
ABSTRACT

Artificial Neural Networks (ANNs) represent an excellent tool that has been used to develop highly accurate models in numerous real-world problem domains. However, despite the proven advantages of ANNs, the notorious difficulty to understand how they arrive at a particular decision has led to a barrier to a more widespread acceptance of them, especially in some industry domains.

In recent years, some works have been developed with the aim of redressing this general problem of lack of ANN explanation capability. In particular, a substantial part of these works have focused on a line of investigation involving the development of techniques for extraction of the hidden knowledge in ANNs. This body of work is actually referred to as *rule extraction* and this name reflects the fact that these works have been largely concentrated on translating ANN hypotheses into inference-rule languages.

Following the line of rule extraction investigation, this thesis presents a new methodology for fuzzy rule extraction from ANN, with full theoretical support and practical validation. The methodology, called Transparent Fuzzy Rule Extraction from Neural Networks (*TFRENN*), was developed with the goal of overcoming the main limitation of some previous methodologies, i.e. the extraction of transparent fuzzy rules. The importance of obtaining a transparent rule set must be underlined, because it allows full understanding by humans of the hidden knowledge captured by ANNs when trained to fit data in a given problem.

The efficiency of *TFRENN* methodology and the importance of transparent fuzzy systems for knowledge discovery are verified by the application of the methodology in Transformer Incipient Fault Diagnosis using DGA (Dissolved Gas-in-oil Analysis). Many diagnosis systems based on ANN have been developed and presented in the literature with good results. However, these systems have had difficulty to be accepted by utilities, perhaps due to the lack of ANN behavior explanation. The results presented in this thesis shown that the application of *TFRENN* methodology to Transformer Incipient Fault Diagnosis can overcome this limitation.

The *TFRENN* approach allowed producing a very good tool for fault diagnosis, with results better than the ones published for comparable techniques. Furthermore, it allowed the discovery of new rules for classifying faults, which led to building a new diagnosis table useful for practical purposes. This new table is an improvement to the diagnosis table published by IEC, which is actually one of the tables most used to transformer incipient fault diagnosis.

RESUMO

As Redes Neurais Artificiais constituem uma excelente ferramenta que vem sendo largamente utilizada para o desenvolvimento de modelos, com um alto grau de exactidão, em um grande número de aplicações. Contudo, apesar das conhecidas vantagens das redes neuronais, a evidente dificuldade encontrada para se compreender como ela chega a uma determinada decisão constitui um impedimento para sua mais larga aceitação, especialmente em alguns domínios industriais.

Recentemente, alguns trabalhos vêm sendo desenvolvidos com o objectivo de solucionar o problema de falta de capacidade de explanação das redes neuronais. Em particular, uma grande parte destes trabalhos tem sido direccionada para uma linha de investigação envolvendo o desenvolvimento de técnicas para extracção do conhecimento escondido nas redes neuronais. Este tipo de pesquisa vem actualmente sendo referenciado como *extracção de regras* e este nome reflecte o facto de que estes trabalhos estão largamente concentrados na translação da hipótese fornecida pela rede neuronal para uma linguagem baseada em regras de inferência.

Seguindo esta linha de pesquisa, esta tese apresenta uma nova metodologia para extracção de regras difusas de uma rede neuronal artificial com um completo suporte teórico e validação prática. A metodologia, chamada TFRENN (Extracção de regras difusas transparentes de uma rede neuronal), foi desenvolvida com o principal objectivo de extrair regras difusas transparentes de redes neuronais. A importância de se obter regras transparentes deve ser ressaltada devido ao facto de que estas permitem uma completa compreensão, por parte dos seres humanos, do conhecimento capturado pela rede neuronal treinada para um determinado problema.

A eficiência da metodologia, bem como a importância de regras difusas transparentes para a descoberta do conhecimento são verificadas através da aplicação da metodologia em um sistema de diagnóstico de faltas incipientes em transformadores usando a análise dos gases dissolvidos em óleo. Muitos sistemas de diagnóstico baseados em redes neuronais que vêm sendo desenvolvidos e apresentados na literatura têm apresentado bons resultados. Contudo, talvez devido a falta de capacidade de explanação das redes neuronais, tais sistemas têm tido sua aceitação dificultada pelas empresas. Os resultados apresentados nesta tese mostram que a aplicação da metodologia TFRENN para o diagnóstico de faltas incipientes em transformadores pode superar este problema.

A metodologia TFRENN permitiu o desenvolvimento de uma óptima ferramenta para o diagnóstico de faltas, com resultados superiores que os já publicados usando técnicas semelhantes. Adicionalmente, esta técnica permitiu ainda a descoberta de novas regras para classificação de faltas, as quais proporcionaram a construção de uma nova tabela para diagnóstico de faltas, útil em aplicações práticas. Os resultados desta nova tabela mostraram-se superiores que os apresentados pela tabela publicada pelo IEC, que é actualmente uma das mais utilizadas para diagnóstico de faltas incipientes em transformadores.