Rede de Sensores para Controle Inteligente de Ambientes

Artejose Revoredo da Silva[1], **João C. Xavier-Júnior**[1], **Ivanovitch Medeiros Dantas da Silva**[1]

[1]Instituto de Metrópole Digital – Universidade Federal do Rio Grande do Norte – (UFRN)
Av. Senador Salgado Filho, 3000 – 59.078-970 – Natal – RN

{ivan,jcxavier}@imd.ufrn.br, artejose.rs@gmail.com

Abstract. *In general, laboratories or labs are workplaces that provide controlled conditions for experiments and measurements to be performed. For this reason, controlling temperature and humidity is an important requirement that needs to be achieved in order to guarantee the reproducibility of processes carried out in labs. Aiming to introduce efficient environmental controlling mechanisms, we present in this paper an intelligent environmental control system based on sensors network and pattern classification. Our system uses the information generated by the sensors to make decisions for constant controlling of temperature and relative humidity within the lab.*

Resumo. *Laboratórios são ambientes que necessitam ser controlados, principalmente devido ao fato de que nesses ambientes são realizados processos químicos. Controlar a temperatura e a umidade relativa do ar é tarefa primordial para que os processos desenvolvidos dentro de tais laboratórios possam ser repetidos dentro dos parâmetros exigidos. Buscando propor mecanismos de controle para ambientes, neste artigo, apresentamos um sistema baseado em rede de sensores e classificação de padrões para controle inteligente de ambientes. Nosso protótipo utiliza as informações gerados pelos sensores para tomar decisões relativas ao controle constante da temperatura e da umidade relativa do ar.*

1. Introdução

A evolução exponencial das tecnologias tem provocado o surgimento de soluções que não eram pensadas apenas há alguns anos atrás. Entre as áreas emergentes destaca-se a Internet das Coisas. Em um futuro muito breve a maioria dos dispositivos que nos cercam estarão conectados na Internet através de uma grande rede de sensores. Esse cenário irá provocar o surgimento de uma quantidade colossal de dados. Esses por sua vez, precisarão ser processados para gerar informação útil. Nesse contexto, surge uma nova e promissora área de pesquisa, a Ciência dos Dados. Essa nova frente de pesquisa irá impactar diretamente a industria e especificamente, a automação. Técnicas de classificação de dados e tomada de decisões podem ser empregadas como uma ferramenta da Ciência dos Dados para gerar sistemas automatizados de controle e monitoramento de ambientes.

Baseado no cenário anterior, o respectivo trabalho busca desenvolver um sistema inteligente para automatizar o controle e monitoramento térmico de labo-

ratórios de análises químicas a partir de uma rede de sensores e algoritmos de classificação de dados. Em laboratórios químicos, controlar as variações de temperatura e umidade resulta diretamente na melhoria da qualidade das amostras trabalhadas, assim como a economia da própria energia elétrica utilizada para abastecer aparelhos de ar-condicionados e desumificadores. As ferramentas da Ciência dos Dados serão utilizadas para auxiliar a tomada de decisão no controle de tais equipamentos.

O restante deste artigo está dividido em 5 seções. A Secção II descreve alguns conceitos teóricos importantes e alguns estudos relacionados ao tema deste artigo. A seção III discute os conceitos relacionados. Na Seção IV, a metodologia utilizada no trabalho experimental do presente trabalho é apresentada, enquanto que uma análise dos resultados fornecidos pela análise empírica é mostrada na Seção V. Finalmente, a Seção VI apresenta as considerações finais deste trabalho.

2. Trabalhos Relacionados

Os sistemas de controle e monitoramento de ambientes inteligentes apresentam diversos desafios, englobando aspectos de desempenho (rede, latência, confiabilidade), segurança e de tomada de decisão. O referido trabalho está focado nos desafios de tomada de decisão a partir de uma rede de sensores para o controle térmico de laboratórios químicos. Na literatura, as principais contribuições utilizam aspectos de automação *on/off*, ligando ou desligando luzes e ar-condicionados. Aspectos de tomada de decisão embarcadas em redes de sensores são pouco explorados.

No trabalho apresentado em [da Silva 2013], uma solução de automação residencial baseado na plataforma Arduino foi implementada. Controle de iluminação, abertura e fechamento de portas, alarmes e câmeras foram utilizadas para validação da proposta. Em outro trabalho [Maycon J. C. Mesquita 2014], um procedimento de controle manual de equipamentos residenciais através de um sistema web foi implementado. No entanto, em ambos os trabalhos citados não houve a exploração dos aspectos de tomada de decisão em conjunto com rede de sensores para uma automação mais eficiente.

Um sistema para monitoramento de concreto baseado em redes de sensores de temperatura e umidade foi apresentado em [Barroca et al. 2013]. Na mesma linha de pesquisa, foi desenvolvido em [Abraham and Li 2014] uma rede de sensores para o controle da qualidade do ar em ambiente fechados. Ambos não descreveram técnicas de controle inteligente (classificação de padrões) na tomada de decisões.

Outros trabalhos que, de forma geral utilizaram sensores de temperatura e que implementaram algum tipo de inteligência, focaram em aplicações diversas. [Vera-Repullo et al. 2015] utilizou sensores para capturar informações de temperatura, umidade e solo. A partir de tais informações, aplicou regras pré-definidas para controlar a quantidade e duração da irrigação de vegetais. Em outro trabalho [Yuwono et al. 2015], foi utilizado seleção não supervisionada de atributos e um algoritmo de agrupamento inspirado no Enxame de Partículas, do inglês *Particle Swarm Optimization (PSO)*, para detecção automática de falhas e diagnóstico em sistemas de ventilação e de ar-condicionado.

A partir da discussão acima, é importante salientar que nenhum dos trabalhos encontrados na literatura utilizou algoritmos de agrupamento para encontrar

os grupos que são particulares de um ambiente específico tampouco algoritmos de classificação para tomar decisões. Além disso, poucos utilizam rede de sensores de temperatura e umidade.

3. Controle Térmico em Laboratórios

Os laboratórios de análise química têm características bem específicas, pois podem variar de acordo com o prédio que guarda suas instalações. Adicionalmente, existem exigências que precisam ser seguidas, nomeadamente o controle da temperatura e umidade relativa do ar. O controle dessas variáveis é de suma importância haja vista que influencia diretamente no resultado das amostras assim como danificar os equipamentos do laboratório. O objeto principal deste trabalho, o controle de ambientes, está sendo testado em um dos laboratórios de análises químicas da UFRN.

Devido a fatores tão importantes, é essencial distribuir vários sensores pelo laboratório, de modo que todo o espaço esteja coberto por eles. A Figura 1 descreve o cenário utilizado no trabalho. Perceba que cinco sensores (azul claro) foram distribuídos pelo espaço físico, sendo que três ficaram na parte à esquerda (maior número de equipamentos) e dois ficaram na parte à direita (menor número).

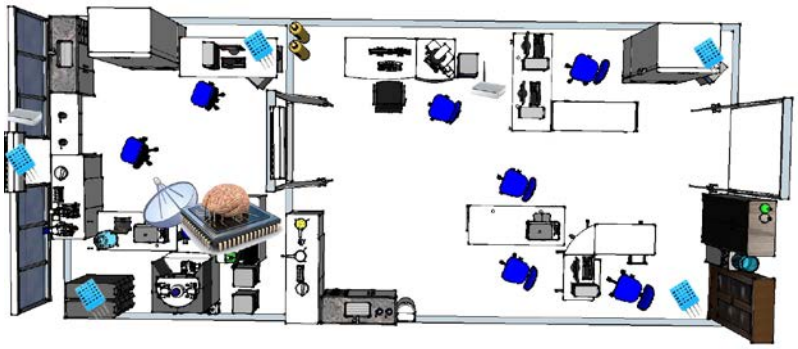

Figure 1. Distribuição dos sensores no laboratório.

Em relação aos aspectos de hardware, existem diversos sensores de fácil aquisição que podem ser utilizados para o monitoramento de temperatura e umidade [Margolis 2011, McRoberts 2011]. A Figura 2 descreve os sensores mais utilizados para esse fim [McRoberts 2011], nomeadamente, o LM35 (sensor de precisão de temperatura) e o DHT11 (sensor de umidade e temperatura). Em testes realizados no laboratório, o sensor de temperatura LM35 apresentou melhor precisão que o sensor DHT11. Por essa razão, o sensor DHT11 foi utilizado apenas para monitorar a umidade.

Figure 2. Sensores mais utilizados para medição de temperatura e umidade.

Um aspecto também importante a ser mencionado é a questão da comunicação entre os sensores. No cenário descrito na Figura 1, os cinco sensores de

temperatura e umidade são conectador em uma topologia em estrela a um centralizador ou comumente chamado na área de redes de sensores de *sink*. Tal dispositivo irá coletar todos os dados e compartilhar através de um enlace Wifi a um banco de dados remoto. O centralizador é um hardware dedicado e construído a partir de microcontroladores ATMEGA e com software embarcados. Essa solução é uma das contribuições do trabalho. A Figura 3 descreve uma visão do centralizador.

Figure 3. Prototipo de monitoração e controle de temperatura e umidade.

4. Metodologia de Experimentos

Como pode ser visualizado na Figura 1, os sensores foram distribuídos em cinco pontos estratégicos da laboratório, visando desta forma, cobrir as áreas mais críticas com relação a temperatura e umidade. O tempo de aferição foi setado em 10 minutos, ou seja, a cada 10 minutos os sensores enviavam dados para o controle que os armazenava em uma base dados. Também utilizamos tempos de amostra iguais a 5 e 15 minutos, porém verificamos que não houve alteração significativa de temperatura e umidade. Depois disso, o protótipo ficou em execução por 50 horas, e com isso armazenou-se um total de 300 registros. Cada registro gerado é composto por 10 (dez) atributos, sendo cinco referente as temperaturas e cinco as umidades.

O objetivo principal desse primeiro experimento era gerar uma base de dados, para que a partir dela, se descobrisse algum padrão de comportamento de temperatura e umidade dentro do laboratório de análises químicas escolhido como caso de uso. O método utilizado para descobrir padrões na base de dados foi o de agrupamento ou *clustering*. Nesse método, os valores dos atributos são utilizados conjuntamente com uma medida de distância, para dessa forma, separar os registros ou instâncias em grupos ou *clusters*.

Na fase de agrupamento, escolhemos três métodos de agrupamentos bastante conhecidos na literatura, e a partir deles criamos partições de k grupos variando de dois a seis grupos. Dessa forma, gerou-se cinco bases clusterizadas (com k grupos) para cada método de agrupamento, perfazendo um total de vinte bases (4 métodos x 5 partições), já que um dos métodos foi usado duas vezes. Mudando-se para isso, alguns parâmetros probabilísticos. O passo seguinte foi avaliar as partições criadas. Para essa tarefa, escolheu-se dois índices internos de validação de agrupamentos, já que as bases não possuíam classes ou rótulos. Os índices geraram valores para cada uma das vinte bases. A partir de tais valores, escolheu-se as três melhores bases geradas por cada um dos quatro métodos de agrupamento. Essas bases foram usadas na fase seguinte.

Na fase de classificação, escolheu-se seis métodos de classificação bastante conhecidos na literatura. Além da melhor base (partição), de acordo com os índices, pegou-se as três melhores bases resultantes de cada método de agrupamento, perfazendo um total de 12 (doze) bases. Aplicamos validação cruzada para cada uma das bases. O objetivo principal dessa fase era verificar o desempenho dos métodos de classificação, considerando o percentual de acerto, aqui chamado de acurácia. Baseado na acurácia de cada um dos seis classificadores sobre cada uma das 12 bases resultantes da fase anterior, calculou-se a média e o desvio padrão para que fosse possível averiguar se os desempenhos eram significativamente diferentes. A partir do teste estatístico realizado, chegou-se ao método de classificação mais adequado para classificar (rotular) os registros futuros.

Na fase de controle, identificamos o conjunto de ações possíveis de controle para o ambiente em questão, e as variações de estados (possíveis valores da classe). Desse modo, o controle passou a executar as ações relacionadas aos estados, assim como, a desfaz as ações anteriores. Dentre os valores da classe, identificamos cinco possíveis valores, que são: frio, normal, calor1, calor2 e perigo. Para cada valor, há pelo menos uma ação vinvulada, que visa controlar os aparelhos de ar-condicionado e desumificador.

As subseções seguintes mostrarão em detalhes os métodos utilizados nas fases de agrupamento e classificação relatadas acima.

4.1. Algoritmos de Agrupamentos

Existe uma grande diversidade de algoritmos de agrupamento que podem ser encontrados na literatura [Xu and Wunsch 2005]. Nesta subseção serão apresentados os algoritmos utilizados na fase de agrupamento deste trabalho.

- k-Médias: o algoritmo k-Médias é comummente utilizado em tarefas de agrupamento, principalmente por de fácil implementação. Ele se baseia na ideia de centroides, onde o centroide representa a instância média de um grupo, ou seja, a instância que é a média de todas as outras dentro de um mesmo grupo. Os elementos são agrupados em k grupos com base em alguma medida de similaridade. Geralmente, é utilizado a distância Euclidiana.
- Expectância Máxima: o algoritmo Expectância Máxima, do inglês *Expectation-Maximization* - EM, é uma técnica usada para estimar funções de máxima verossimilhança a partir de dados incompletos, de modo que se os dados estão incompletos, pode-se utilizar os casos em que foram observadas as variáveis para aprender a predizer seus valores quando não observados[Faceli et al. 2011].
- Hierárquico Aglomerativo: os algoritmos de agrupamento hierárquicos geram uma sequência de partições aninhadas, semelhante a uma estrutura hierarquia do tipo árvore [Faceli et al. 2011]. Há dois tipos de algoritmos: os divisivos e os aglomerativos. Na primeira, os objetos são alocados em um único grupo, sendo divididos sucessivamente até o número de grupos escolhido previamente. Já no outro tipo, inicialmente, cada objeto representa um grupo, sendo aglomerados até o número de grupos desejado.

4.2. Medidas para Validação

Como forma de validar as 20 bases criadas a partir da utilização dos algoritmos de agrupamentos citados na subseção anterior, utilizou-se os índices Davies-Bouldin (DB) e Silhueta [Halkidi et al. 2002].

Índice Davies-Bouldin (DB): esta medida calcula a compacticidade dos grupos formados, através da dispersão intra-grupos e entre os grupos. Dado uma partição $C = \{C_1, C_2, ..., C_k\}$, o índice DB pode ser definido da seguinte forma:

$$DB(C) = \frac{1}{k} \sum_{i=1}^{k} \max_{j=1,...,k|j \neq i} \left\{ \frac{\Delta(C_i) + \Delta(C_j)}{\delta(C_i, C_j)} \right\} \tag{1}$$

em que k é o número de *clusters*; $\Delta(C_i)$ é a distância intra-*cluster* para o *cluster* i e $\delta(C_i, C_j)$ é a distância entre os centróides dos *clusters* C_i e C_j.

A distância intra-*cluster*, $\Delta(C_i)$, representa a distância média de todos os elementos do *cluster* para o seu centróide e pode ser definida como sendo:

$$\Delta(C_i) = \frac{1}{|C_i|} \sum_{x \in C_i} (x - z_i)^2 \tag{2}$$

em que x é uma instância de i e z_i é o centróide do grupo i.

Silhueta: a medida de silhueta mede a qualidade dos *clusters* com base na proximidade entre as instâncias de um cluster e na distância das instâncias de um *cluster* com relação aos outros *clusters*. As silhuetas mostram quais instâncias estão bem situados dentro dos seus *clusters* e quais não estão. Seja $a(x_i)$, a distância média da instância x_i em relação a todas as outras instâncias do *cluster* C_i; $d(x_i, C_j)$, a distância média da instância x_i em relação as instâncias do *cluster* C_j; e $b(x_i)$, a menor distância média de x_i em relação a todos os demais *clusters*, pode ser defina como sendo:

$$b(x_i) = \lim_{C_i \neq C_j} d(x_i, C_j) \tag{3}$$

Desse modo, a silhueta de uma instância, $s(x_i)$, pode ser definida como sendo:

$$s(x_i) = \begin{cases} 1 - \frac{a(x_i)}{b(x_i)}, & a(x_i) < b(x_i) \\ 0, & a(x_i) = b(x_i) \\ \frac{b(x_i)}{a(x_i)} - 1, & a(x_i) > b(x_i) \end{cases} \tag{4}$$

4.3. Algoritmos de Classificação

Os algoritmos de classificação, também chamados de algoritmos preditivos, são aplicados em problemas onde as instâncias de uma base de dados, representados por um

conjunto de atributos, precisam ser enquadrados em um conjunto pré-definidos de possíveis rótulos (classes). Há um bom número de algoritmos de classificação propostos na literatura. Os algoritmos utilizados neste trabalho são descritos a seguir.

- k-Nearest-Neighbor (k-NN): o algoritmo k-NN, também conhecido como k vizinho mais próximo, é um método baseado em instância que classifica uma instância de teste baseada na a classe referente as instâncias mais próximas pertencentes ao conjunto de treinamento [Mitchell 1997].
- Árvore de decisão (AD): esse algoritmo utiliza a técnica de dividir para conquistar na solução de problemas de classificação. Seu funcionamento consiste em, dado um conjunto de treinamento, gerar uma árvore de decisão como saída. Essa árvore será utilizada para classificar as instâncias de teste [Berry and Linoff 2004].
- Rede Neural Artificial: o funcionamento das redes neurais artificiais é baseado no sistema nervoso. Uma RNA é composta por neurônios artificiais que estão conectados em camadas. Os neurônios podem estar organizados em três camadas: de entrada, escondida e de saída. O Multilayer Perceptron (MLP) é um dos tipos de RNAs mais conhecidos na literatura [Berry and Linoff 2004].
- Máquina de vetor de suporte (SVM): o algoritmo SVM é embasado na teoria de aprendizado estatístico, que permite a obtenção de classificadores com boa capacidade de generalização [Lorena and de Carvalho 2007].
- Naive bayes (NB): é um algoritmo probabilístico que é baseado no teorema de Bayes, originado na Estatística. Para que esse método funcione corretamente é necessário garantir que os dados sejam independentes entre si [Mitchell 1997].

5. Resultados dos Experimentos

Nesta seção, apresentamos os resultados experimentais obtidos nas fases de agrupamento e de classificação. É importante lembrar que a base de dados utilizada foi gerada pelo próprio protótipo, que recebeu dados de cinco sensores de temperatura e umidade. Cada sensor gera dois atributos, totalizando 10 (dez) atributos. O protótipo funcionou por 50 horas, totalizando 300 registros. Os valores de temperatura variaram de 15 a 47 graus Celcius. Já os valores de umidade variaram de 20 a 60%.

Como já foi mencionado anteriormente, utilizamos dois índices internos para avaliar as partições criadas na fase de agrupamento. A Tabela 1 mostra os valores do índice DB correspondente a cada uma das bases (partições) criadas, assim como também a média e o desvio padrão dos referidos valores. Observe que para os valores do índice DB, quanto menor for o valor gerado, melhor será a partição. Os experimentos mostraram, que de acordo com o índice DB, o algoritmo Hierárquico Aglomerativo gerou a melhor partição com 5 (cinco) grupos.

A Tabela 2 mostra os valores do índice Silhueta correspondente a cada uma das bases (partições) criadas, e também a média e o desvio padrão dos referidos valores. É importante enfatizar, que para os padrões do índice Silhueta, quanto maior for o valor, melhor será a partição. Os experimentos mais uma vez mostraram, que de acordo com o referido índice, o algoritmo Hierárquico Aglomerativo gerou a melhor partição com 5 (cinco) grupos.

Table 1. Resultados experimentais obtidos para o índice DB.

Índice DB							
Grupos (k)	**2**	**3**	**4**	**5**	**6**	**Média**	**D. Padrão**
EM	37,5822	42,1109	48,3575	45,4935	37,2736	42,1635	4,8568
HA	43,4795	39,3663	40,7793	**33,5883**	34,2792	38,2985	4,2566
k-Média01	38,3968	41,1507	48,3575	46,2688	42,6041	43,3555	3,9850
k-Média02	48,9702	38,2874	41,4610	40,7931	41,9508	42,2924	3,9902

Table 2. Resultados experimentais obtidos para o índice Silhueta.

Índice Silhueta							
Grupos (k)	**2**	**3**	**4**	**5**	**6**	**Média**	**D. Padrão**
EM	0,4756	0,3500	0,3083	0,3462	0,4024	0,3765	0,0647
HA	0,4601	0,5126	0,5069	**0,5599**	0,5310	0,5141	0,0366
k-Média01	0,4683	0,3663	0,3083	0,2487	0,2798	0,3343	0,0865
k-Média02	0,3275	0,2423	0,3409	0,2349	0,3819	0,3055	0,0643

Analisando as melhores partições, percebeu-se que houve um padrão similar de comportamento entre os grupos gerados. Devido a esse fato, decidiu-se utilizar as três melhores partições geradas por cada algoritmo de agrupamento, e não somente a melhor, de acordo com os índices de validação. Assim, as partições com 4, 5 e 6 grupos foram utilizadas nessa fase (classificação), ilustrada na Tabela 3. Na primeira linha depois do cabeçalho desta tabela, onde se lê EM_4k, na verdade significa: partição com 4 grupos gerada pelo algoritmo EM (Expectância Máxima). Ou seja, primeiro o nome do algoritmo de agrupamento, depois o parâmetro probabilístico (semente), se for o caso, e por último, o número de grupos. Como foram utilizadas as três melhores de cada algoritmo, temos um total de 12 bases (3 bases x 4 algoritmos).

Além das bases, a referida tabela ainda mostra os cinco algoritmos de classificação (colunas) utilizados nos experimentos. Vale salientar que o algoritmo k-NN foi utilizado duas vezes, com $k = 1$ e $k = 3$. Dessa forma, para cada uma das 12 bases, anotou-se o comportamento de cada classificador, usando-se para tal a acurácia (percentual de acerto). Ainda na mesma tabela, pode-se verificar a média geral da acurácia de cada classificador, assim como o desvio padrão. Note, que o classificador MLP obteve a melhor média entre os classificadores. Em segundo lugar ficou o 1k-NN com 97,01149 de média. A Árvore de Decisão (AD) obteve a pior média entre todos.

A fim de comparar o desempenho dos algoritmos de classificação, aplicou-se um teste estatístico, chamado teste de hipótese (t-test unicaudal), com um nível de confiança de 95% ($\propto = 0,05$) [Fisher 1925]. Através deste teste estatístico, foi possível analisar o desempenho de cada classificador, comparando-os entre si. Dessa forma, concluiu-se que o desempenho do melhor (MLP) não foi estatisticamente significante, se comparado com o segundo melhor (1k-NN). Na verdade, o MLP foi estatisticamente melhor apenas que o AD. Logo, qualquer um dos quatro classificadores (MLP, k-NN, SVM e NB) podem ser implementado no protótipo, para quando em funcionamento, possa vir a classificar as novas instâncias (registros) geradas pelos sensores de temperatura e umidade.

Contudo, como é de conhecimento geral, alguns classificadores são mais com-

Table 3. Resultados experimentais obtidos para os seis classificadores.

Acurácia (%) dos Classificadores						
Bases	**1k-NN**	**3k-NN**	**AD**	**MLP**	**SVM**	**NB**
EM_4k	95,1724	92,7586	95,1724	96,2069	93,1034	**97,9310**
EM_5k	97,2414	97,2414	93,7931	95,1724	87,9310	**98,9655**
EM_6k	96,2069	96,8966	94,1379	96,5517	94,8276	**99,6552**
HA_4k	98,9655	98,9655	98,9655	**99,3103**	98,9655	97,9310
HA_5k	98,9655	98,9655	97,5862	**99,3103**	98,9655	97,5862
HA_6k	98,9655	98,9655	96,8966	**99,3103**	98,9655	97,9310
kM10s_4k	96,5517	96,2069	92,7586	97,2414	**97,5862**	94,1379
kM10s_5k	96,2069	96,5517	95,1724	96,8966	**97,2414**	92,4138
kM10s_6k	96,2069	**96,8966**	94,1379	95,1724	95,5172	93,1034
kM12000s_4k	97,5862	96,2069	96,5517	97,2414	**98,2759**	94,4828
kM12000s_5k	96,2069	96,5517	91,0345	96,8966	**97,2414**	95,1724
kM12000s_6k	95,8621	95,1724	93,7911	**97,2414**	96,2069	91,7241
Média	97,01149	96,78161	94,99983	**97,21264**	96,23563	95,91953
Desvio Padrão	1,32603	1,76033	2,20807	1,45153	3,18582	2,72644

plexos que outros. Além disso, alguns são computacionalmente mais custosos que outros. Logo, para utilização em nosso protótipo, o k-NN nos parece bastante apropriado, considerando a facilidade de implementação e o tamanho inicial da base (número de instâncias e de atributos).

Depois de finalizadas as fases de agrupamento e classificação utilizadas para criar a base de treinamento e para escolher o classificador mais adequado para ser implementado, chega o momento de pensarmos nos controles que serão vinculados aos valores da classe. Dessa forma, o sistema irá classificar as novas instâncias em função do aprendizado durante o treinamento, e executará as ações vinculadas aos possíveis valores da classe (frio, normal, calor1, calor2 e perigo).

6. Conclusão e Trabalhos Futuros

A principal contribuição deste trabalho é propor um mecanismo de controle para ambientes, baseado em rede de sensores e classificação de padrões. A ideia principal é utilizar as informações geradas pelos sensores para tomar decisões relativas ao controle constante da temperatura e da umidade relativa do ar. Como caso de uso, escolhemos um dos laboratórios de análises químicas da UFRN.

A fim de provar que nosso trabalho é factível, dividimos os experimentos em duas fases. Na primeira, também chamada de fase agrupamento, usamos três algoritmos de agrupamento para gerar a base de treinamento e teste. As partições criadas foram avaliadas a partir de dois índices de validação (DB e Silhueta). Já na fase de classificação, utilizamos cinco algoritmos de classificação para avaliar seus desempenhos sobre as três melhores partições criadas por cada algoritmo de agrupamento. Depois de analisados os resultados de ambas as fases, identificamos o conjunto de ações de controle relacionado a cada valor de estado (classe) e suas variações.

Como trabalhos futuros, precisamos comparar o desempenho de nosso sistema de controle de ambientes com outros sistemas de controle que utilizam soluções clássicas.

References

Abraham, S. and Li, X. (2014). A cost-effective wireless sensor network system for indoor air quality monitoring applications. *Procedia Computer Science*, 34(0):165 – 171.

Barroca, N., Borges, L. M., Velez, F. J., Monteiro, F., Górski, M., and Castro-Gomes, J. (2013). Wireless sensor networks for temperature and humidity monitoring within concrete structures. *Construction and Building Materials*, 40(0):1156 – 1166.

Berry, M. J. and Linoff, G. S. (2004). *Data Mining Techniques: for Marketing, Sales, and Customer Relationship Mangement*. Wiley Publishing, Inc, Indianapolis, Indiana, USA, 2nd edition.

da Silva, L. T. (2013). Uma proposta de automação residencial utilizando a plataforma de prototipagem eletrônica arduino. *Revista Tecnologias em Projeção*.

Faceli, K., Lorena, A. C., Gama, J., and de Carvalho, A. C. (2011). *Inteligência Artificial: uma Abordagem de Aprendizado de Máquina*. LTC - Livros técnicos e científicos Editora Ltda.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver & Boyd.

Halkidi, M., Batistakis, Y., and Vazirgiannis, M. (2002). Clustering validity checking methods: part ii. *SIGMOD Rec.*, 31 (3):19–27.

Lorena, A. C. and de Carvalho, A. C. P. L. F. (2007). Uma introdução às support vector machines. *Revista de Informática Teórica e Aplicada (RITA)*, 14(2):43 – 67.

Margolis, M. (2011). *Arduino Cookbook*. O'Reilly Media, USA, 2nd edition.

Maycon J. C. Mesquita, Jose C. G. Neto, W. R. M. A. (2014). Sistema de controle e supervisão residencial utilizando plataforma arduino.

McRoberts, M. (2011). *Beginning Arduino*. Apress, USA, 1st edition.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill, Inc., New York, NY, USA, 1st edition.

Vera-Repullo, J., Ruiz-Peñalver, L., Jiménez-Buendía, M., Rosillo, J., and Molina-Martínez, J. (2015). Software for the automatic control of irrigation using weighing-drainage lysimeters. *Agricultural Water Management*, 151(0):4 – 12. New proposals in the automation and remote control of water management in agriculture: agromotic systems.

Xu, R. and Wunsch, II, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3):645–678.

Yuwono, M., Guo, Y., Wall, J., Li, J., West, S., Platt, G., and Su, S. W. (2015). Unsupervised feature selection using swarm intelligence and consensus clustering for automatic fault detection and diagnosis in heating ventilation and air conditioning systems. *Applied Soft Computing*, 34(0):402 – 425.