**Como definir seu problema de aprendizado de máquina**

O primeiro passo em qualquer projeto é definir seu problema. Você pode usar os algoritmos mais poderosos e brilhantes disponíveis, mas os resultados não terão sentido se você estiver resolvendo o problema errado.

Neste post você aprenderá o processo para pensar profundamente sobre o seu problema antes de começar. Este é, indiscutivelmente, o aspecto mais importante da aplicação do aprendizado de máquina.

**Quadro de definição de problemas**

Uso uma estrutura simples ao definir um novo problema para resolver com aprendizado de máquina. A estrutura me ajuda a entender rapidamente os elementos e motivação para o problema e se o aprendizado de máquina é adequado ou não.

O quadro envolve responder a três perguntas em diferentes graus de minuciosidade:

* **Passo 1**: Qual é o problema?
* **Passo 2**: Por que o problema precisa ser resolvido?
* **Passo 3**: Como eu resolveria o problema?

**Passo 1: Qual é o problema**

O primeiro passo é definir o problema. Eu uso uma série de táticas para coletar essas informações.

**Descrição informal**

Descreva o problema como se estivesse descrevendo para um amigo ou colega. Isso pode fornecer um ótimo ponto de partida para destacar áreas que você pode precisar preencher. Ele também fornece a base para uma descrição de uma frase que você pode usar para compartilhar sua compreensão do problema.

Por exemplo: *eu preciso de um programa que me diga quais tweets receberão retweets.*

**Formalismo**

Em um post anterior [do blog definindo aprendizado de máquina](https://machinelearningmastery.com/what-is-machine-learning/) você aprendeu sobre o formalismo de aprendizado de máquina de Tom Mitchell. Aqui está novamente para refrescar sua memória.

*Diz-se que um programa de computador aprende com a experiência E em relação a alguma classe de tarefas T e medida de desempenho P, se seu desempenho em tarefas em T, medido por P, melhora com a experiência E.*

Use este formalismo para definir o *T*, *P* e *E* para o seu problema.

Por exemplo:

* **Tarefa** (*T*): Classifique um tweet que não tenha sido publicado como indo para receber retweets ou não.
* **Experiência** (*E*): Um corpus de tweets para uma conta onde alguns têm retweets e outros não.
* **Desempenho** (*P*): Precisão de classificação, o número de tweets previstos corretamente de todos os tweets considerados como uma porcentagem.

**Suposições**

Crie uma lista de suposições sobre o problema e está formulando. Estas podem ser regras de polegar e informações específicas de domínio que você acha que o levarão a uma solução viável mais rapidamente.

Pode ser útil destacar questões que podem ser testadas contra dados reais porque avanços e inovação ocorrem quando suposições e melhores práticas são demonstradas como erradas diante de dados reais. Também pode ser útil para destacar áreas da especificação do problema que podem precisar ser desafiadas, relaxadas ou apertadas.

Por exemplo:

* As palavras específicas usadas no tweet importam para o modelo.
* O usuário específico que retweets não importa para o modelo.
* O número de retweets pode importar para o modelo.
* Tweets mais antigos são menos preditivos do que tweets mais recentes.

**Problemas semelhantes**

Que outros problemas você já viu ou pode pensar que são como o problema que você está tentando resolver? Outros problemas podem informar o problema que você está tentando resolver, destacando limitações em sua frase do problema, como dimensões do tempo e deriva conceitual (onde o conceito que está sendo modelado muda ao longo do tempo). Outros problemas também podem apontar algoritmos e transformações de dados que poderiam ser adotados para detectar o desempenho da verificação.

Por exemplo: Um problema relacionado seria a discriminação por spam de e-mail que usa mensagens de texto como dados de entrada e precisa de decisão de classificação binária.

**Passo 2: Por que o problema precisa ser resolvido?**

O segundo passo é pensar profundamente sobre por que você quer ou precisa resolver o problema.

**Motivação**

Considere sua motivação para resolver o problema. Que necessidade será cumprida quando o problema for resolvido?

Por exemplo, você pode estar resolvendo o problema como um exercício de aprendizagem. Isso é útil para esclarecer, pois você pode decidir que não quer usar o método mais adequado para resolver o problema, mas em vez disso você quer explorar métodos que você não está familiarizado para aprender novas habilidades.

Alternativamente, você pode precisar resolver o problema como parte de um dever no trabalho, em última análise para manter o seu emprego.

**Benefícios da solução**

Considere os benefícios de ter o problema resolvido. Quais recursos ele habilita?

É importante deixar claro sobre os benefícios do problema que está sendo resolvido para garantir que você capitalize sobre eles. Esses benefícios podem ser usados para vender o projeto aos colegas e à gestão para obter compra e recursos adicionais de tempo ou orçamento.

Se isso te beneficia pessoalmente, então fique claro sobre quais são esses benefícios e como você saberá quando os tiver. Por exemplo, se é uma ferramenta ou utilidade, então o que você será capaz de fazer com esse utilitário que você não pode fazer agora e por que isso é significativo para você?

**Uso da solução**

Considere como a solução para o problema será usada e que tipo de vida você espera que a solução tenha. Como programadores, muitas vezes pensamos que o trabalho é feito assim que o programa é escrito, mas na verdade o projeto está apenas começando sua vida útil de manutenção.

A forma como a solução será usada influenciará a natureza e os requisitos da solução que você adota.

Considere se você está procurando escrever um relatório para apresentar resultados ou se deseja operacionalizar a solução. Se você quiser operacionalizar a solução, considere os requisitos funcionais e não funcionais que você tem para uma solução, assim como um projeto de software.

**Passo 3: Como eu resolveria o problema?**

Nesta terceira e última etapa da definição do problema, explore como você resolveria o problema manualmente.

Liste passo a passo quais dados você coletaria, como você os prepararia e como você projetaria um programa para resolver o problema. Isso pode incluir protótipos e experimentos que você precisaria realizar que são uma mina de ouro, porque eles destacarão perguntas e incertezas que você tem sobre o domínio que poderia ser explorado.

Esta é uma ferramenta poderosa. Ele pode destacar problemas que realmente podem ser resolvidos satisfatoriamente usando uma solução implementada manualmente. Ele também libera importantes conhecimentos de domínio que foram presos até agora, como onde os dados são realmente armazenados, quais tipos de recursos seriam úteis e muitos outros detalhes.

Colete todos esses detalhes à medida que eles ocorrem a você e atualize as seções anteriores da definição do problema. Especialmente as suposições e regras do polegar.

Consideramos uma solução especificada manualmente antes ao descrever problemas complexos no [porquê o aprendizado de máquina importa](https://machinelearningmastery.com/machine-learning-matters/).

**Resumo**

Neste post você aprendeu o valor de ser claro sobre o problema que você está resolvendo. Você descobriu uma estrutura de três passos para definir seu problema com táticas práticas em etapas:

* **Passo 1: Qual é o problema?** Descrever o problema informalmente e formalmente e listar suposições e problemas semelhantes.
* **Passo 2: Por que o problema precisa ser resolvido?** Liste sua motivação para resolver o problema, os benefícios que uma solução proporciona e como a solução será usada.
* **Passo 3: Como eu resolveria o problema?** Descreva como o problema seria resolvido manualmente para liberar o conhecimento do domínio.