Aplicação de inferência difusa em bioinformática para identificação de SNPs

Wagner Arbex[1], Fabrízzio Condé de Oliveira[2], Luís Alfredo Vidal de Carvalho[3]

[1]Empresa Brasileira de Pesquisa Agropecuária
Rua Eugênio do Nascimento, 610 - 36038-330 - Juiz de Fora - MG

[2]Universidade Salgado de Oliveira
Av. dos Andradas, 731 - 36036-000 - Juiz de Fora - MG

[3]Universidade Federal do Rio de Janeiro
Centro de Tecnologia - Bloco H-319 - 21945-970 - Rio de Janeiro - RJ

arbex@cnpgl.embrapa.br, fabrizzioconde@gmail.com, alfredo@ufrj.br

Abstract. *Research involving the discovery of single nucleotide polymorphisms (SNPs) requires bioinformatics tools to be applied to different cases, with the ability to analyze "reads" from different sources, levels of coverage and to establish reliable measures. These tools work with different methodologies on different attributes, however, it is expected similar results, even when dealing with a same data set, but it's not unusual to provide different results, which leads to uncertainty in decision making, when the results are discordant.*

Resumo. *A investigação de polimorfismos de base única necessita de ferramentas de bioinformática que devem ser aplicadas a diferentes casos, com capacidade para analisar seqüências de diferentes fontes, níveis de cobertura e que consigam medidas confiáveis. Essas ferramentas trabalham com diferentes metodologias, sobre diferentes atributos, contudo, espera-se que apresentem resultados similares, ao tratarem um mesmo conjunto de dados, mas, não é incomum fornecerem resultados diferentes, o que produz incerteza na tomada de decisão, quando os resultados são discordantes.*

1. Introdução

Polimorfismos de base única (*single nucleotide polymorphisms* - SNPs) são pares de bases em uma única posição no DNA genômico, que se apresentam com diferentes alternativas nas seqüências e podem ser encontrados no genoma de indivíduos isoladamente ou em grupos de indivíduos, em alguma população. A individualidade é conseqüência da expressão do código genético, ou seja, em sua essência, as seqüências de nucleotídeos formam as moléculas e seqüências de DNA, RNA e proteínas, que, por sua vez, interagem e formam as células, as quais também, interagem e formam os tecidos, os órgãos, até que, finalmente, formam os indivíduos. Essa é a importância dos SNPs, pois, em síntese, a alteração de um único nucleotídeo, uma única base, em uma dada seqüência, pode alterar a formação de proteínas e o conjunto dessas alterações pode provocar variações nas características dos indivíduos.

Esse texto apresenta um modelo matemático e computacional para tomada de decisão, desenvolvido e implementado com o fuzzyMorphic.pl [Arbex 2009], aplicado à investigação de SNPs em seqüências expressas de cDNA, que utiliza-se de lógica difusa para a implementação de um sistema de inferência, auxiliar à tomada de decisão, partindo de resultados prévios, obtidos por diferentes ferramentas de descoberta de SNPs e que apresentam resultados possivelmente conflitantes. O modelo é aplicado para auxiliar na tomada de decisão, nos casos em que as informações sejam divergentes e, também, na confirmação de informações coincidentes.

2. Inferência difusa como suporte à decisão

A subjetividade no raciocínio em geral, utilizada no cotidiano, sendo transmitida e perfeitamente compreendida entre interlocutores, é expressa em "termos e variáveis lingüísticas" [Zadeh 1973] e não é expressa sob a lógica clássica ou qualquer abordagem matemática tradicional. O uso de, p. ex., adjetivos comuns que representam imprecisão ou incerteza, tais como, *alto*, *baixo* ou, relações e agrupamentos, como, *conjunto das pessoas altas*, não podem ser expressos por essas abordagens, a menos que seja definido, com exatidão, o conceito ou o valor que determine a altura, a partir da qual, uma pessoa pode ser considera alta.

Os termos e variáveis lingüísticas aumentam a complexidade dos sistemas computacionais frente à capacidade trabalharem com números, valores exatos, discretos e, por vezes, excludentes, o que sugere a idéia de que, trabalhar com valores incertos, possibilita a modelagem de sistemas complexos, mesmo que se reduza a precisão do resultado, mas não retira a credibilidade. Se as incertezas, quando consideras isoladamente, são indesejáveis, quando associadas a outras características, em geral, permitem a redução da complexidade do sistema e aumentam a credibilidade dos resultados obtidos [Klir and Yuan 1995].

As abordagens clássicas são falhas para valores limítrofes e, portanto, resultados matemática e logicamente precisos, porém, questionáveis, podem ser encontrados. P. ex., o *Polyphred score* (*PPS*) estabelece seis classes com intervalos precisos (Tab. 1) [Nickerson et al. 2008] e, supondo que fossem determinados os *scores* 70 e 89 para dois pontos, então, para ambos, seria considerada a taxa de 35% de verdadeiros positivos na decisão desses pontos virem a ser SNPs (Classe 4).

Tabela 1. Classes definidas pelo PPS [Nickerson et al. 2008].

Classe	PPS	Taxa de verdadeiros positivos
1	99	97%
2	95 - 98	75%
3	90 - 94	62%
4	70 - 89	35%
5	50 - 69	11%
6	0 - 49	1%

Essa decisão, lógica e matematicamente precisa, pode ser questionada devido à subjetividade que a envolve, visto que, 70 e 89, se encontram nos limites da classe a qual pertencem e, portanto, muito próximos de diferentes interpretações. Todavia, as abordagens clássicas da lógica e da matemática não possuem as ferramentas necessárias para tratar valores limítrofes, imprecisão ou incerteza. Um valor limítrofe acarretará dúvidas na "decisão" de o ponto ser, ou não, considerado polimórfico, o que sugere um sistema de inferência difusa (SID) para o tratamento dessa incerteza.

O problema de valores limítrofes, em geral, não é tão simples quanto parece, do contrário, as abordagens clássicas poderiam facilmente resolvê-lo, mas, ao aproximar-se do raciocínio subjetivo para a interpretação e a extração de uma resposta, uma decisão, torna-se complexo e a aparente simplicidade é conferida pela modelagem por lógica difusa e seu embasamento na teoria dos conjuntos difusos. A subjetividade intrínseca ao raciocínio trata situações complexas, mediante imprecisão, incerteza ou aproximação e, então, são utilizados "operadores humanos", também de natureza imprecisa, que são expressos por termos ou variáveis lingüísticas, o que, em geral, não permite uma solução em termos exatos, mas, pode propor uma classificação, agrupamento ou agregação qualitativa em categorias ou possíveis conjuntos de soluções [de Almeida and Evsukoff 2005].

A teoria dos conjuntos difusos e a lógica difusa são adequadas para representar a informação imprecisa e caso seja possível organizar os operadores humanos em regras da forma *se ANTECEDENTE então CONSEQÜENTE*, o raciocínio subjetivo pode ser descrito em um algoritmo computacionalmente executável [Tanscheit 2007] capaz de classificar, de modo impreciso, as variáveis que participam dos termos antecedentes e conseqüentes das regras, em conceitos qualitativos, o que representa a idéia de variável lingüística [de Almeida and Evsukoff 2005]. Assim, como sistemas capazes de processar informações imprecisas e qualitativas, os modelos de inferência difusa são adequados à situação de tomada de decisão [de Almeida and Evsukoff 2005].

3. Descrição do modelo e do SID para identificação de SNPs

Em geral, as etapas de um SID são: a fuzzificação, que converte os dados "precisos" (*crisps*) de entrada em valores difusos; a inferência, propriamente dita; e a defuzzificação, que converte os resultados difusos em grandezas numéricas precisas. No modelo proposto, consideram-se como valores de entrada, as probabilidades, previamente determinadas, de o ponto vir a ser um SNP e o valor de qualidade do ponto na seqüencia consenso. Os *Casos 1* e *2* serão utilizados ao longo do texto para demonstrar o modelo, assumindo, para o *Caso 1*, 99% e 96%, quanto as probabilidades e 43 de qualidade e, para o *Caso 2*, os valores são, respectivamente, 94%, zero e 50.

3.1. Fuzzificação

Avalia-se um valor de entrada por sua "função de pertinência", o que determina um "grau de pertinência" (GP) do valor para a sua função e as funções de pertinência adotadas foram baseadas:

1. no *PPS* (Tab. 1), com a função de pertinência definida pela variável linguística *probabilidade*, com os termos (Exps. 1 e 2): *improvável* (P_{IM}), *pouco provável* (P_{PP}), *medianamente provável* (P_{mP}), *provável* (P_{PR}), *muito provável* (P_{MP}) e *altamente provável* (P_{AP});
2. na qualidade das bases do consenso – o *Phrap quality score* (PQS) – que varia entre 4 e 90 e sua função de pertinência (Exps. 3) define a variável linguística *qualidade*, nos termos: *ruim* (Q_R), *boa* (Q_B) e *ótima* (Q_O).

$$P_{IM}(x) = \begin{cases} 1 & x \leq 49 \\ \frac{59-x}{59-49} & 49 < x < 59 \\ 0 & x \geq 59 \end{cases} \quad P_{PP}(x) = \begin{cases} 0 & x \leq 25 \\ \frac{x-25}{50-25} & 25 < x < 50 \\ 1 & 50 \leq x \leq 69 \\ \frac{79-x}{79-69} & 69 < x < 79 \\ 0 & x \geq 79 \end{cases} \quad P_{mP}(x) = \begin{cases} 0 & x \leq 60 \\ \frac{x-60}{70-60} & 60 < x < 70 \\ 1 & 70 \leq x \leq 89 \\ \frac{91,5-x}{91,5-89} & 89 < x < 91,5 \\ 0 & x \geq 91,5 \end{cases} \quad (1)$$

$$P_{PR}(x) = \begin{cases} 0 & x \leq 80 \\ \frac{x-80}{90-80} & 80 < x < 90 \\ 1 & 90 \leq x \leq 94 \\ \frac{96-x}{96-94} & 94 < x < 96 \\ 0 & x \geq 96 \end{cases} \quad P_{MP}(x) = \begin{cases} 0 & x \leq 92,5 \\ \frac{x-92,5}{95-92,5} & 92,5 < x < 95 \\ 1 & 95 \leq x \leq 98 \\ \frac{99-x}{99-98} & 98 < x < 99 \\ 0 & x \geq 99 \end{cases} \quad P_{AP}(x) = \begin{cases} 0 & x \leq 96,5 \\ \frac{x-96,5}{99-96,5} & 96,5 < x < 99 \\ 1 & x \geq 99 \end{cases} \tag{2}$$

$$Q_R(x) = \begin{cases} 1 & x \leq 20 \\ \frac{30-x}{30-20} & 20 < x < 30 \\ 0 & x \geq 30 \end{cases} \quad Q_B(x) = \begin{cases} 0 & x \leq 15 \\ \frac{x-15}{30-15} & 15 < x < 30 \\ 1 & 30 \leq x \leq 40 \\ \frac{70-x}{70-40} & 40 < x < 70 \\ 0 & x \geq 70 \end{cases} \quad Q_O(x) = \begin{cases} 0 & x \leq 40 \\ \frac{x-40}{50-40} & 40 < x < 50 \\ 1 & x \geq 50 \end{cases} \tag{3}$$

Os resultados da fuzzificação para o *Caso 1*, $PPS_1 = 99$, $PPS_2 = 96$ e $PQS = 43$, em suas respectivas funções de pertinência, podem ser vistos nas Tabs. 2 e 3 e as Figs. 1 e 2 representam graficamente seus conjuntos difusos.

Tabela 2. GPs para a variável probabilidade, para o Caso 1.

	PPS_1	PPS_2
Improvável	0	0
Pouco provável	0	0
Medianamente provável	0	0
Provável	0	0
Muito Provável	0	1
Altamente provável	1	0

Tabela 3. GPs para a variável qualidade, para a Caso 1.

	PQS
Ruim	0
Boa	0,9
Ótimo	0,3

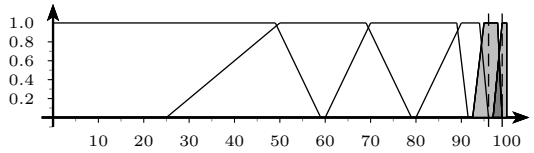

Figura 1. Fuzzificação para a variável probabilidade, no Caso 1.

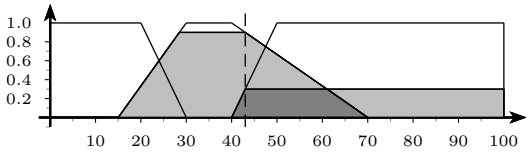

Figura 2. Fuzzificação para a variável qualidade, no Caso 1.

Para o *Caso 2*, o resultado da fuzzificação para $PPS_1 = 94$, $PPS_2 = 0$ e $PQS = 50$, pode ser visto nas Tabs. 4 e 5 com as representações nas Figs. 3 e 4.

Tabela 4. GPs para a variável probabilidade, para o Caso 2.

	PPS_1	PPS_2
Improvável	0	1
Pouco provável	0	0
Medianamente provável	0	0
Provável	1	0
Muito Provável	0,6	0
Altamente provável	0	0

Tabela 5. GPs para a variável qualidade, para o Caso 2.

	PQS
Ruim	0
Boa	0,67
Ótimo	1

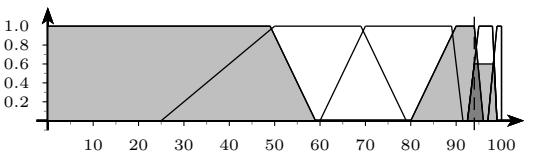

Figura 3. Fuzzificação para a variável probabilidade, no Caso 2.

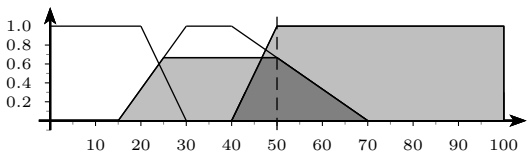

Figura 4. Fuzzificação para a variável qualidade, no Caso 2.

A *probabilidade*, para o *Caso 1*, é expressa pelos termos *muito provável* e *altamente provável*, e a *qualidade*, pelos termos *bom* e *ótimo* e, essas mesmas variáveis do *Caso 2*, pelos termos *improvável*, *provável*, *muito provável*, *bom* e *ótimo*.

3.2. Inferência

A inferência executa operações sobre os conjuntos difusos, com a combinação dos antecedentes das regras, a implicação e a aplicação do *modus ponens* generalizado, sendo, esse procedimento, feito em dois passos: a "agregação", que corresponde ao operador lógico E que executa a intersecção entre conjuntos e, portanto, determina o mínimo entre os valores disparados pelas regras, seguido da "composição".

Os modelos ("máquinas") de inferência adequados para esse SID, são os modelos de Mamdani ou de Larsen, visto que são sensíveis ao disparo de múltiplas regras sobre o conjunto de saída, quando, então, inicia-se o procedimento de defuzzificação, que começa com o segundo passo da inferência, a "composição", que é equivalente ao operador lógico OU e executa a união entre conjuntos, na qual o maior valor entre os mínimos resultantes da agregação é considerado para a defuzzificação.

Foram estabelecidas trinta e seis regras de inferência (Tab. 6), sendo que em metade dessas seus antecedentes são avaliados pelas variáveis *probabilidade* (PPS_1) e *qualidade* e, a outra metade, é avaliada pelas variáveis *probabilidade* (PPS_2) e *qualidade*. Essas regras, relacionam termos de entrada com a função de saída, expressa pelos termos *SNP descartado*, *SNP não confirmado* e *SNP confirmado*.

Tabela 6. Regras de inferência utilizadas no SID.

se	improvável	*e*	qualidade ruim	*então*	SNP descartado	(R_1)
	pouco provável		qualidade ruim		SNP descartado	(R_2)
	medianamente provável		qualidade ruim		SNP descartado	(R_3)
	provável		qualidade ruim		SNP descartado	(R_4)
	muito provável		qualidade ruim		SNP descartado	(R_5)
	altamente provável		qualidade ruim		SNP descartado	(R_6)
	improvável		qualidade boa		SNP descartado	(R_7)
	pouco provável		qualidade boa		SNP descartado	(R_8)
	medianamente provável		qualidade boa		SNP não confirmado	(R_9)
	provável		qualidade boa		SNP não confirmado	(R_{10})
	muito provável		qualidade boa		SNP confirmado	(R_{11})
	altamente provável		qualidade boa		SNP confirmado	(R_{12})
	improvável		qualidade ótima		SNP descartado	(R_{13})
	pouco provável		qualidade ótima		SNP descartado	(R_{14})
	medianamente provável		qualidade ótima		SNP não confirmado	(R_{15})
	provável		qualidade ótima		SNP não confirmado	(R_{16})
	muito provável		qualidade ótima		SNP confirmado	(R_{17})
	altamente provável		qualidade ótima		SNP confirmado	(R_{18})

No *Caso 1*, as funções de pertinência (Exps. 1, 2 e 3), resultam em $P_{MP} = 1$, para PPS_2, $P_{AP} = 1$, para PPS_1, $Q_B = 0,9$ e $Q_O = 0,3$ (Tabs. 2 e 3 e Figs. 1 e 2), então, a agregação é feita entre Q_B e Q_O, o que resulta no termo *ótima* para a variável *qualidade*. Os demais valores obtidos são iguais e, assim, não aplica-se a agregação, o que resulta em *muito provável* (PPS_2) e *altamente provável* (PPS_1), para *probabilidade*, que disparam as regras R_{17} e R_{18}.

Para o *Caso 2*, após a agregação, toma-se $P_{IM} = 1$ (PPS_2), $P_{MP} = 0,6$ (PPS_1) e $Q_B = 0,67$ que são levados à máquina de inferência, que dispara R_7 e R_{11}.

O modelo de inferência mapeia os antecedentes, resultantes da agregação, no termo conseqüente, que, para os modelos de Mamdani ou Larsen, representa uma função de saída em termos lingüísticos, exatamente como uma função de pertinência.

A função de saída que foi estabelecida, reduz as seis classes definidas para o *PPS* aos termos *SNP descartado* (SNP_D), *SNP não confirmado* (SNP_{NC}) e *SNP confirmado* (SNP_C), que, então, compõem a variável linguística *SNP* (Exps. 4):

$$SNP_D(x) = \begin{cases} 1 & x \leq 20 \\ \frac{30-x}{30-20} & 20 < x < 30 \\ 0 & x \geq 30 \end{cases} \quad SNP_{NC}(x) = \begin{cases} 0 & x \leq 15 \\ \frac{x-15}{30-15} & 15 < x < 30 \\ 1 & 30 \leq x \leq 40 \\ \frac{70-x}{70-40} & 40 < x < 70 \\ 0 & x \geq 70 \end{cases} \quad SNP_C(x) = \begin{cases} 0 & x \leq 40 \\ \frac{x-40}{50-40} & 40 < x < 50 \\ 1 & x \geq 50 \end{cases} \tag{4}$$

As regras R_{17} e R_{18}, disparadas no *Caso 1*, são processadas como:

1. R_{17} tem como antecedentes o valor *muito provável*, com $GP = 1$, e o valor *ótima*, com $GP = 0,3$; assim, a aplicação da regra mapeia o conseqüente *SNP confirmado*, com $GP = 1$ e $GP = 0,3$, isto é $SNP_C = 1$ e $SNP_C = 0,3$;
2. R_{18} tem como antecedentes o valor *altamente provável*, com $GP = 1$, e o valor *ótima*, com $GP = 0,3$; então, da mesma forma, mapeia o conseqüente *SNP confirmado*, com $GP = 1$ e $GP = 0,3$, isto é $SNP_C = 1$ e $SNP_C = 0,3$.

Com a aplicação das duas regras, cujos resultados foram coincidentes, apenas o termo *SNP confirmado* foi mapeado e o procedimento de composição deve ser tomado somente sobre esse termo. A composição busca o máximo entre os GPs de cada termo, no caso, somente sobre o termo *SNP confirmado*, fazendo $SNP_C = 1$.

Para o *Caso 2*, são disparadas as regras R_7 e R_{11}, que avaliam os valores antecedentes $P_{IM} = 1$ e $Q_O = 0,67$, para R_7, e $P_{MP} = 0,6$ e $Q_O = 0,67$, para R_{11}. A regra R_7 mapeia na função de saída o valor *SNP descartado*, com $GP = 1$ e $GP = 0,67$, enquanto a regra R_{11} mapeia na função de saída o valor *SNP não confirmado*, com $GP = 0,67$ e $GP = 0,6$. O termo *SNP confirmado* não foi mapeado, logo o procedimento de composição aplicado aos demais termos resulta em *SNP descartado*, com $GP = 1$ ($SNP_D = 1$), e *SNP não confirmado*, com $GP = 0,67$ ($SNP_{NC} = 0,67$).

As Figs. 5 e 6 representam, respectivamente, a aplicação das regras de inferência sobre a função de saída (Exps. 4) para os *Casos 1* e *2*.

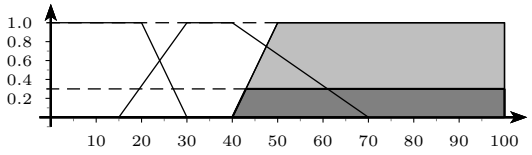

Figura 5. Aplicação das regras de inferência para o Caso 1.

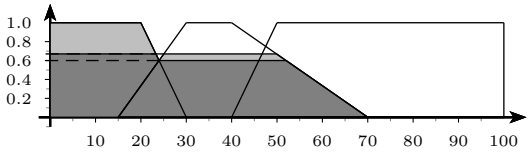

Figura 6. Aplicação das regras de inferência para o Caso 2.

3.3. Defuzzificação

A defuzzificação executa a composição, que determina os valores que representam cada um dos conjuntos mapeados na função de saída, e, a partir desses, calcula um valor preciso (VP), obtido com a aplicação do método de defuzzificação.

Para o modelo proposto, o método de defuzzificação deve considerar múltiplos disparos, pois o valor da qualidade da base no consenso é utilizada como um "valorizador" dos valores de probabilidade confrontados (PPS_1 e PPS_2). Assim, havendo

disparos múltiplos, esses devem ser avaliados, pois, servem à idéia de valorizar os conjuntos difusos estabelecidos na função de saída. Para esse fim, deve ser utilizado o método centro de máximo (*center of maximum* - COM) e, a partir dos modelos de inferência, aplica-se o método de defuzzificação adequado ao problema. Como o fuzzyMorphic.pl permite a inferência pelos modelos de Mamdani e Larsen, ambos podem ser aplicados e, juntamente com os valores tomados da composição, definem os valores para a defuzzificação.

O COM (Exp. 5), trata-se de uma média ponderada, onde o numerador é o somatório dos valores da composição (h_i), isto é, a altura do conjuntos de saída, multiplicados pelos valores no universo de discurso (u_i), encontrados pelo modelo de inferência, do seu respectivo conjunto de saída, e o denominador é o somatório das alturas (h_i).

Para o *Caso 1*, o VP (Exp. 6) e sua representação (Fig. 7) são iguais para os modelos de Mamdani e Larsen, mas, para o *Caso 2*, como conseqüência desses modelos, a defuzzificação apresenta diferentes resultados (Exps. 7 e 8 e Figs. 8 e 9).

$$VP = \frac{\sum h_i \cdot u_i}{\sum h_i} \qquad (5)$$

$$VP_{C_1} = \frac{75 \cdot 1}{1} = 75 \qquad (6)$$

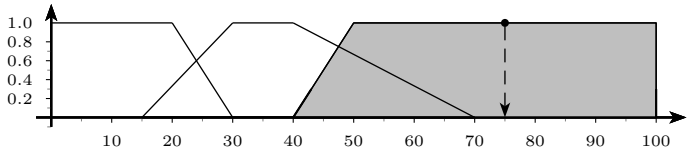

Figura 7. Aplicação do modelo de inferência, para o Caso 1.

$$VP_{C_2} = \frac{(10 \cdot 1) + (37,475 \cdot 0,67)}{1 + 0,67} = 21,02 \qquad (7)$$

$$VP_{C_2} = \frac{(10 \cdot 1) + (35 \cdot 0,67)}{1 + 0,67} = 20,03 \qquad (8)$$

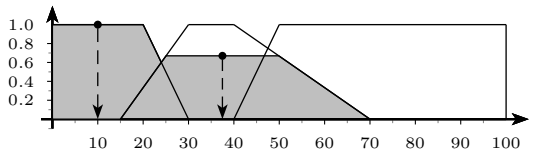

Figura 8. Aplicação do modelo de Mamdani para o Caso 2.

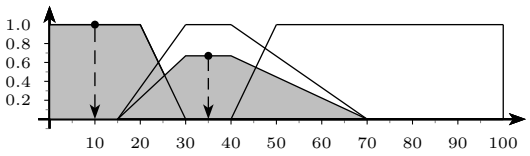

Figura 9. Aplicação do modelo de Larsen para o Caso 2.

3.4. Discussão sobre o modelo de inferência para identificação de SNPs

O *Caso 1* inicia com resultados prévios similares, 99% e 96% de probabilidades do ponto vir a ser um SNP, entretanto, o *Caso 2*, parte de resultados divergentes, 94% e zero. O SID incluiu um novo atributo, a qualidade da base no consenso, 43 e 50, para os *Casos 1* e *2*, respectivamente, ampliando as possibilidades de investigação e utilizando-se deste como um "valorizador" para a tomada da decisão. Assim, aos resultados prévios de o ponto vir a ser um SNP, acrescenta-se a qualidade do ponto, utilizando-os como as variáveis do modelo que permite a determinação de uma das três possibilidade excludentes: a confirmação do SNP, a eliminação dessa possibilidade ou, uma situação intermediária, sem elementos conclusivos para a confirmação ou o descarte dessa possibilidade.

A análise desses casos fornece elementos suficientes para apresentar o modelo, contudo, resultados efetivos são obtidos mediante a análise de conjuntos de dados, quando os valores inferidos a partir do modelo, podem, então, ser agrupados, determinando os conjuntos de pontos que melhor se ajustam às possibilidades investigadas. Estabelecer grupos é uma tarefa complexa, pois procura-se dizer como são e em quantas classes os dados se distribuem, sem o conhecimento a priori dos mesmos e, caso os valores se distribuam equitativamente no espaço, não caracterizando qualquer categoria, as classes podem não existir, uma vez que são definidas com base na semelhança entre os elementos, cabendo a verificação das possíveis classes para avaliar a existência de algum significado útil [de Carvalho 2005].

4. Conclusões

Critérios fixos e precisos de classificação, em geral, não são adequados, quando um estudo apresenta resultados próximos à divisão das classes, o que pode ser tratado por SIDs, que também são convenientes e possuem capacidade para tratar problemas que apresentam incerteza ou imprecisão para a tomada de decisão.

Ao adicionar um novo atributo aos resultados prévios, o modelo de inferência é capaz de decidir, de forma única, entre suas três possibilidades e, então, agrupá-las a partir de um algoritmo não-supervisionado e com estabelecimento dinâmico do número de grupos, esperando que o resultado desse agrupamento confirme o particionamento do conjunto em três grupos, não necessitando de limites fixos e precisos para a identificar possíveis SNPs.

Referências

Arbex, W. (2009). fuzzyMorphic.pl. 1 CD. Perl. Ambiente UNIX-like com GUI e interpretador Perl 5.0 ou posterior.

de Almeida, P. E. M. and Evsukoff, A. G. (2005). *Sistemas fuzzy*, pages 169–202. Manole, Barueri.

de Carvalho, L. A. V. (2005). *Datamining: a mineração de dados no marketing, medicina, economia, engenharia e administração*. Ciência Moderna, Rio de Janeiro.

Klir, G. J. and Yuan, B. (1995). *Fuzzy sets and fuzzy logic: theory and applications*. Prentice Hall, Upper Saddle River.

Nickerson, D. A., Taylor, S. L., Kolker, N., Sloan, J., Bhangale, T., Stephens, M., and Robertson, I. (2008). *Polyphred users manual*. University of Washington, Seattle. Version 6.15 Beta.

Tanscheit, R. (2007). *Sistemas fuzzy*, pages 229–264. Thomson Learning, São Paulo.

Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. on Systems, Man, and Cybernetics*, SMC-3:28–44.