5 dias? - Sim O paciente tomou a vacina contra a gripe no último ano? - Sim</p>		

**Figura 3. Diagnóstico final**

No exemplo acima, mostramos a execução do sistema enfatizando as principais diferenciações entre o modelo proposto e a TCP que são: critério de parcimônia singular, árvores hierárquicas de sintomas, fatores relevantes e suas contribuições e como os CONINDESUs auxiliam na geração do diagnóstico final.

## 5. Conclusão

Este trabalho se concentrou nos aspectos técnicos de adaptação da TCP para o problema dado. Iremos discutir modelos alternativos para o problema e os planos de avaliação do sistema.

Uma alternativa para o modelo de TCP descrito neste artigo seria um modelo padrão de diagnóstico Bayesiano (por exemplo [1]). Um sistema baseado neste modelo iria computar a probabilidade de uma doença dados os sintomas, como:

$$P(d/M) = \frac{P(M/d)P(d)}{P(M)} \approx P(d) \prod_{m_i \in M} P(m_i/d) \quad (1)$$

onde  $M$  é o conjunto de sintomas,  $d$  uma doença em particular.  $P(d)$  é a prevalência de uma doença, e  $P(m_i/d)$  é a probabilidade da doença  $d$  causar o sintoma  $m_i$ .

Modelos Bayesianos para diagnóstico possuem semântica clara e um forte histórico na pesquisa de sistemas de apoio à decisões [5]. Mas estes dependem criticamente da força das probabilidades causais  $P(m_i/d)$ . Apesar da existência de alguns sistemas para diagnóstico de doenças infecciosas baseados no modelo Bayesiano (por exemplo o Gideon <http://www.gideononline.com/>), os autores, com a ajuda de alguns epidemiologistas brasileiros, não conseguiram obter estas probabilidades em nenhuma literatura. A utilização da TCP pode ser vista como uma simplificação do modelo Bayesiano, onde  $P(m_i/d)$  assume o valor de 0 ou 1, quando  $d$  pode ou não ser causa de  $m_i$ .

O sistema apresentado neste artigo está sob avaliação de um grupo de epidemiologistas. Enquanto escrevemos este artigo, o grupo está tentando validar o conhecimento e eurísticas utilizadas no sistema, ou sob as definições de [2], uma avaliação preditiva. Estas avaliações usam casos de teste de laboratórios onde os resultados são conhecidos, e verificam se as sugestões do sistema são compatíveis com as dos especialistas.

Infelizmente, não possuímos uma forma metodológica de realizar esta avaliação preditiva, e portanto não temos uma visão clara dos seus resultados. A partir do momento que o conhecimento seja validado (por repetição do ciclo de críticas dos especialistas e a re-engenharia da base de conhecimento pela equipe técnica), os usuários em potencial poderão ser envolvidos na avaliação do sistema.

## Referências

- [1] S. A. Berger and U. Blackman. Computer program for diagnosing and teaching geographic medicine. *Journal of Travel Medicine*, pages 199–203, 1995.
- [2] D. Borenstein. Towards a practical method to validate decision support systems. *Decision Support Systems*, 1998.
- [3] Secretaria de vigilância em saúde. *Guia de vigilância epidemiológica*. Ministério da Saúde, 6th edition, 2005.
- [4] P. Wang J. A. Reggia, D. Nau and Y. Peng. A formal model for diagnostic inference. *Information Sciences*, 37, 1985.
- [5] R. A. Miller. Medical diagnostic decision support systems-past, present, and future: a threaded bibliography and brief commentary. *Journal of the American Medical Informatics Association*, pages 1(2):8–27, 1994.
- [6] Y. Peng and J. A. Reggia. *Abductive Inference Models for Diagnostic Problem-Solving*. Springer-Verlag, New York, 1st. edition, 1990.
- [7] J. Wainer and A. M. Rezende. A temporal extension to the parsimonious covering theory. Technical report, Institute of Computing, University of Campinas, Nov 1996.
- [8] J. Wainer and S. Sandri. A fuzzy temporal extension to the parsimonious covering theory. *The Seventh Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 9–26, 1998.