

# Mineração de dados aplicada a *dataset* de partidas de *kendô*

Guilherme Brizzi<sup>1</sup>, Mathias Eckert Recktenvald<sup>1</sup>

<sup>1</sup>Curso de Ciência da Computação –  
Universidade Federal de Santa Maria (UFSM)

## 1. Introdução

O Kendô é uma arte marcial japonesa que surgiu quando o país passava por um rápido processo de modernização. Assim como o futuro de uma nação em transformação é difícil de prever, o resultado de um duelo de Kendô também pode ser difícil de prever.

Contudo, assim como os governantes das nações podem usar estatísticas para realizar decisões bem informadas, os praticantes do Kendô podem analisar grandes números de partidas para obter padrões e compreender como melhorar seu desempenho.

Dessa forma, este trabalho objetiva usar o processo de KDD (*Knowledge Discovery from Data*) sobre um *dataset* descrevendo partidas de Kendô. Buscou-se responder às perguntas originais da proposta do trabalho, assim como obter demais conhecimento útil relativo a essa arte marcial.

## 2. Metodologia e Desenvolvimento

### 2.1. Pré-processamento

Iniciou-se o KDD sobre o *dataset* a partir do pré-processamento dos dados. Essa fase visou principalmente à limpeza e à organização dos dados em um formato mais próximo ao utilizado pelo algoritmo de *Apriori*.

O *dataset* original contém 202 entradas, cada uma com os seguintes campos:

**Tabela 1. Descrição do conjunto de dados de ippons**

| Coluna               | Descrição   |
|----------------------|---|
| [índice não nomeado] | Contador geral.   |
| match_id             | Identificador da luta (cada número representa uma luta distinta).   |
| ippon_number         | Número sequencial do ippon dentro da luta.  |
| seconds_between      | Tempo (em segundos) desde o início da rodada até a ocorrência do golpe.   |
| ippon_taken          | Indica se o ippon foi sofrido (1) ou não (0). Ou seja, o adversário executou um golpe que valeu ponto para ele, não para o atleta cuja perspectiva é vista. |
| men                  | Indica se o golpe marcado foi um <i>men</i> (cabeça).   |
| kote                 | Indica se o golpe marcado foi um <i>kote</i> (antebraço).   |
| do                   | Indica se o golpe marcado foi um <i>dō</i> (tronco).  |
| tsuki                | Indica se o golpe marcado foi um <i>tsuki</i> (garganta).   |

Entretanto, nem todos desses atributos são úteis para o processo de mineração de dados. Índices, via de regra, tendem a não oferecer quaisquer dados úteis. Portanto, iniciou-se a limpeza retirando os atributos de índice ou assemelhados.

Os campos `match_id`, `ippon_number` assim como o [índice não nomeado] foram, portanto, retirados, pois atuam meramente para indexação. Para tal, mesclaram-se as duas entradas de cada partida<sup>1</sup> - ou seja, as entradas de mesmo `match_id` e `ippon_number = 1` e `ippon_number = 2`.

Ademais, realizou-se a limpeza de registros claramente ruidosos, os quais foram descartados. Verificou-se se apenas um golpe estava marcado, se o tempo era razoável e se os valores tinham tipos válidos.

Com as entradas ruidosas removidas, ficou-se com 198 registros - os quais, após a mescla descrita anteriormente, resultaram em 99 linhas de dados úteis, dispostos com a seguinte estrutura:

**Tabela 2. Descrição do conjunto de dados após limpeza e reorganização**

| Coluna                         | Descrição  |
|--------------------------------|--|
| <code>seconds_between_1</code> | Tempo (em segundos) desde o início da rodada até a ocorrência do <b>primeiro golpe</b> . |
| <code>ippon_taken_1</code>     | Indica se o <b>primeiro ippon</b> foi sofrido (1) ou não (0).                            |
| <code>men_1</code>             | Indica se o <b>primeiro golpe</b> marcado foi um <i>men</i> (cabeça).                    |
| <code>kote_1</code>            | Indica se o <b>primeiro golpe</b> marcado foi um <i>kote</i> (antebraço).                |
| <code>do_1</code>              | Indica se o <b>primeiro golpe</b> marcado foi um <i>dō</i> (tronco).                     |
| <code>tsuki_1</code>           | Indica se o <b>primeiro golpe</b> marcado foi um <i>tsuki</i> (garganta).                |
| <code>seconds_between_2</code> | Tempo (em segundos) desde o início da rodada até a ocorrência do <b>segundo golpe</b> .  |
| <code>ippon_taken_2</code>     | Indica se o <b>segundo ippon</b> foi sofrido (1) ou não (0).                             |
| <code>men_2</code>             | Indica se o <b>segundo golpe</b> marcado foi um <i>men</i> (cabeça).                     |
| <code>kote_2</code>            | Indica se o <b>segundo golpe</b> marcado foi um <i>kote</i> (antebraço).                 |
| <code>do_2</code>              | Indica se o <b>segundo golpe</b> marcado foi um <i>dō</i> (tronco).                      |
| <code>tsuki_2</code>           | Indica se o <b>segundo golpe</b> marcado foi um <i>tsuki</i> (garganta).                 |

## 2.2. Aplicação do *Apriori* e obtenção de regras

Desenvolveu-se um *script* em Python para aplicação do algoritmo *Apriori* e obtenção das regras de associação usando a biblioteca *mlxtend*. Este algoritmo está anexado em `/code/main.py`. Para executá-lo, basta rodar o comando `python main.py`. A saída estará no arquivo `output.txt`.

Foram criadas funções para obter informações visando a responder cada questão. Outrossim, também foram programadas funções adicionais buscando encontrar padrões não contemplados pelas perguntas originais. Na próxima subseção, expor-se-ão os resultados obtidos.

<sup>1</sup>Excetuando-se entradas ruidosas, cada partida tem apenas duas rodadas.

### 2.3. Interpretação e respostas às questões

#### 2.3.1. Qual golpe, em regra, com mais de 90% de confiança, implica o não uso de outro golpe?

Encontrou-se que o golpe que, usado na primeira rodada, implica o não uso de outro golpe na segunda é o *men* que implica no não uso do *tsuki*. Essa regra possui confiança de 93.47% e suporte de 43.43%. Existem outras regras com confiança maior que 90% para esse quesito, porém seu suporte é menos da metade da regra  $men \Rightarrow \neg tsuki$ .

**Tabela 3. Regras relacionando golpes ao não uso de outros golpes**

| Antecedent | Consequent | ASup  | CSup  | Sup   | Conf  | Lift  |
|------------|------------|-------|-------|-------|-------|-------|
| (men_1)    | (tsuki_2)  | 0.465 | 0.919 | 0.434 | 0.935 | 1.017 |
| (do_1)     | (do_2)     | 0.162 | 0.828 | 0.162 | 1.000 | 1.207 |
| (do_1)     | (tsuki_2)  | 0.162 | 0.919 | 0.152 | 0.938 | 1.020 |

#### 2.3.2. Qual a confiança da regra: um atleta que aplicou um *men* levou um *ippon*?

Obteve-se um valor de confiança de 45.16% para atletas que aplicaram um *men* levarem um *ippon*.

**Tabela 4. Regras relacionando golpe *men* a resultados**

| Antecedent | Consequent    | ASup  | CSup  | Sup   | Conf  | Lift  |
|------------|---------------|-------|-------|-------|-------|-------|
| (men)      | (ippon_taken) | 0.470 | 0.439 | 0.212 | 0.452 | 1.028 |

#### 2.3.3. Qual(is) golpe(s), em regra, com mais de 85% de confiança, tende a demorar mais de 25 segundos após o primeiro golpe?

O golpe em questão é o *tsuki*, que, quando ocorrido na segunda rodada, demora mais de 25 segundos com 87.5% de confiança.

**Tabela 5. Regras relacionando o golpes da segunda rodada a tempo de execução maior que 25 segundos**

| Antecedent | Consequent | ASup  | CSup  | Sup   | Conf  | Lift  |
|------------|------------|-------|-------|-------|-------|-------|
| (tsuki_2)  | (over25)   | 0.081 | 0.747 | 0.071 | 0.875 | 1.171 |

#### 2.3.4. Se o primeiro golpe demorou mais do que 15 segundos, e o segundo golpe demorou mais do que 25 segundos. Qual golpe provavelmente não foi usado com sucesso? Mostre a regra que suporta sua resposta.

Provavelmente se trata de um *men*. A falha na aplicação desse golpe se mostrou como a regra mais forte, conforme disposto a seguir:

**Tabela 6. Regras relacionando ataques lentos a ataques falhos**

| Antecedent     | Consequent    | ASup  | CSup  | Sup   | Conf  | Lift  |
|----------------|---------------|-------|-------|-------|-------|-------|
| (slow_attacks) | (failed_men)  | 0.495 | 0.212 | 0.111 | 0.224 | 1.058 |
| (slow_attacks) | (failed_kote) | 0.495 | 0.157 | 0.061 | 0.122 | 0.782 |
| (slow_attacks) | (failed_do)   | 0.495 | 0.061 | 0.040 | 0.082 | 1.347 |

### 2.3.5. Padrões adicionais

O processo de mineração revela um padrão geral interessante da amostra de partidas de Kendô: via de regra, perder a primeira rodada implica na vitória na segunda.

Com 65.95% de confiança, o segundo ataque acerta se o primeiro falha. A seguinte regra descreve essa relação:

**Tabela 7. Relação entre *ippon\_taken\_1* e *not\_ippon\_taken\_2***

| Antecedent      | Consequent          | ASup  | CSup  | Sup   | Conf  | Lift  |
|-----------------|---------------------|-------|-------|-------|-------|-------|
| (ippon_taken_1) | (not_ippon_taken_2) | 0.475 | 0.596 | 0.313 | 0.660 | 1.107 |

Esse padrão é curioso, visto que, de modo geral, esperamos que um atleta derrotado seja derrotado novamente. É comum prever a performance de um clube, seleção ou atleta pelo seu desempenho passado - tanto historicamente, quanto dentro da partida.

Tal faceta, portanto, pode ser interpretada como uma característica própria do Kendô: é possível que a luta permita revanches e recuperações mais do que outros esportes. Contudo, é possível que se trate tão somente de um indicador da artificialidade do *dataset* em questão, ao passo que seu conteúdo reflete o oposto de regras que seriam esperadas para dados desportivos.

## 3. Conclusão

Nesse sentido, a aplicação do processo de KDD ao conjunto de dados de Kendô possibilitou a identificação de regras de associação relevantes, revelando padrões inesperados nas partidas analisadas. Apesar das limitações do *dataset*, os resultados obtidos mostram como a mineração de dados pode auxiliar na compreensão de dinâmicas presentes em artes marciais.