O Papel da Estatística em outras áreas

Erica Castilho Rodrigues

04 de Maio de 2017

Site: http://www.iceb.ufop.br/deest/
e-mail: ericacastirodrigues@gmail.com

Aplicações

Distribuição de Pontos

- 3 provas - 30 pontos cada;
- Listas - 10 pontos.

Livro Texto

- ESTATISTICA APLICADA E PROBABILIDADE PARA ENGENHEIROS (Quarta edição) - MONTGOMERY, DOUGLAS C. e RUNGER, GEORGE C.

Objetivos

Ao final deste capítulo você deve ser capaz de:

- Identificar o papel da estatística na resolução de problemas.
- Discutir como a variabilidade afeta os dados coletados e usados para tomada de decisão.
- Diferenciar estudos enumerativos e analíticos.
- Discutir métodos de coletas de dados.

- Identificar as vantagens de experimentos planejados.
- Diferenciar modelos mecanicistas de modelos empíricos.
- Discutir como probabilidade e modelos de probabilidade são usados em engenharia e em ciência.

O que é Estatística?

- Antigamente: originada do termo em latin *status*.
- Se referia apenas às informações sobre os estados.

Nos dias de hoje...

- **Estatística** se refere a todo tipo de informação:
  - medidas de temperatura;
  - satisfação do consumidor;
  - casos de doença.
- A **Estatística** analisa e interpreta esses dados.

- Desde a antiguidade, os governos têm se interessado por informações sobre:
  - populações e riquezas.
- Existem relatos de levantamentos feitos na China;
  - há mais de 2000 anos A. C.
- Pesquisas arqueológicas mostram que:
  - os faraós fizeram uso sistemático de informações de caráter estatístico.

- No século XVI surgiram as primeiras análises sistemáticas de fatos como:
  - batizados, casamentos e funerais.
- Originaram as primeiras tabelas e os primeiros números relativos.
- No século XVIII tais análises adiquiriram aspecto verdadeiramente científico.
- As tabelas se tornaram mais completas.
- Surgiram representações gráficas e os cálculos de probabilidade.
- Verificou-se que a estatística poderia ser usada para:
  - tomar decisões com base na análise de dados.

- A estatística fornece métodos para:
  - coleta,
  - organização,
  - descrição,
  - análise,
  - interpretação de dados.
- Permite canalisar as informações para um objetivo.

O Pensamento Estatístico

- Objetivo do engenheiro ou de um pesquisador:
  - resolver problemas de interesse da sociedade através da aplicação do método científico.

- Método de engenharia ou método científico $\Rightarrow$ abordagem para formular e resolver problemas.

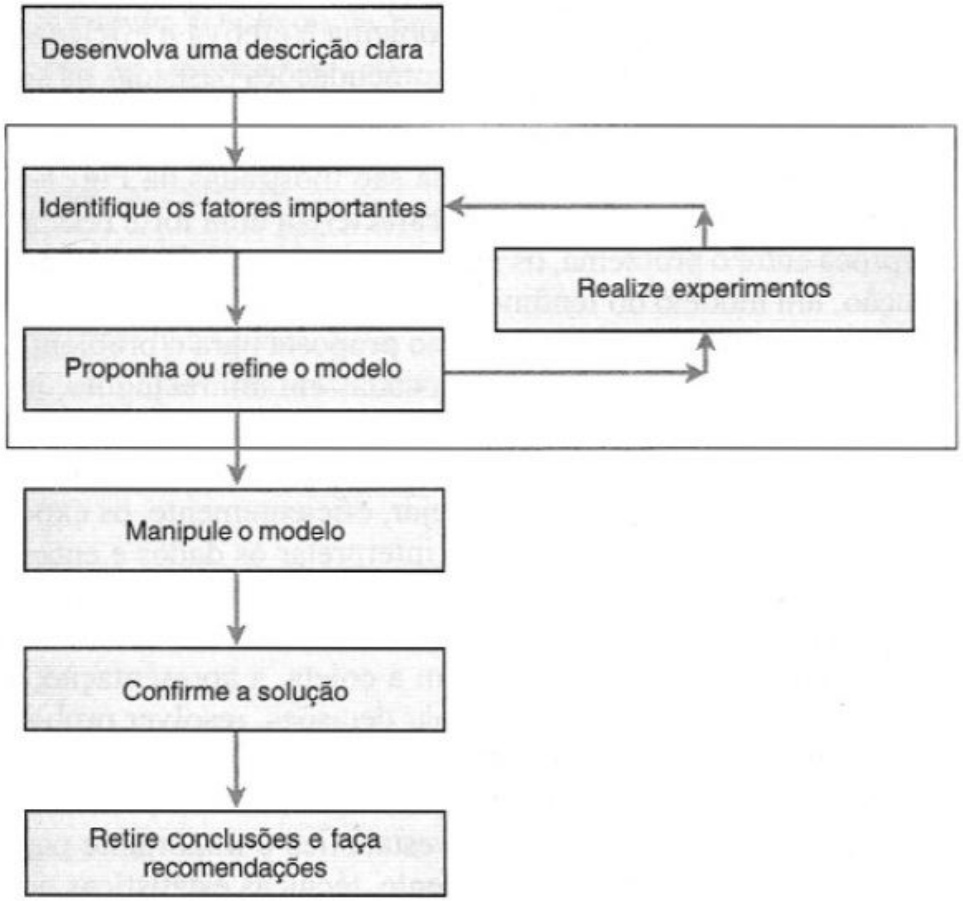


Fig. 1.1 O método de solução de um problema.

O que engenheiros e pesquisadores devem saber:

- planejar eficientemente experimentos;
- coletar dados;
- analisá-los e interpretá-los;
- entender como os dados estão relacionados com o modelo proposto em estudo.

Estatística lida com:

- coleta, apresentação e análise de dados;
- uso desses dados para tomar decisões e resolver problemas.

- O conhecimento de estatística é importante para qualquer profissional.
- Técnicas estatísticas podem ser uma ajuda poderosa no planejamento de novos produtos e sistemas.
- Podem ser usadas melhorando projetos existentes e planejando, desenvolvendo e melhorando processos de produção.
- Pondem ser usadas para melhorar métodos de avaliação dos alunos. (TRI)

- Métodos estatísticos nos ajudam a entender a variabilidade.
- Variabilidade $\Rightarrow$ sucessivas observações de um sistema não produzem o mesmo resultado.
- Um mesmo aluno faz duas provas com mesmo nível de dificuldade e obtém resultados distintos.
- Pensamento estatístico $\Rightarrow$ nos ensina a incorporar essa variabilidade na tomada de decisão.

Exemplo: Desempenho do consumo de gasolina de um carro

- O desempenho não é o mesmo para cada tanque de combustível.
- Depende de fatores como:
  - tipo de estrada usada;
  - mudanças nas condições do veículo;
  - marca da gasolina;
  - condições climáticas.

- São fatores potenciais de variabilidade.
- Estatística nos fornece uma estrutura para
  - descrever a variabilidade;
  - aprender quais fontes são mais importantes.

Exemplo: projeto de um conector de náilon para aplicação automotiva

- Especificação da espessura da parede: 3/32 polegadas.
- Está inseguro quanto a força de remoção.
- Produz oito unidades do protótipo e mede suas forças de remoção.
- Valores observados: 12,9; 13,4; 12,3; 13,6; 13,5; 12,6; 13,1.
- Nem todos protótipos tem a mesma força de remoção.

- Consideramos a força de remoção como uma variável aleatória X.
- Podemos pensar em X como

$$X = \mu + \epsilon$$

  onde μ é uma constante e ϵ é uma perturbação aleatória.
- A constante permanece a mesma em cada medida.
- O valor de ϵ pode mudar, devido a
  - mudanças no ambiente;
  - diferenças na peças individuais;

- O diagrama de pontos permite visualizar localização e dispersão dos dados.

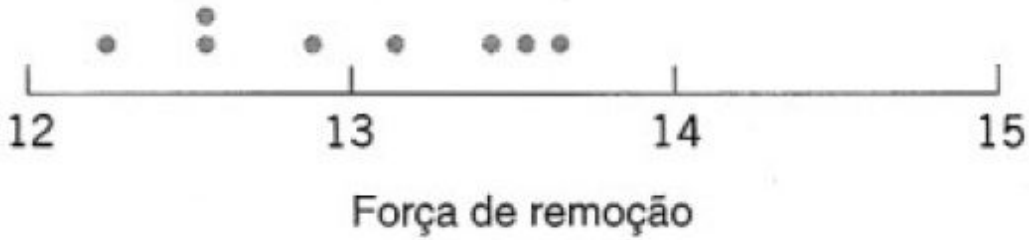


Fig. 1.2 Diagrama de pontos dos dados da força de remoção, quando a espessura da parede for 3/32 polegada.

Inferência Estatística

- raciocinar a partir de um conjunto de medidas para casos mais gerais.

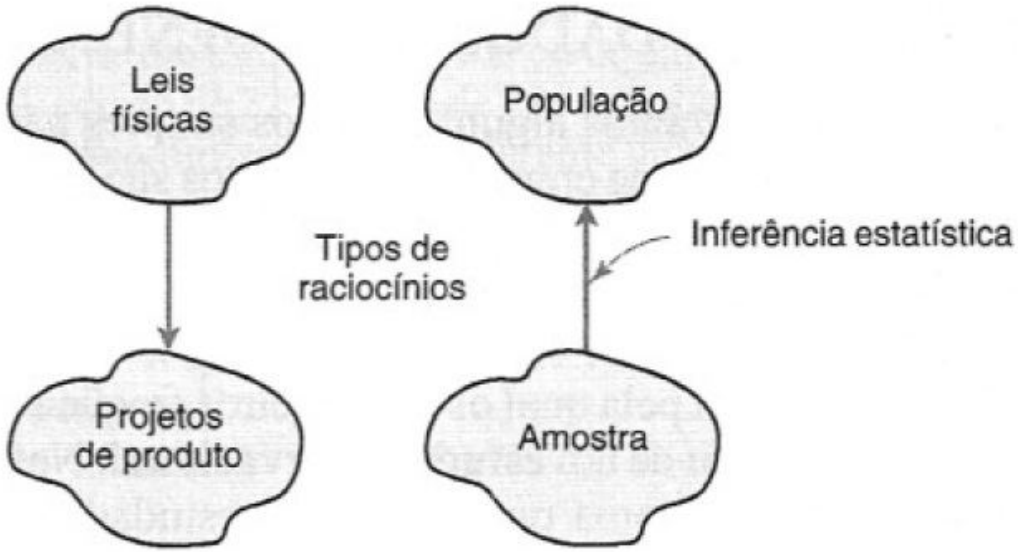


Fig. 1.6 Inferência estatística é um tipo de raciocínio.

- O raciocínio baseado em alguns objetos pode resultar em erros:
  - erros amostrais.
- Se a amostra é selecionada adequadamente,
  - esses riscos podem ser quantificados e um tamanho apropriado de amostra é determinado.

Coleta de dados

Três métodos básicos de coletar dados são:

- estudo **retrospectivo** usando dados históricos;
- estudo de **observação**;
- um **experimento planejado**.

Estudo Retrospectivo

- Utiliza dados arquivados.
- Esses dados podem conter informação relativamente de pouca utilidade para o problema.
- Alguns dados relevantes podem ser omitidos.
- Podem haver erros de transcrição ou registro, resultando em *outliers* (valores não-usuais).

Exemplo:

- Deseja-se medir a concentração de acetona em um destilado.
- Fatores que afetam:
  - temperatura do refervedor,
  - temperatura do condensado,
  - taxa de refluxo.
- Uma amostra de dados arquivados é utilizada.
- Pode-se estar interessado em descobrir relações entre as duas temperaturas e a taxa de refluxo sobre a concentração de acetona.

Estudo de observação

- Observa-se o processo perturbando-o tão pouco quanto possível.
- Ocorrem por um curto período de tempo.
- Assim, variáveis que não são rotineiramente medidas podem ser incluídas.

No exemplo de destilação:

- O engenheiro poderia planejar uma forma de medir as temperaturas e a taxa de refluxo quando as medidas de concentração são feitas.
- Poderia ser possível também medir concentrações da corrente de alimentação e estudar seu impacto.

Experimentos planejados

- O engenheiro faz variações deliberadas em variáveis controláveis do sistema.
- Observa os dados de saída resultante.
- Faz uma inferência sobre quais variáveis são responsáveis pelas mudanças observadas.

Exemplo:

- No exemplo do conector de náilon.
- O engenheiro escolhe duas espessuras da parede e mede a força de remoção para cada uma delas.
- Nesse tipo de **experimento comparativo** o interesse é em determinar se existe diferença entre os dois projetos.

Observando Processos ao Longo do Tempo

- Dados coletados ao longo do tempo.
- É útil plotar os dados versus tempo em um gráfico de **série temporal**.
- Nesse tipo de gráfico, fenômenos que possam afetar o sistema se tornam mais visíveis.
- O conceito de estabilidade pode ser mais bem julgado.

Exemplo:

- No exemplo da destilação.
- Poderíamos fazer um gráfico de série temporal da concentração de acetona como esse mostrado abaixo.

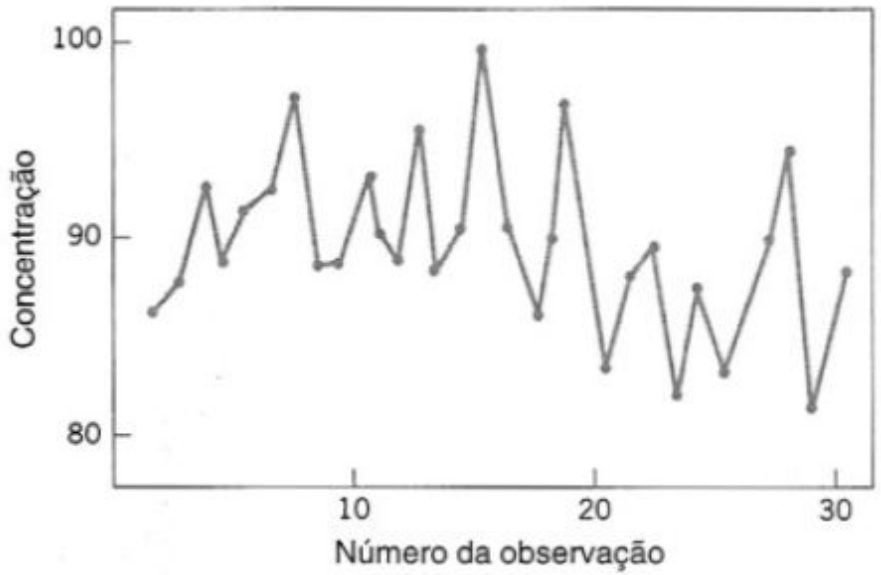


Fig. 1.15 Um gráfico temporal de concentração provê mais informações do que o diagrama de pontos.

Probabilidade e Modelos de Probabilidade

- Vimos que as decisões são tomadas tendo como base os objetos de uma amostra.
- Precisamos saber o quão bem essa amostra representa a população.
- Devemos quantificar os riscos de decisões tomadas tendo como base a amostra.
- Os **modelos de probabilidade** ajudam a quantificar esses riscos.

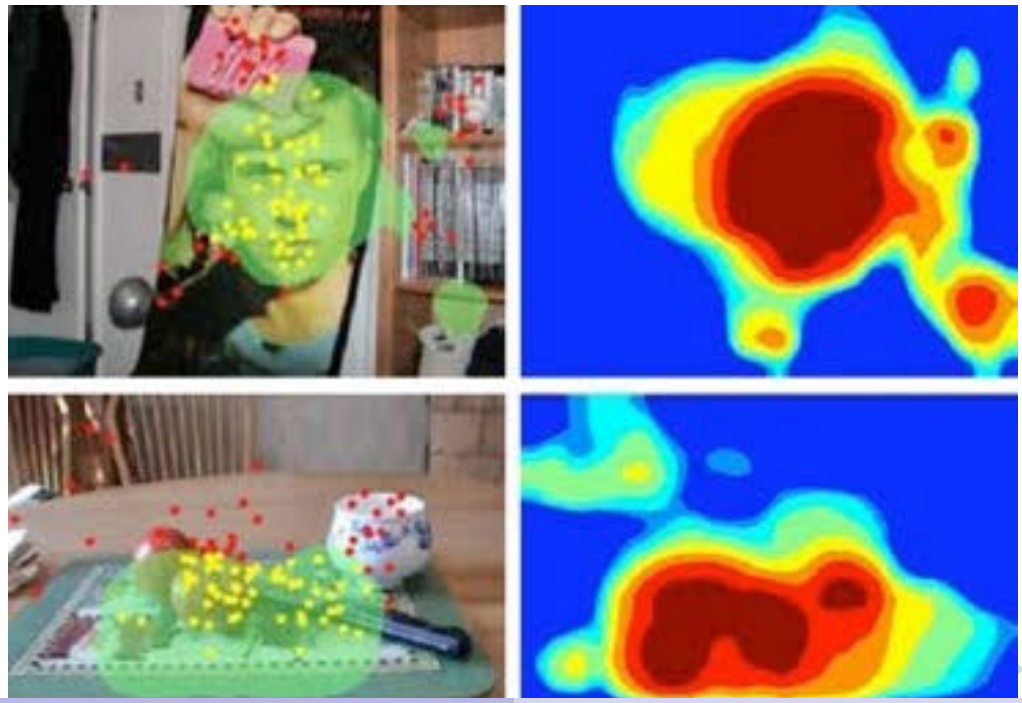
- Suponha que temos um lote com 25 pastilhas.
- Se todas forem defeituosas ou todas boas:
  - as pastilhas da amostra serão todas defeituosas ou todas boas.
- Suponha que só uma pastilha é defeituosa.
- Uma amostra retirada pode ou não conter essa pastilha defeituosa.
- Para quantificarmos os riscos de uma pastilha defeituosa ser detectada devemos usar:
  - um modelo de probabilidade
  - uma método para selecionar a amostra.

Outras aplicações Interessantes...

What and where: A Bayesian inference theory of attention

- Como o cérebro humano reconhece objetos?
- Pesquisadores do Institute for Brain Research (MIT) desenvolveram um modelo matemático para descrever como o cérebro identifica objetos visualmente.
- O modelo é capaz de predizer como o cérebro reage a certas tarefas de percepção.
- É também capaz de melhorar sistemas de reconhecimento automáticos.

- O objetivo: inferir o que é e onde está um objeto em uma cena.
- Essas duas tarefas são executadas em partes distintas do cérebro.
- Porém são integradas durante a análise da imagem.
- A questão é:
  - como essas duas tarefas são combinadas.



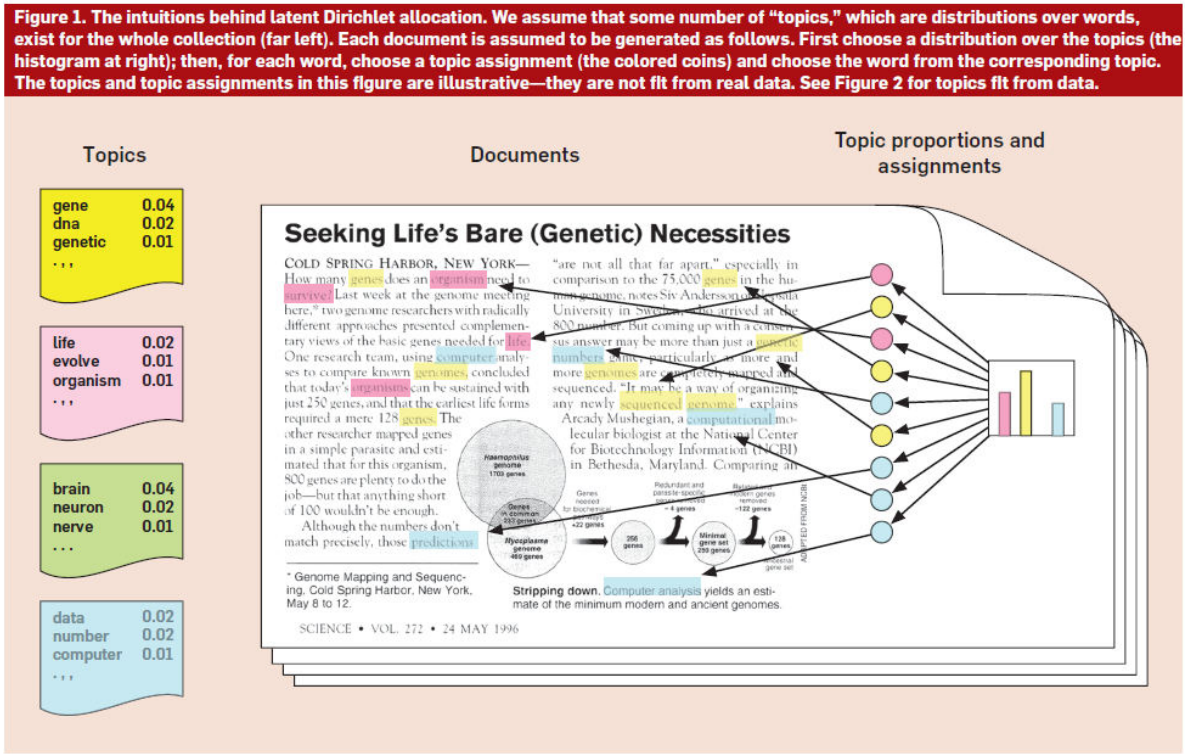
- Coloca uma distribuição *a priori* para a descrição da cena.
- Usa a informação a *priori* que tem sobre o espaço:
  - iluminação global,
  - identidade da cena,
  - objetos presentes.
- Observa os píxels da imagem.
- A partir dessas duas informações consegue inferir a probabilidade de cada objeto e sua posição.


Free viewing
(uniform priors)
Search for cars and pedestrians
(learned priors)
car search
pedestrian search
Fixations
Model posterior
Fixations
Model posterior

Surveying a suite of algorithms that offer a solution to managing large document archives.

- Como resumir a informação contida em vários textos?
- Sites de internet, blogs, redes sociais possuem um volume enorme de informação.
- Precisamos de um método automático para organizar, procurar e entender esse grande volume de informação.
- Podemos ter um método automático de inferir os assuntos abordados em um texto.

- Cada tipo de tópico tem um conjunto de palavras relacionadas.
- De acordo com as palavras que aparecem no texto podemos inferir os tópicos mais prováveis.
- Obtemos uma distribuição de probabilidades para os tópicos do texto - **Probabilistic Topic Models**.



- Essa metodologia pode ser aplicada para grandes coleções de documentos.
- É possível recuperar informações sobre documentos em tempo real.
- Podem ser adaptados para vários tipos de dados.
- São usados em:
  - identificação de padrões em dados genéticos;
  - análise de imagens;
  - análise de redes sociais.

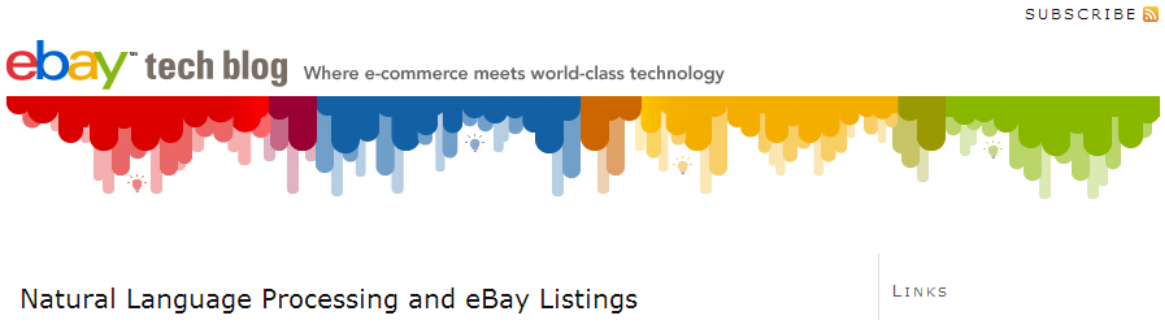
Named-entity recognition

- O objetivo é classificar os elementos de um texto em:
  - pessoa;
  - organizações;
  - locais;
  - valores monetários.
- Temos uma frase como essa:

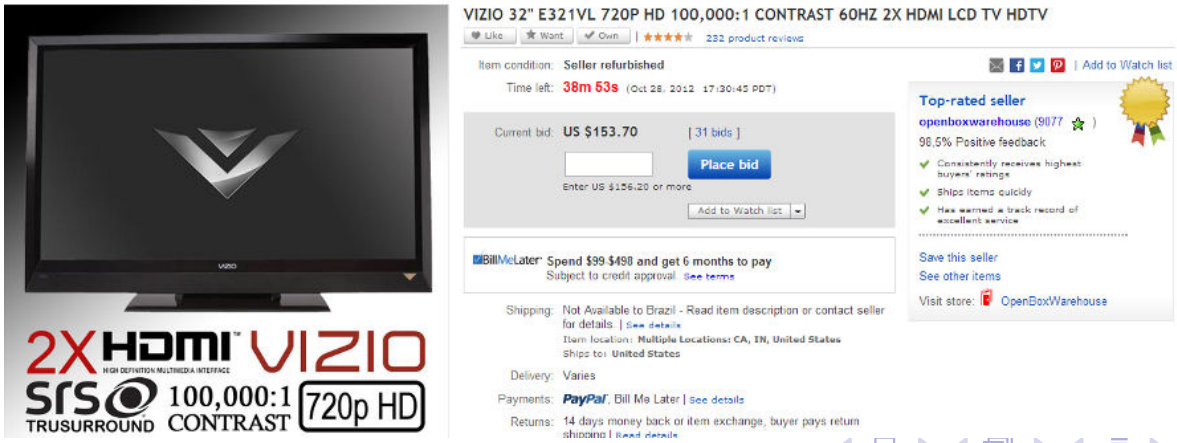
  Jim comprou 300 ações da Acme Corp em 2006.

  queremos obter automaticamente a seguinte classificação
  - Jim - pessoa;
  - 300 - quantidade;
  - Acme Corp.- organização;
  - 2006 - data.

- As palavras são classificadas de acordo com suas características:
  - se começam com letra maiúscula;
  - qual tipo de palavras que aparecem antes e depois;
  - se estão no início ou final da frase.
- Estima-se a probabilidade de cada palavra pertencer a cada uma das classes.



- O site *ebay* utiliza essa metodologia.
- O usuário pode entrar com qualquer tipo de texto descrevendo o produto vendido.
- Informações relevantes sobre o produto são extraídas do texto.

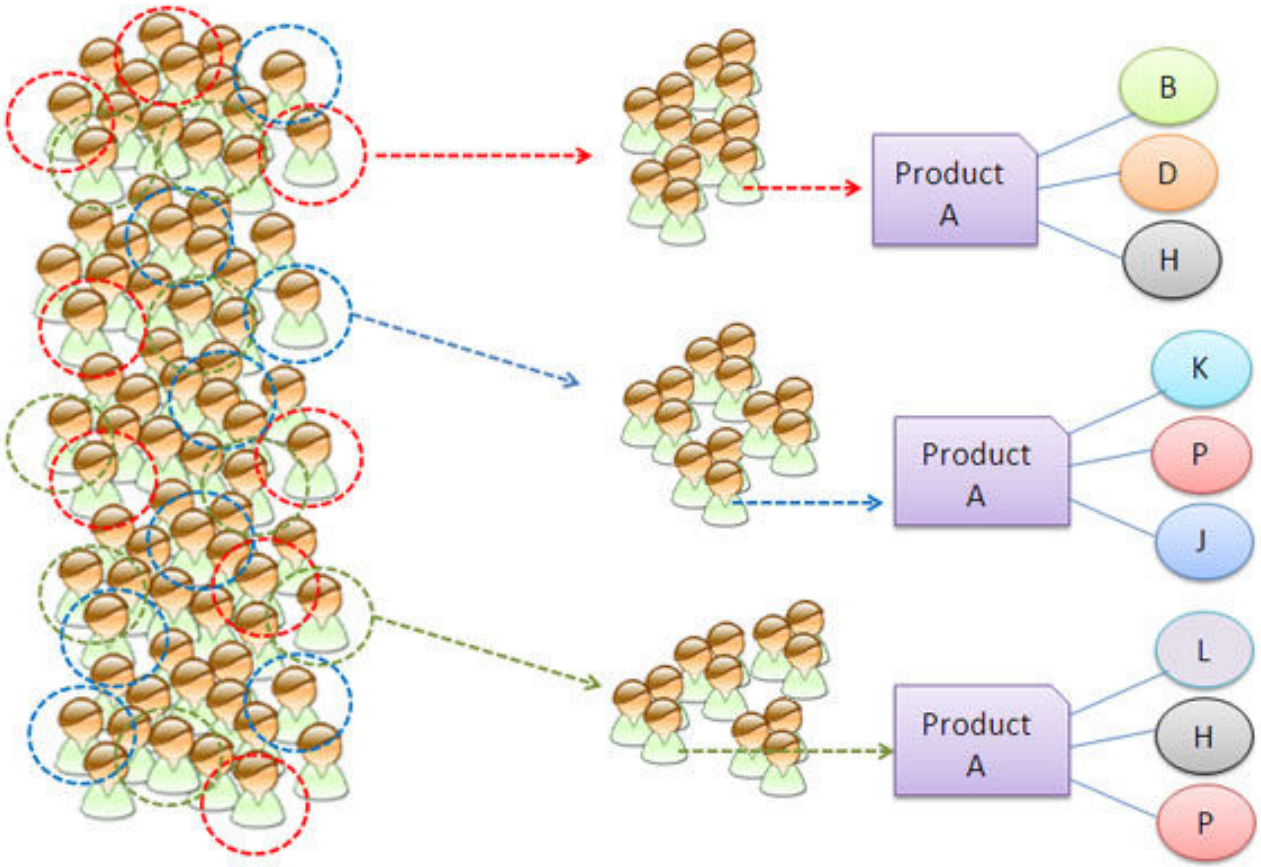


- Algumas dificuldades encontradas:
  - o textos são relativamente curtos;
  - geralmente são só uma lista de palavras sem estrutura gramatical;
  - contém abreviações e erros de digitação.

Bayesian-inference Based Recommendation in Online Social Networks

- Sistemas de recomendação na internet são essenciais.
- Em sites de compras, são apresentados ao consumidor apenas produtos que lhe interessam.
- *Netflix* ofereceu um prêmio de um milhão de dólares à uma equipe que melhorou o sistema de recomendação em 10%.
- A metodologia proposta pode ser usada para recomendar produtos e serviços em geral.

- Sites conhecidos que utilizam sistemas de recomendação:
  - Amazon, iTunes (music), Netflix, etc.



- Consideram que os usuários estão conectados por uma rede social.
- Os amigos dividem entre si suas opiniões sobre o produto.

- O histórico das preferências dos usuários são armazenadas.
- Quando o usuário quer comprar um produto;
  - ele requisita a opinião de seus amigos.
- Com base nessas respostas o método apresenta o quão recomendável é para o usuário o produto procurado.