

Emprego de Algoritmos de Clusterização e Sistemas Especialistas na Determinação de Assinaturas Radar

MARCELO MENEZES PIMENTEL¹ e PAULO MARCELO TASINAFFO¹

¹Instituto Tecnológico de Aeronáutica, Pça Mal. Eduardo Gomes, 50 – Jd. das Acácias – S. J. dos Campos – SP

Resumo — O processo de gestão de assinaturas radar carece de ferramentas adequadas à automatização da produção de conhecimento a cerca do potencial de GE estrangeiro. Com isso, o lapso para a manutenção dos registros de assinaturas radar tem aumentado, criando condições inseguras para a programação dos sistemas de ESM ou autodefesa dos vetores militares. As tecnologias da Inteligência Artificial oferecem excelentes soluções para diversos problemas relacionados à produção de conhecimento, contrapondo-se às limitações dos seres humanos no campo da classificação e gerenciamento de padrões. Empregando-se algoritmos de clusterização e sistemas especialistas é possível otimizar a produção de conhecimento nessa área.

Palavras-chaves — Radar, Inteligência Artificial, clusterização, sistemas especialistas, sistemas de produção, Guerra Eletrônica.

I. INTRODUÇÃO

Assegurar o uso do espectro eletromagnético por nossas forças e impedir, reduzir ou prevenir seu uso contra os interesses do país, é o que a Política de Guerra Eletrônica de Defesa [1] preconiza para as atividades de guerra eletrônica nas Forças Armadas, revelando o alto grau de comprometimento que os setores de inteligência devem ter com relação à produção de conhecimento no que diz respeito à real capacidade nacional e estrangeira de uso do espectro eletromagnético para guerra eletrônica.

À frente da coordenação das atividades de guerra eletrônica dos três comandos militares, o Ministério da Defesa tem desenvolvido esforços no sentido de promover um maior intercâmbio de informações, através da integração dos bancos de dados de guerra eletrônica das FFAA, com o intuito de melhorar a condução das atividades de inteligência nesse campo. Apesar disso, de nada adianta aumentar a disponibilidade de dados sem o arcabouço tecnológico adequado para que os processos sejam capazes de abstrair a informação oportuna e necessária à produção de conhecimento.

A Inteligência Artificial – IA apresenta diversas tecnologias que vêm sendo utilizadas para a solução de velhos e novos problemas das demais ciências, onde o desenvolvimento de agentes inteligentes voltados para a descoberta de padrões interessantes em grandes massas de dados [2] [3]– encampados sob a denominação de Mineração de Dados – é uma necessidade comum a quase todas. Outra importante contribuição da IA é a tecnologia para produção de sistemas especialistas voltados para o controle de ações, desde que fatos e regras de um domínio de problema constituam as bases de uma engrenagem alimentada por novos fatos que, por sua vez, irão disparar as ações das regras, gerando outros fatos e ações de forma cíclica.

A investigação do ferramental tecnológico da Inteligência Artificial aplicável ao problema de classificação e controle das assinaturas radar levou à realização de alguns experimentos sobre análise de agrupamentos de sinais radar com emprego de algoritmos de clusterização e à investigação de sistemas especialistas para determinação de assinaturas radar.

Assim, na seção 2 a problemática que envolve o processo de determinação e gestão das assinaturas radar será caracterizada dentro de uma perspectiva que naturalmente aponta para a investigação de soluções suportadas pelas tecnologias da Inteligência Artificial. A seção 3 introduz o problema de classificação não-supervisionada, reduzido à análise de algoritmos de clusterização. A seção 4 apresenta uma noção geral sobre os fundamentos que regem soluções pautadas no emprego de sistemas especialistas. A seção 5, apresenta uma proposta de arquitetura híbrida para a solução do problema de determinação e gestão de assinaturas radar, reunindo as duas tecnologias das seções anteriores. Finalmente, as conclusões apresentadas remetem a uma breve reflexão.

II. O PROBLEMA DE DETERMINAÇÃO DE ASSINATURAS RADAR

Dentre as dificuldades para o levantamento do potencial de guerra eletrônica de forças hostis, uma delas corresponde à determinação de assinaturas radar a partir de conjuntos de suas emissões, coletadas durante a realização das missões de reconhecimento eletrônico da ELINT. Este problema subdivide-se em dois:

- como descobrir os modos de operação dos radares num espaço amostral ruidoso onde os seus sinais fazem parte de distribuições desconhecidas?
- a partir das hipóteses admitidas como solução para o subproblema anterior, como identificar os emissores correspondentes aos modos de operação estabelecidos, considerando-se as ambigüidades provocadas pela possibilidade de dois ou mais emissores operarem com as mesmas características?

Atualmente, a realização dessas tarefas depende, quase que exclusivamente, da experiência dos especialistas de guerra eletrônica, cuja perspicácia para a identificação e correlação de dados afins é fundamental para a manutenção dos registros de assinaturas dos radares de interesse. A fig. 1 retrata os principais parâmetros que caracterizam os sinais radar [4]. A partir da diferença entre os valores medidos para cada parâmetro, o analista verifica a similaridade entre os elementos constantes da amostra e distribuindo-os em grupos,

de tal forma que elementos similares estejam no mesmo grupo, enquanto os dissimilares estejam em grupos diferentes. Esta é, praticamente, a definição para o problema de clusterização apresentado na próxima seção.

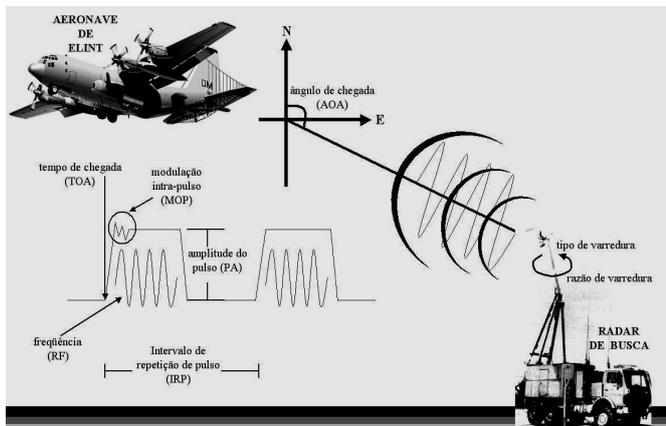


Fig. 1 – Parâmetros característicos dos sinais radar detectados em missões de ELINT.

No segundo caso, uma vez que um modo de operação seja caracterizado por um determinado grupo de sinais, nem sempre o conhecimento que se tem do potencial de guerra eletrônico de forças neutras ou hostis, permite imediata associação aos seus prováveis emissores, dentro de limites aceitáveis de confiabilidade. Corre-se o risco, então, de que o acúmulo de potenciais associações postergadas provoque distorções, em virtude da incapacidade humana de lidar, simultaneamente, com um número elevado de informações. Provavelmente, isto não aconteceria caso estas possíveis associações fossem evidenciadas, automaticamente, à medida que a inserção de novos dados satisfizesse as condições para tal. Esta idéia será abordada na seção que trata de sistemas especialistas.

III – ALGORITMOS DE CLUSTERIZAÇÃO

Segundo a literatura [3]-[5]-[6] a formação de agrupamentos, mais conhecida na comunidade científica por clusterização (neologismo de *clustering*, cujo significado é “agrupamento”) é um problema de classificação não-supervisionada. Para a resolução de problemas desse tipo, graças ao caráter multidisciplinar da Inteligência Artificial diversos campos da ciência, tais como o da Lógica, Matemática, Biologia, Psicologia, Neurociência, Linguística, etc. oferecem o embasamento teórico para o desenvolvimento das tecnologias disponíveis.

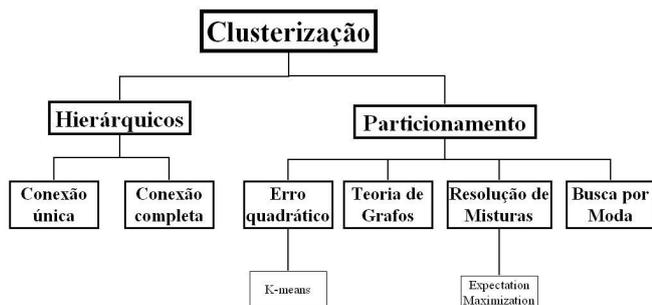


Fig. 2 – Taxonomia de Jain(1999) para Clusterização.

Em Jain [5] pode-se ver uma taxonomia para os métodos tradicionais de clusterização apresentada na fig. 2,

onde os métodos se dividem em duas categorias principais denominadas de acordo com a configuração das partições produzidas, ou seja, enquanto na abordagem *hierárquica* as partições são aninhadas em série, na abordagem de *particionamento* elas se resumem numa única partição.

Além destas tradicionais técnicas, outras mais recentes fazem parte do conjunto de ferramentas que a IA disponibiliza para que sejam selecionadas de acordo com as características de cada problema e dos conjuntos de dados que se tem a mão. Embora os métodos escolhidos para experimentação tenham sido o *K-means* e o *Expectation Maximization (EM)*, métodos tais como, vizinho mais próximo (*nearest neighbor clustering*); clusterização nebulosa (*fuzzy clustering*); redes neurais artificiais; abordagens evolutivas; abordagens baseadas em busca; e incorporação de restrições de domínio; podem ser vistos em [2]-[5]-[6]-[7]-[8].

A identificação correta do número de grupos presentes na amostra é considerado um subproblema cuja solução influi bastante na qualidade da partição resultante, e dois níveis de complexidade quanto à determinação do número de grupos (clusters) dividem o problema da clusterização em duas categorias: *K-Clustering* – onde o número esperado de clusters é previamente fornecido como sendo o parâmetro *k* do algoritmo, e o PCA (problema de clusterização automática) – onde encontrar o número de grupos ideal faz parte do problema. No primeiro caso, a complexidade dos algoritmos é bem menor, já que a determinação do número “*k*” de clusters é de responsabilidade dos especialistas da área de aplicação que se baseiam na experiência para inferir o número de grupos (clusters) existentes num determinado conjunto de dados [8].

Uma atividade de agrupamento envolve, tipicamente, cinco passos: (a) adequação da representação de amostras; (b) definição de proximidade – por meio de distâncias ou medidas de similaridade adequadas; (c) agrupamento – propriamente dito; (d) abstração de dados; e (e) análise dos resultados. Em geral, os dois últimos passos exigem menos rigor quando os resultados apresentam significado evidente e, por isso, podem ser considerados opcionais no processo [5].

A aplicação dos passos anteriormente citados foi aplicada num estudo de caso envolvendo a utilização de algoritmos de clusterização sobre um conjunto de sinais radar. Para isso, foi utilizado o framework WEKA (um acrônimo para *Waikato Environment for Knowledge Analysis*) que se trata de um ambiente para verificação de desempenho de algoritmos de aprendizado de máquina (*machine learning*), desenvolvido em linguagem Java pelos pesquisadores da universidade de Waikato [10], e que possui diversos algoritmos de clusterização, dentre os quais foram selecionados o *SimpleKmeans* e o *EM (Expectation Maximization)*, em virtude da simplicidade do primeiro e da capacidade de estimação do número de clusters do segundo.

A fig. 3 apresenta a interface do Weka Explorer integrante do framework WEKA. Nela pode-se ver um histograma da amostra utilizada, constituída de 17 modos de operação de 9 tipos diferentes de emissores, distribuídos em 165 instâncias de sinais radar. Cada instância da amostra foi representada por um vetor de 3 dimensões relativas aos parâmetros de RF (*radio frequency*), PRF (*pulse repetition frequency*) e PW (*pulse width*). Apesar das novas tecnologias incorporadas ao projeto dos sensores permitir a extração de

parâmetros intrapulso a partir das medições sobre modulações (intencionais ou não) da frequência portadora, estudos recentes ainda buscam incorporar estas características na identificação de emissores a partir de um único pulso [9], porém estes dados não foram disponibilizados para fins deste estudo.

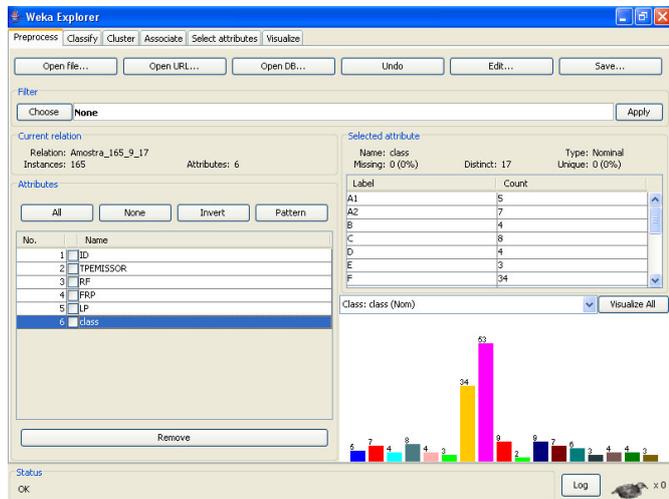


Fig. 3 – Interface do Weka Explorer carregada com o conjunto de instâncias de sinais radar.

Os resultados dos experimentos estão resumidos na tabela 1, onde se pode ver que mesmo tendo que estimar o número de clusters da amostra, o algoritmo EM apresentou melhores resultados em média, chegando no seu melhor resultado igual ao melhor de SimpleKmeans, sempre ajustado para $k=9$ (9 tipos de emissores).

TABELA 1 – RESULTADOS SIMPLEKMEANS VS. EM

Iter.	Atribuições erradas		Porcentagem de erros	
	SimpleKmeans	EM	SimpleKmeans	EM
1	48	38	29.0909	23.0303
2	43	38	26.0606	23.0303
3	45	51	27.2727	30.9091
4	30	31	18.1818	18.7879
5	30	39	18.1818	23.6364
6	69	42	41.8182	25.4545
7	64	52	38.7879	31.5152
8	55	32	33.3333	19.3939
9	34	43	20.6061	26.0606
10	48	30	29.0909	18.1818
Média	46.6	39.6	28.2424	24

Percebe-se que 3 dimensões é um número insuficiente para permitir a distinção entre modos de operação, como se pode observar na imagem (capturada de uma das telas do Weka Explorer) da fig. 4, cuja melhor estimativa para o número de clusters, feita pelo algoritmo EM, considera 9 clusters em vez de 17.

Verificando-se mais tarde a forte tendência de divisão do tipo G em dois grupos, junto a especialistas de guerra eletrônica, viu-se que a existência uma bipolarização em torno de dois valores para PW das instâncias desse tipo de emissor, apontavam para a possibilidade de se tratarem de dois modos de operação distintos, sendo bastante desejável que resultados desse tipo voltem a acontecer.

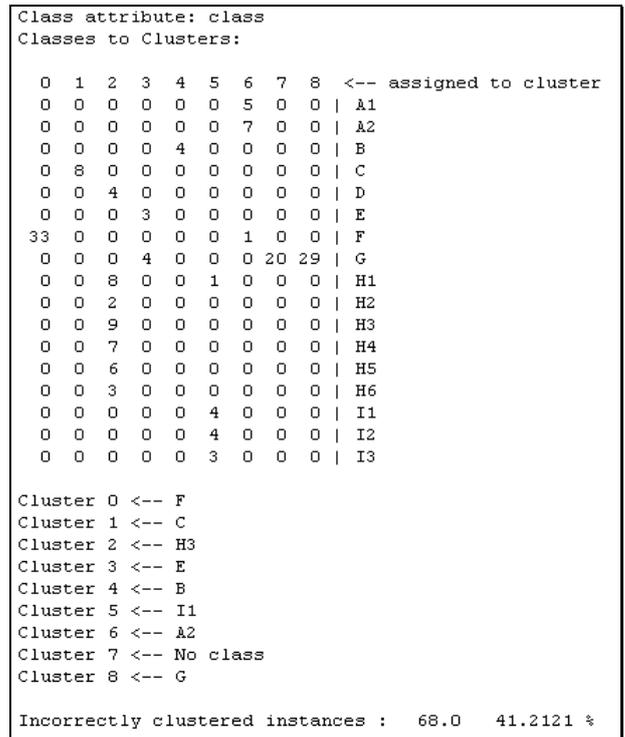


Fig. 4 – Distribuição obtida com a clusterização *Expectation Maximization*.

IV- SISTEMAS ESPECIALISTAS

A terminologia de sistemas especialistas, remonta do sucesso obtido com uma aplicação baseada na implementação de regras heurísticas, capturadas junto a alguns especialistas de Química, para resolver o problema de análise de compostos químicos orgânicos, baseando-se na interpretação da saída de um espectrômetro de massa. Esta aplicação recebeu o nome de DENDRAL e foi considerado como o primeiro sistema especialista [11]-[12].

Em razão do sucesso obtido, em 1977 Feigenbaum [3] introduziu uma das primeiras definições para sistemas especialistas: “*sistemas especialistas são sistemas que solucionam problemas que são resolvíveis apenas por pessoas especialistas (que acumularam conhecimento exigido) na resolução destes problemas*”. De outra forma, pode-se dizer, então, que programas de computador que tentam resolver problemas emulando o raciocínio humano de um especialista, aplicando conhecimentos específicos e inferências, são ditos sistemas especialistas.

Um exemplo de regras do mesmo domínio de conhecimento utilizado para o desenvolver o DENDRAL [3], é transcrito a seguir:

SE existem dois picos em x_1 e x_2 , tais que

- $x_1 + x_2 = M + 28$ (M é a massa da molécula inteira);
- $x_1 - 28$ é um pico;
- $x_2 - 28$ é um pico;
- no mínimo um entre x_1 e x_2 é alto.

ENTÃO existe um subgrupo cetona.

Esse exemplo é ilustra bem o uso de regras empregadas para o reconhecimento do subgrupo cetona ($C=O$) que pesa 28 unidades de massa.

A utilização da estrutura de regras baseadas em cláusulas SE... ENTÃO fez com que os sistemas especialistas se tornassem bastante conhecidos, também, pelos termos “sistemas baseados em conhecimento” ou “sistemas baseados em regras”. Segundo [13] os termos “sistema especialista” e “sistema baseado em conhecimento” são intercambiáveis, pois a tênue diferença entre ambos reside no fato de que a abordagem do primeiro envolve a aquisição de conhecimento proveniente de um *especialista* humano, enquanto no segundo não se dá tanta ênfase a um processo de aquisição de conhecimento particular.

Um sistema especialista é composto basicamente por uma base de conhecimento (regras e fatos), um mecanismo de inferência e uma interface com o usuário. Numa representação mais contemporânea, desde que um sistema especialista esteja inserido numa arquitetura capaz de interagir com o ambiente, ele pode ser visto na forma de um *agente baseado em conhecimento* [3]-[14].

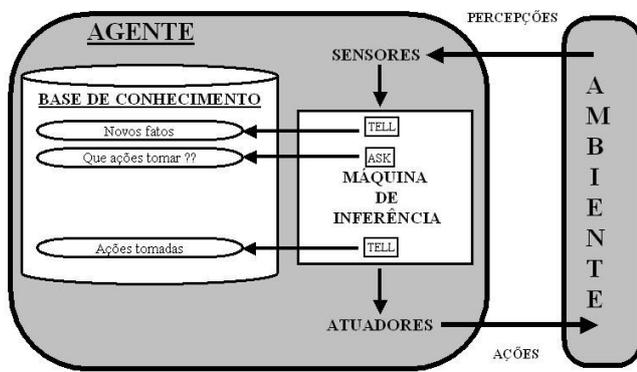


Fig. 5 – Arquitetura de agente baseado em conhecimento

A fig. 5, apresenta um esquema de um agente baseado em conhecimento, o qual é capaz de interagir com o ambiente através de sensores e atuadores. No interior desse agente, o seu *programa* utiliza mecanismos de inferência para executar ações, à medida que fatos novos, juntamente com fatos já conhecidos satisfaçam as premissas das regras da mesma base de conhecimento. Nesse ciclo, o mecanismo de inferência comunica a ação tomada e os novos fatos gerados à base de conhecimento, antes mesmo de interagir com o ambiente externo. Isso propicia a incorporação do aprendizado de novas hipóteses a partir de novos fatos, através de algoritmos de encadeamento para trás (*backward chaining*) – para buscar evidências de possíveis conclusões – ou de encadeamento para frente (*forward chaining*), para produzir conclusões a partir de evidências.

Provavelmente, a linguagem mais conhecida para programação em lógica é a Prolog [3], através da qual é possível a implementação de conhecimento através de regras que seguem os fundamentos teóricos da lógica de primeira ordem. Para isso é desejável que o conhecimento seja representado através de cláusulas definidas, ou seja, regras do tipo: **se** A e B, **então** C. Além da Prolog, outras linguagens para a implementação de sistemas especialistas se encontram disponíveis e o uso de alguns *shells* (*frameworks* para programação), tais como o CLIPS e o JESS, também são bastante difundidos, pois seus mecanismos de inferência são implementados por algoritmos bastante eficientes.

O *shell* eleito para a implementação do sistema especialista empregado na arquitetura proposta para gestão das assinaturas radar, foi o JESS, em virtude dos seguintes fatores: suporte à linguagem Java; possibilidade de criação de novas funções através de sua linguagem nativa; e, principalmente, pelo algoritmo Rete que faz com que o mecanismo de inferência do JESS seja considerado bastante rápido, conforme comprovam os resultados obtidos numa plataforma HotSpot da Sun com um processador Pentium III 800MHz. Nela, as tarefas de execução de 80.000 regras, execução de 600.000 operações de correspondência de padrões e adição de mais de 100.000 fatos à memória de trabalho, despenderam, cada uma, o tempo de apenas 1 segundo [15].

Um exemplo bastante simples mostra o que seria um conjunto mínimo de regras capaz de prover a classificação de círculos, quadrados e triângulos no espaço bidimensional de atributos da amostra apresentada na fig. 6.

- Se $(Y \geq y_3)$ então é triângulo;
- Se $(Y \geq y_2)$ e $(Y \leq y_3)$ então é quadrado;
- Se $(Y \leq y_2)$ então é círculo.

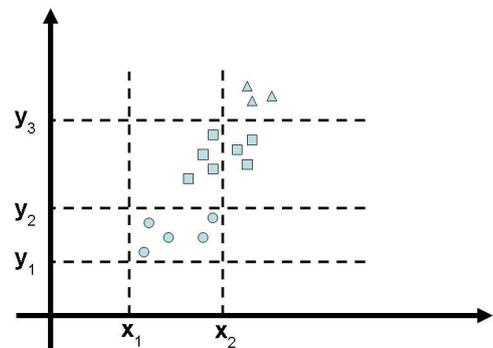


Fig. 6 – Classificação de um conjunto de amostras num espaço bidimensional de atributos

V- ARQUITETURA PROPOSTA

Diante da problemática que envolve o processo de determinação e gestão das assinaturas radar, apresentada na seção 2 e diante das potencialidades das ferramentas de IA descritas nas demais seções, o presente trabalho apresenta a proposta de adoção de arquiteturas que considerem mais de uma tecnologia de Inteligência Artificial, tais com a apresentada nesta seção, onde é considerado o emprego concomitante do algoritmo de clusterização EM e um sistema especialista desenvolvido em JESS para a realização de inferências sobre as saídas daquele, conforme o esquema da fig. 7.

Acompanhando a figura 7, o processo proposto inicia a partir da inserção dos sinais radar, detectados pelos sensores de ELINT numa base de dados. Esses sinais, juntamente com outros que não puderam ainda ser classificados, são então submetidos ao algoritmo de clusterização para fins de determinação dos *clusters*. Os *clusters* formados passam a representar as hipóteses iniciais sobre a distribuição das instâncias dos sinais radar.

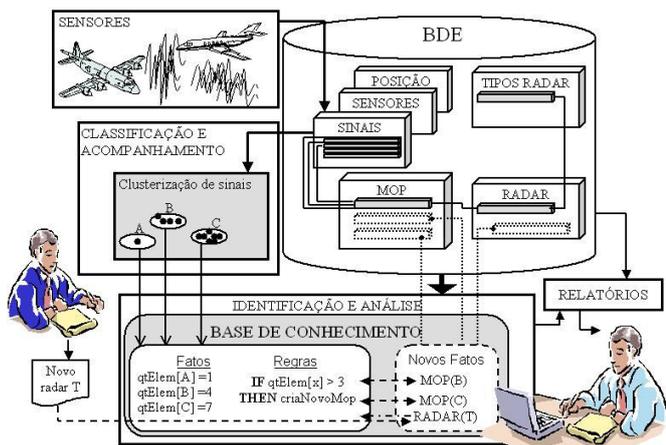


Fig. 7 – Arquitetura híbrida de gestão de assinaturas radar, com emprego de algoritmos de clusterização e sistema especialista.

Incluindo-se aquelas hipóteses na base de conhecimento do sistema especialista, elas passam a constituir novos fatos. Caso a conjunção desses fatos com outros já conhecidos complete as premissas das regras, os algoritmos de encadeamento para a frente ou para trás do mecanismo de inferência irá validar ou rejeitar as hipóteses da clusterização. Uma vez validadas, novos fatos são gerados, os quais podem completar as premissas de outras regras, provocando, novamente, o disparo de outras ações. Estas novas ações podem incluir:

- a correlação entre os modos de operação recém identificados com outros pré-existentes (refinamento de assinaturas);
- a determinação de um novo modo de operação para um radar de acordo com as características deste (identificação de assinatura);
- correlação de radares à medida que forem caracterizados na base de conhecimento (resgatando modos de operação sem relacionamento);
- identificação de intenções de movimento (pela discriminação de posições de radares móveis);
- etc.

VI- CONCLUSÃO

A busca pela excelência no campo da guerra eletrônica almejada pelo Ministério da Defesa, não pode pautar-se tão somente pelo aumento da disponibilidade de dados, pois, como foi visto, é necessário que os processos atuais sejam suportados por tecnologia adequada e oportuna para a extração e gestão de conhecimento.

Nesse sentido, vimos como é possível abstrair automaticamente as distribuições que indicam a presença de diferentes tipos de emissores, a partir de uma amostra de sinais radar, para fins de caracterização de seus modos de operação e consequente determinação de suas assinaturas, utilizando, para isso, algoritmos de clusterização, cujos resultados justificaram plenamente o seu emprego.

Contudo, a forte dependência de outras informações para a correta associação entre modos de operação e seus prováveis emissores fez despontar uma solução com base no emprego de sistemas de produção e algoritmos de clusterização juntos numa arquitetura híbrida capaz de gerar

resultados satisfatórios na condução das atividades de apoio à guerra eletrônica, ao longo do tempo.

De outro modo, a capacidade de processamento da enorme quantidade de dados esbarraria nas limitações humanas. Se em tempo de paz essas limitações já causam um lapso grande entre as atualizações de informação, em caso de conflito o sistema provavelmente entraria em colapso.

Apesar disso, a impossibilidade de abstrair os parâmetros intrapulse dos sinais, evidencia a necessidade de maiores investimentos nos setores de pesquisa e desenvolvimento para que o Brasil alcance patamares tecnológicos mais altos e se liberte da dependência de soluções estrangeiras de defesa que limitam o estabelecimento de doutrinas mais modernas de emprego.

REFERÊNCIAS

- [1] BRASIL. Ministério da Defesa. Portaria Normativa Nº 333/MD, Brasília, março, 2004.
- [2] HAN, J. et KAMBER, M. Data Mining: Concepts and Techniques. Academic Press, Morgan Kaufmann Publishers, 2001.
- [3] RUSSEL, S. et NORVIG, P. Inteligência Artificial. Editora Campus, 2004.
- [4] BRASIL. Ministério da Aeronáutica. MMA 500-2: Manual de Fundamentos de Guerra Eletrônica. Brasília, DF, 1997. Reservado.
- [5] JAIN, A. K.; DUBES, R. C. Algorithms for clustering data. Englewood Cliffs, New Jersey, Prentice Hall, 1988.
- [6] JAIN, A.K.; MURTY M.N.L; FLYNN P.J. Data Clustering: A Review. In ACM Computing Surveys, Vol 31, No. 3, 264-323. 1999.
- [7] WITTEN, I. H.; FRANK E. Data mining: practical machine learning tools and techniques. Morgan Kaufmann, San Francisco, 2005.
- [8] GHOSH, J. Scalable clustering. Edited by Nong Ye in Handbook of data mining. Lawrence Erlbaum Associates, Inc., Publishers, Mahwah, New Jersey, London. 2003.
- [9] KAWALEC, A; OWCZAREK, R. Specific emitter identification using intrapulse data. In European Radar Conference, Amsterdam. 2004.
- [10] UNIVERSIDADE DE WAIKATO. WEKA 3: Data Mining Software in Java. Disp. em <<http://www.cs.waikato.ac.nz/~ml/weka/index.html>>. Acesso em: 20 ago. 2005.
- [11] FEIGENBAUM, E. A.; BUCHANAN, B. G.; LEDERBERG, J. On generality and problem solving: A case study using the DENDRAL program. In Meltzer, B. and Michie, D., editors, Machine Intelligence 6, pages 165-190. Edinburgh University Press, Edinburgh, Scotland. Feigenbaum, E. A. and Feldman, J., editors, 1971.
- [12] LINDSAY, R. K.; BUCHANAN, B. G.; FEIGENBAUM, E. A.; LEDERBERG, J. Applications of Artificial Intelligence for Organic Chemistry: The DENDRAL Project. McGraw-Hill, New York, 1980.
- [13] MILES, J. A. H. Artificial intelligence and command and control. Ed. em HARRIS, C. J. Application of artificial intelligence to Command & Control systems. Peter Peregrinus Ltd, Londo, United Kingdom, 1988.
- [14] TASINAFFO, P. M. Apostila de IA, 2005. Disp. em: <<http://www.comp.ita.br/~tasinafo/>> Acesso em: 29/10/2006.
- [15] FRIEDMANN-HILL, E. Jess in Action. Manning, Greenwich. 2003