0cc8456950acc93553591f4a79b138bbeaf0.pdf

Análise Preditiva Privada em Dados Médicos Criptografados

Mais e mais empresas e indivíduos confiam seus dados em serviços em nuvem e terceirizam tarefas computacionais em seus dados para provedores de serviços de terceiros. Isso suscita preocupações sobre a privacidade das informações sensíveis, uma vez que os dados são armazenados em centros de dados externos ou de base. Em particular, no setor da saúde, os registros de pacientes pessoais sensíveis precisam ser confidenciais. A privacidade das informações confidenciais pode ser garantida, se for codificada pelo proprietário dos dados antes de ser carregada para um serviço na nuvem. Dessa forma, apenas o proprietário de dados legítimo pode acessar os dados, descriptando-o usando sua chave privada de descriptografia. Mas a criptografia limita a possibilidade de terceirizar computação nas informações armazenadas externamente, especialmente se o data center não tiver acesso à chave de descriptografia, uma vez que a chave, para esquemas de criptografia padrão, é necessária para descriptografar os dados de modo a realizar uma computação em isto.

No entanto, exatamente essas tarefas computacionais são muitas vezes cruciais para o valor comercial de manter bancos de dados de informações do cliente ou do paciente. Por exemplo, um hospital pode querer ser avaliado em seu desempenho com base nos registros de saúde de seus pacientes, mas pode não querer divulgar os detalhes de todos os registros de pacientes. Em outro exemplo, um paciente pode querer usar um serviço na web que armazene e mantenha todos os seus registros médicos em um lugar centralizado, mas ela pode não confiar no serviço da nuvem para manter confidencial seu condiçoe de dados de saúde privada. Ainda assim, ela pode querer obter informações sobre seu estado de saúde, como uma previsão de se ela irá ou não contrair uma doença específica.

Cenários para computação privada em dados médicos:

Registros de pacientes:

Muitas opções comerciais existem hoje para armazenamento e acesso on-line aos registros médicos eletrônicos dos pacientes (EMR).

Em [1], foi proposta criptografia controlada por paciente (PCE), permitindo que os pacientes terceirizassem o armazenamento de seus registros pessoais em forma criptografada. Esta proposta incluiu a possibilidade de usar uma estrutura de chave hierárquica para permitir o compartilhamento de chaves apenas para ramos seletivos da gravação (por exemplo, apenas para registros dentários) e uma proposta para usar criptografia pesquisável para permitir que o paciente procure de forma particular e remota o registro. Em [2], os autores propõem a utilização de criptografia baseada em atributos para alcançar uma maneira mais flexível e dinâmica de compartilhar permissões em partes seletivas de um EMR. Em [3], a criptografia homomórfica foi proposta para permitir uma certa quantidade de computação nos dados médicos criptografados do paciente, incluindo, por exemplo, dados que foram carregados em forma criptografada de várias fontes, como monitores de saúde, laboratórios ou médicos. O resultado criptografado da computação pode ser retornado ao paciente para descriptografar, ou para qualquer outra parte designada com quem o paciente tenha compartilhado chaves de descriptografia, como um médico ou membro da família.

Bases de dados médicas para pesquisa:

Quando uma grande quantidade de dados de vários pacientes diferentes, relevantes para uma determinada doença (como o câncer) está disponível, a construção de modelos preditivos a partir desses dados usando o aprendizado de máquina pode contribuir substancialmente para a ciência médica e para o bem público. Tais modelos podem ajudar a identificar os fatores de risco genéticos e ambientais e constituir a base para a análise preditiva. No entanto, as preocupações com a privacidade, mesmo quando os dados são anonimizados, precisam ser ponderadas no equilíbrio (ver [4] para um estudo recente sobre a re-identi fi cação de dados neste contexto em que certos indivíduos de alto perfil são re-identificados).

Bancos de dados disponíveis publicamente. Esta configuração permite o maior acesso aos dados, criando assim o maior desafio de privacidade. As diferentes técnicas de preservação da privacidade são potencialmente aplicáveis. A privacidade global é implementada diretamente em dados não criptografados e funciona fornecendo respostas aproximadas para consultas para obscurecer parcialmente a informação exata. Veja [6] para obter uma visão geral das diferentes técnicas de privacidade.

Bancos de dados compilados de forma privada. Os dois cenários a seguir são mais adequados do que o cenário acima para criptografia homomórfica. Em primeiro lugar, as empresas privadas que desenvolvem e implantam ferramentas para pesquisa genética acumulam dados através de seus negócios. Dependendo dos acordos com os clientes, eles podem estar obrigados a manter a confidencialidade dos dados do paciente quando a computação de terceirização em tais bancos de dados. Em segundo lugar, hospitais ou clínicas também mantêm registros de pacientes, mas são vinculados por regulamentos e acordos de pacientes para lidar com esses dados com confiança. Em ambas as configurações, o hospital ou empresa pode procurar melhorar seus negócios, aprendendo com seus dados. Um provedor de serviços em nuvem pode obter armazenamento e / ou análise sobre esses dados privados, aplicando técnicas de aprendizado de máquina para aprender com os dados para construir modelos preditivos. A criptografia homomórfica permite a terceirização de computação nos dados, criptografada sob a chave do hospital ou da empresa.

3e18340408a291897c8c6ea7a0f07289fe54.pdf

Armazenando e Usando Dados de Saúde em uma Nuvem Privada Virtual

A crescente adoção de registros de saúde eletrônicos fornece uma fonte rica para a mineração de dados, a fim de identificar padrões e tendências nos dados de cuidados de saúde. Muitas comunidades nos Estados Unidos têm ou estão formando trocas de informações de saúde (HIE). Os intercâmbios de informações de saúde servem como um hub comum para troca de dados. Essas organizações permitem às organizações de saúde transferir dados, contatando um hub comum, o HIE, em vez de manter conexões com inúmeras organizações de pares.

Esperamos que essa tendência cresça, uma vez que as universidades (e outras organizações de pesquisa) podem chegar a um acordo legal com o HIE para troca de dados em vez de negociar com numerosas organizações. O desafio fundamental para permitir tais parcerias de dados é fornecer um ambiente seguro para o compartilhamento de dados e aderir ao HIPAA, conforme necessário.

Um sistema para suportar a mineração de dados de informações de saúde deve ser um servidor privado acessível a apenas alguns indivíduos autorizados, com suporte apropriado para implementar a auditoria e outros requisitos para a conformidade HIPAA.

1471-2105-12-S12-S7.pdf

Novas ameaças à privacidade dos dados de saúde

uma grande quantidade de dados de saúde foi digitalizada para reduzir o custo e melhorar a qualidade e eficiência dos cuidados de saúde. Em vários cuidados, as informações de saúde do paciente, como dados demográficos, problemas, medicamentos, notas de progresso, dados laboratoriais e histórico médico são registradas no formato de registros eletrônicos de saúde (EHRs). A disponibilidade e a legibilidade dos registros eletrônicos não só facilitam o compartilhamento de informações em atividades relacionadas ao cuidado, mas também reduzem os erros médicos e o tempo de serviço. Com a ampla adoção de EHRs, a segurança e a privacidade dos dados de saúde digitalizados tornam-se extremamente críticas, considerando que os dados médicos são altamente sensíveis aos pacientes.

O que ainda complica o problema é que, com a grande quantidade de dados de saúde digitalizados, há sempre uma demanda para publicar os dados para um uso mais inteligente. Imensos volumes de EHRs são publicados todos os anos para uso secundário, como pesquisa médica, saúde pública, administração governamental e outros serviços relacionados à saúde [2].

Um EHR típico consiste em um conjunto de atributos de identificador (por exemplo, nome, SSN), atributos de quase-identificador (por exemplo, gênero, código postal) e atributos sensíveis (por exemplo, doenças). Para proteger a privacidade dos proprietários de discos, os EHRs precisam ser desidentificados [3-6] ou anonimizados [7-10] antes de serem publicados.

As pesquisas atuais sobre proteção da privacidade do paciente em sistemas de informações de saúde são centralizadas em torno da proteção do EHR - isto é, para proteger a informação do paciente de serem abusados ​​por usuários autorizados ou ser acessados ​​por pessoas fora de uso não autorizadas ou ser re-identificados a partir de dados de saúde publicados para uso secundário .

Os regulamentos sobre a divulgação de dados são estabelecidos e legislados para enfrentar a ameaça de privação de infração e melhorar a proteção de dados relacionados à saúde [11]. Enquanto isso, as tecnologias da informação, tais como controle de acesso, criptografia, verificação de integridade de arquivos, firewalls e mecanismos anti-vírus, defendem tanto o abuso interno quanto o acesso não autorizado aos dados de saúde eletrônicos. Várias técnicas de identificação de identificação [3-6] e anonimização de dados [7-10] também foram propostas para desinfetar os conjuntos de dados na publicação e garantir que os dados não divulguem informações sobre os dados confidenciais.

No entanto, à medida que a Web ganha popularidade e toca muitos aspectos de nossa vida diária, ela se torna a maior fonte de informações pessoais. Os adversários possuem armas poderosas e conhecimento rico, que de alguma forma são fornecidos pelas próprias vítimas.

é razoável que possamos levantar a questão: "quando um invasor possui uma pequena quantidade de informações (possivelmente imprecisas) de fontes relacionadas à saúde e associe essas informações com informações acessíveis ao público de fontes on-line, a probabilidade de o atacante seja capaz de descobrir a identidade do paciente alvo e quais são os potenciais riscos de privacidade " [Primeiro, uma grande quantidade de registros públicos foram acessados ​​on-line, incluindo listas telefônicas, registro de eleitores, registros de nascimento / morte, etc. Embora alguns deles apliquem certas restrições para se defender contra abusadores, ainda é relativamente fácil ou barato rastrear / baixar tais registros. Mais recentemente, sites de redes sociais online, como o Facebook e o MySpace, surgiram para atrair com êxito um grande número de usuários, que voluntariamente colocam suas informações pessoais em sites de redes sociais online para compartilhar com pessoas. Em segundo lugar, com a nova sofisticação das técnicas de recuperação de informação e os avanços das técnicas de busca nos motores de busca, torna-se inesperadamente fácil realizar a extração na escala da Web de informações pessoais dos usuários que estão prontamente disponíveis em várias redes sociais online (por exemplo, [12- 16]). Como resultado, adversários maliciosos ou curiosos poderiam facilmente aproveitar essas técnicas para coletar a informação privada dos outros, que está prontamente disponível em registros públicos online ou em várias redes sociais.]

Ataques em registros de saúde:

Uma proteção efetiva no EHR publicado é a identificação e anonimização de dados. No entanto, mesmo com os dados higienizados, os atributos sensíveis que pertencem a um indivíduo podem ser aprendidos com outros atributos não-sensíveis em combinação com conhecimento externo (por exemplo, lista de registro de eleitores, listas telefônicas, etc.). Os riscos de tais ataques de reidentificação foram intensamente estudados, o que mostra que a quantidade e os tipos de conhecimento externo de um invasor desempenham um papel importante no raciocínio sobre privacidade na publicação de dados [9,10,17,18].

não é fácil, senão impossível, que um editor de dados conheça o conhecimento externo que o atacante possui. Portanto, a pesquisa atual sobre a publicação de dados de preservação da privacidade estuda o problema de uma perspectiva teórica, fazendo suposições sobre o conhecimento de fundo do invasor, quantificando o conhecimento externo, independentemente do seu conteúdo e desinfectando os dados para garantir que a quantidade de divulgação esteja abaixo de um limite especificado [17 , 18]. Como resultado, essa proteção, por um lado, não leva em conta que uma grande quantidade de conhecimento externo é acessível aos adversários de várias fontes on-line (por exemplo, redes sociais), por outro lado, pode distorcer muito os dados e seus usos secundários.

é de grande importância investigar os tipos e quantidades de conhecimento externo que um atacante poderoso possui ou infeste do imenso volume de dados eletrônicos de múltiplos recursos on-line.

Outro potencial ataque de privacidade relaciona-se com a questão da falta de influência de dados médicos sensíveis. O pessoal de entrega de cuidados de saúde pode violar as regras de privacidade, divulgando ou roubando registros de saúde privados para usos não autorizados.

Este é um abuso / infração típico com acesso autorizado aos dados. Mais frequentemente, os atacantes não têm autorização para acesso a dados. Ou escutar ou transferir informações privadas em trânsito ou penetrar em sistemas de EHR para obter o controle de dados de saúde valiosos. No entanto, esses tipos de ataques são muitas vezes subestimados [20]. Acreditamos que essa subestimação é parcialmente de um mal-entendido fundamental de que a informação revelada por descuido ou uso indevido é apenas uma parte da imagem geral e não causará uma divulgação de privacidade severa.

Ataques de fontes externas:

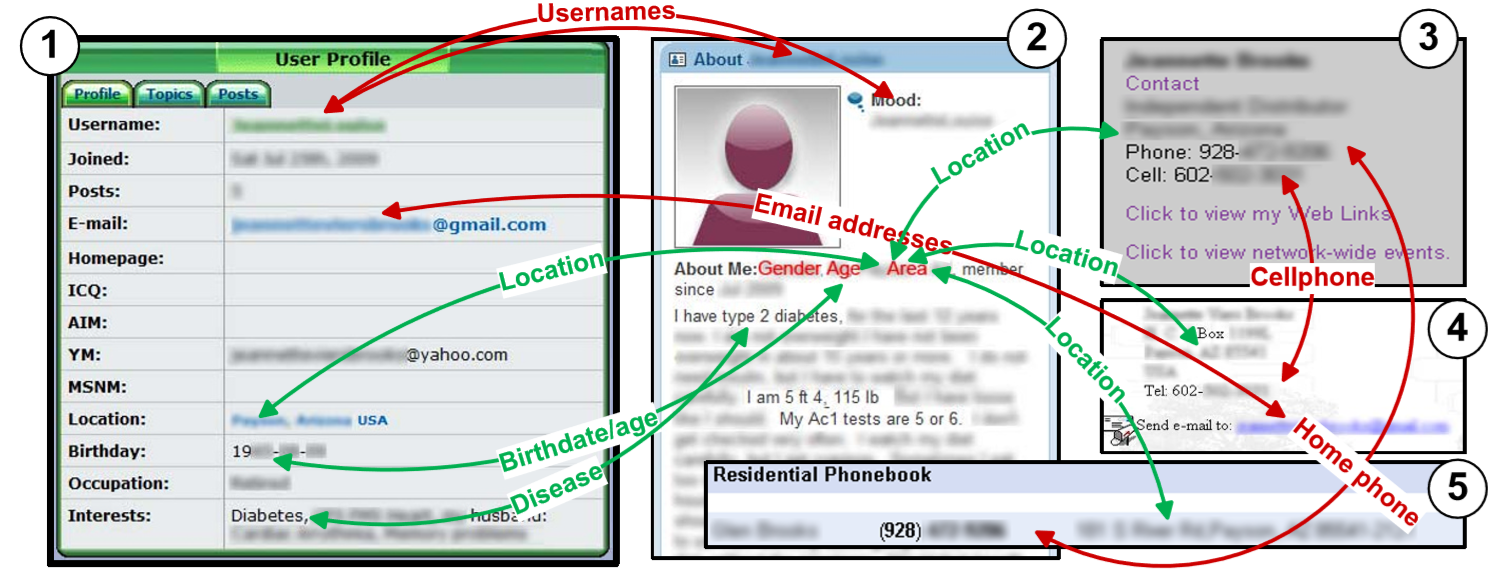
Os usuários de ONS geralmente divulgam voluntariamente suas informações pessoais com detalhes surpreendentes. Por exemplo, os usuários do Linked In listam suas experiências educacionais e de trabalho para buscar possíveis oportunidades de carreira e os usuários do MedHelp compartilham detalhes sobre sua vida e experiências médicas que esperam receber sugestões médicas pertinentes de outros. Ao liberar informações relacionadas à privacidade on-line, um mal-entendido fundamental desses proprietários de informações é que é improvável que vincule informações pertencentes ao mesmo indivíduo de diferentes fontes on-line.

Infelizmente, com os avanços das técnicas de busca e recuperação de informações, é possível que um invasor agregue informações pessoais de um usuário direcionado de diferentes fontes on-line, associando atributos desprotegidos, mas identificáveis ​​ou semi-identificáveis ​​(por exemplo, nomes de conta ou endereço de e-mail idênticos de um usuário descuidado) [21]. Enquanto isso, com esforços governamentais e industriais, uma grande quantidade de registros públicos foram digitalizados e disponibilizados on-line. Os adversários poderiam facilmente acessar e utilizar essas informações para comprometer a privacidade dos outros, especialmente seus dados de saúde altamente sensíveis.

Achamos que é altamente possível que um invasor agregue informações diversas de fontes múltiplas (possivelmente relacionadas à medicina) e associe os atributos para identificar um paciente direcionado com alta confiança.

Estudo de caso do mundo real:

A Figura 1 demonstra um exemplo de um estudo de caso do mundo real com um atacante simulado: "Jean" (cujo nome completo foi descoberto, mas removido aqui para proteção de privacidade) tem diabetes tipo II, então ela participa ativamente de dois sites de rede social online da Medicare , MedHelp (http://www.medhelp.org) e MP e Th1 Discussion Forum (http://www.curemyth1.org). Suponha que estes sejam os dois únicos fatos "triviais" que o atacante conhece sobre Jean: Jean possui diabetes II e Jean possui perfis em ambos OSNs.



os perfis de usuário em ambos os OSNs são acessíveis ao público (após o registro), o primeiro movimento do invasor é rastreá-los de ambos os sites, o que não é uma tarefa impraticável, mesmo para os invasores com recursos limitados e poder de computação. Em seguida, o invasor analisa os perfis rastreados para associar os perfis aos atributos correspondentes. A associação deve ser realizada sob duas premissas: (1) um indivíduo é identificável por um conjunto de atributos e (2) os valores de atributos em tal conjunto de atributos são informações autênticas relativas a um indivíduo.

O segundo pressuposto é válido considerando a natureza das redes sociais médicas em linha. O principal objetivo das redes sociais médicas em linha é servir como plataformas para médicos e pacientes para discutir sintomas, comparar opções de tratamento e trocar avanços médicos. Portanto, para receber avisos médicos imparciais, os usuários geralmente fornecem informações pessoais reais e precisas para a maioria dos campos nos perfis de usuários, embora eles utilizem pseudônimos para se registrarem nos OSNs para fins de proteção de privacidade. Portanto, é razoável acreditar que os valores de atributos não identificáveis nos perfis de usuários são reais.

Para validar o primeiro pressuposto, analisamos os arquivos dos perfis de usuário de ambos os sites de redes sociais médicas:

• O perfil do usuário MedHelp inclui nome de usuário, gênero, idade, localização, tempo combinado, interesse e um campo de texto que permite insumos mais detalhados sobre o usuário e suas condições e problemas médicos específicos.

• O perfil de usuário MP e Th1 contém três tipos de campos: (1) campos diretamente relacionados às atividades de um usuário neste site (por exemplo, nome de usuário, tempo unido); (2) campos sobre seus outros contatos on-line (por exemplo, e-mail, página inicial, ICQ) e (3) campos sobre suas informações pessoais (por exemplo, data de nascimento, ocupação, local e campo de texto para especificar interesses pessoais em informações médicas ).

Ambos os perfis contêm o conjunto de atributos {Gender, Location, Age}. De acordo com um estudo famoso [22] de dados do censo dos EUA de 1990, 87% da população dos EUA é identificável de maneira exclusiva pelo conjunto de atributos (gênero, código postal, data de nascimento) e 53% da população dos EUA é identificada exclusivamente pelo conjunto de atributos {Sexo, Localização, Data de Nascimento}, onde a "localização" se refere à cidade, cidade ou município em que a pessoa reside. Uma vez que as duas redes sociais visam pessoas com interesses particulares, é razoável considerar {Gender, Location, Age} como um quase-identificador neste caso.

A partir dos dados rastreados, o atacante encontra dois perfis "vinculáveis", como mostrado na Figura 1 (1) e Figura 1 (2): 1. O Perfil 1 mostra "meu marido" que indica que o proprietário é uma fêmea, o que é consistente com o gênero mostrado no Perfil 2; 2. Os locais em ambos os perfis são a mesma cidade pequena com aproximadamente 15K de população; 3. A data de nascimento no Perfil 1 é consistente com a idade mostrada no Perfil 2. Além disso, os nomes de pseudo-usuário em ambos os perfis são idênticos (e relativamente únicos), e ambos os perfis demonstram interesses sobre uma mesma doença e sintomas - diabetes tipo II É razoável vincular os dois perfis com um certo nível de confiança e associar os atributos de ambos os perfis ao mesmo indivíduo no mundo real - Jean. Depois disso, mais atributos privados de Jean (por exemplo, visitas médicas, diagnósticos, prescrições e medicamentos) são extraídos de suas postagens nos dois sites de redes sociais médicas e adicionados ao perfil reconstruído de Jean. A partir desta linha de base, continuamos a explorar ainda mais a informação privada da Jean na Web.

• Explore por email: os endereços de e-mail podem efetivamente servir como identificadores únicos. Eles geralmente fornecem uma dica útil para vincular dois perfis. Neste exemplo, com o endereço de e-mail fornecido no Perfil 1, recuperamos o Perfil 4 (como mostrado na Figura 1 (4)) através dos motores de busca da Web. O Perfil 4 inclui um número de telefone (que se encontra como um número de telefone celular em análise posterior) e um P.O. Endereço da caixa. O número de telefone e o endereço indicam a mesma cidade, conforme mostrado nos perfis 1 e 2. • Explore por número de telefone: os números de telefone são comumente considerados como identificadores únicos. Com o número de telefone do Perfil 4, descobrimos o Perfil 3 (como mostrado na Figura 1 (3)), que é uma página jobreada contendo a célula de Jean e os números de telefone residencial. Os perfis 3 e 4 contêm o nome completo de "Jean", e também fornecem uma boa dica sobre sua ocupação. Finalmente, com o número de telefone da casa, também localizamos o registro de Jean na lista telefônica residencial, que mostra o nome do marido e o endereço completo da casa (como mostra a Figura 1 (5)).

neste ataque simulado, adquiremos com sucesso cinco perfis, que provavelmente pertencem a Jean, tanto de sites de redes sociais online quanto de dados online acessíveis ao público.

Ao associar os cinco perfis, recuperamos os tipos de informações privadas sobre Jean, incluindo seu nome completo, data de nascimento, nome do marido, endereço residencial, número de telefone da residência, número de telefone celular, dois endereços de e-mail, ocupação e informações médicas, incluindo teste de laboratório resultados.

Com seu nome completo, mais informações sobre Jean são posteriormente descobertas a partir de várias redes sociais. Finalmente, quando o hospital de Jean publica registros de pacientes des-identificados para apoiar a pesquisa médica, o invasor com conhecimento externo obtido a partir do processo acima é altamente provável que re-identifique o registro de Jean.

Todo o processo inclui três etapas: ataques de atribuição, inferência e agregação. Na atribuição, os atributos identificáveis, semi-identificáveis ​​ou sensíveis são aprendidos / extraídos de várias fontes na web. Em particular, três tipos de recursos on-line são considerados no exemplo: (1) bancos de dados on-line acessíveis ao público: registros de registro de eleitores, listas telefônicas, registros de nascimento e morte, (2) sites de rede social online com atributos identificáveis ​​explícitos (por exemplo, LinkedIn, Facebook , etc.), bem como redes sociais específicas relacionadas à saúde (por exemplo, MedHelp); e (3) motores de busca comerciais, que indexam uma boa parcela da web. Em inferência, mais atributos são descobertos nas atividades e relacionamentos sociais através da aprendizagem estatística ou do raciocínio lógico. Em agregação, os registros recuperados de diferentes fontes que pertencem potencialmente ao mesmo indivíduo estão vinculados sob evidências fortes ou fracas, nas quais evidências fortes incluem identificadores ou quase-identificadores correspondentes e evidências fracas são semelhanças identificadas a partir de uma perspectiva estatística.

várias fontes de informação (por exemplo, redes sociais e registros públicos) poderiam ser utilizadas pelos atacantes. Com ataques de atribuição, inferência e agregação, os ataques são capazes de reconstruir perfis de usuários muito abrangentes, com vários tipos de informações altamente sensíveis e privadas (por exemplo, nomes, números de telefone, datas de nascimento, doenças, resultados de testes de laboratório, etc.).

05654718.pdf

Um quadro geral para mineração de dados médicos

A mineração de dados é a extração não trivial de informações implícitas, anteriormente desconhecidas e potencialmente úteis a partir de dados. Isso abrange uma série de abordagens técnicas diferentes, como agrupamento, resumo de dados, regras de classificação de aprendizado, busca de redes de dependência, análise de mudanças e detecção de anomalias [I].

As aplicações da tecnologia de mineração de dados em domínio médico incluem predição da eficácia dos procedimentos cirúrgicos e descoberta de relações entre medicamentos e dados hospitalares.

a revolução digital forneceu meios relativamente baratos e disponíveis para coletar e armazenar grandes quantidades de dados de pacientes em bancos de dados contendo informações médicas ricas e disponibilizados através da Internet para serviços de Saúde globalmente.

As técnicas de mineração de dados aplicadas nesses bancos de dados descobrem relacionamentos e padrões que ajudam no estudo da progressão da doença. A mineração de dados desempenha um papel importante no diagnóstico médico [3].

AS CARACTERÍSTICAS DOS DADOS MÉDICOS

Como muitas organizações e pessoas podem ter acesso a informações de saúde, há preocupação com a privacidade e a segurança das ferramentas de tecnologia da informação de saúde.

Ao longo dos últimos anos, houve violações de segurança nos hospitais e consultórios médicos, resultando em roubo de identificação médica.

O acesso a suas informações pessoais pode permitir que os ladrões façam faturas por serviços médicos em seu nome [4]. Os dados médicos envolvem alguma informação privada dos pacientes, que ficará indignada quando a informação privada for invadida inesperadamente em suas vidas diárias.

A privacidade é diferente da segurança e confidencialidade, quando indivíduos ou organizações acessam informações privadas sem autorização, cria uma questão de segurança. Enquanto os pesquisadores compartilham informações privadas com pessoas ou instituições não autorizadas, expôs a questão da confidencialidade.

367969507b2209c94a4b23992a50d425dfab.pdf

Rumo a um quadro nacional para o uso secundário de dados de saúde: um livro branco da Associação Americana de Informática Médica

Enquanto o acesso eletrônico rápido e seguro dos pacientes a suas próprias informações de saúde pode levar a um atendimento melhor, mais eficiente e mais personalizado, as demandas proliferam para acesso e análise de dados de saúde fora das configurações clínicas. Os dados de saúde agregados fornecem valor a uma ampla gama de pesquisa, qualidade, saúde pública e aplicações comerciais. Por exemplo, a análise de dados clínicos cuidadosamente controlada sustenta a medida de qualidade e segurança na prestação de cuidados de saúde.

Os usos secundários dos dados de saúde podem melhorar as experiências de cuidados de saúde dos indivíduos, ampliar o conhecimento sobre doenças e tratamentos, fortalecer a compreensão da eficácia e eficiência dos sistemas de cuidados de saúde, apoiar a saúde pública e os objetivos de segurança e ajudar as empresas a atender às necessidades dos clientes.

No entanto, o acesso e o uso de dados de saúde representam desafios éticos, políticos, técnicos e econômicos complexos.

Por exemplo, para atender aos requisitos de saúde pública, preparação para emergências e segurança interna, o governo federal iniciou a coleta de dados em tempo real de salas de emergência e outras fontes - sem diálogo público, com base na autoridade da legislação vigente em saúde pública. Além disso, há relatos de compra e venda de dados não anônimos de pacientes e provedores pela indústria médica - realizados sem o consentimento explícito de pacientes ou médicos.

A AMIA convocou uma reunião de diversas partes interessadas (ou seja, o painel) para discutir toda a gama de tais questões, incluindo, entre outras,:

• Quais são os potenciais benefícios e riscos em relação ao uso secundário de dados de saúde?

• Quem possui dados de saúde e quem tem o direito de acessar os dados e para que fins?

• Quais são os problemas de confiança pública em evolução em relação ao consentimento do paciente para o uso secundário de dados de saúde? Os pacientes têm o direito de auditar ou colocar outras restrições sobre o uso de seus dados, mesmo após a anonimização?

• À luz de graves ameaças à saúde pública, como a gripe aviária, como a sociedade concilia o bem público com os direitos dos indivíduos enquanto pesa considerações de saúde versus privacidade?

• Quais os problemas que podem surgir à medida que as tecnologias inovadoras melhoram a capacidade e a facilidade do compartilhamento generalizado de dados e de usos comerciais adicionais?

• O que pode ser feito para abordar questões decorrentes de uso inapropriado e / ou exploração de compartilhamento de dados?

• Que regulamentos, legislação e / ou políticas e procedimentos são necessários para abordar essas questões?

TERMINOLOGIA:

dados anonimizados - alteração da PHI que torna impossível vincular indivíduos com seus dados.

dados de-identificados - a eliminação de todos os identificadores como enumerados sob HIPAA pelo método de porto seguro (ou seja, nome do paciente, número de registro médico, número de segurança social, e outros campos de dados que apontam diretamente um paciente aos seus dados). Existe potencialmente outra abordagem que envolve ter um estatístico para determinar que a capacidade (probabilidade) de poder combinar dados com outras fontes públicas de informação e identificar com sucesso um indivíduo é extremamente pequena.

registro eletrônico de saúde (EHR) - dados pessoais criados, desenvolvidos, mantidos e / ou fornecidos por clínicos, provedores e provedores de saúde aliados em atendimento direto ao paciente; um aplicativo eletrônico contendo informações de saúde sobre indivíduos que são utilizados por clínicos, provedores e profissionais de saúde aliados para prestar cuidados diretos aos indivíduos.

dados de saúde - dados sobre ou de um indivíduo, como idade da pessoa ou nível de potássio sérico. Em conjunto, os dados de um indivíduo são chamados de informações pessoais de saúde (PHI).

registro de saúde pessoal (PHR) - um aplicativo eletrônico através do qual os indivíduos podem acessar, gerenciar e compartilhar suas informações de saúde, em um ambiente privado, seguro e confidencial; dados pessoais criados, desenvolvidos, mantidos e / ou fornecidos por indivíduos sobre si próprios.

uso primário de dados - o uso da PHI pela organização ou entidade que produziu ou adquiriu esses dados no processo de fornecer atendimento direto em tempo real de um indivíduo.

dados anonimatos reversíveis - a alteração da PHI de forma que a re-identificação possa ser realizada através do acesso a uma chave protegida que permita vincular indivíduos com seus dados somente através de um intermediário confiável.

uso secundário de dados - uso de cuidados não diretos da PHI, incluindo, entre outros, análises, pesquisas, medições de qualidade / segurança, saúde pública, pagamento, certificação ou acreditação de provedores e marketing e outros negócios, incluindo atividades estritamente comerciais.

Quem possui os dados no registro de saúde pessoal da Sra. Powter?

Quando a Sra. Powter sai do Plano de Saúde nº 1, o que acontece com os dados dela?

Quais são os problemas (por exemplo, padrões de intercâmbio de dados, custo) que ocorrem ao transferir dados entre os planos de saúde?

Que outras utilizações secundárias dos dados devem ser permitidas?

Se a Sra. Powter receber permissão para cada instância de uso ou deveria dar permissão global?

os participantes estabeleceram uma distinção entre propriedade de dados e acesso a dados; diferenciado entre PHR e EHR; suscitou preocupações sobre o uso indevido de dados, o consentimento sob coação e lacunas nas proteções HIPAA

**considerou os direitos dos pacientes em relação às suas obrigações em relação aos dados do próprio paciente.**

Conclusão 1: O uso secundário de dados de saúde é generalizado. As apresentações e discussões, bem como a literatura (ver Apêndice D para uma bibliografia selecionada), documentam o uso secundário generalizado, crescente e a reutilização de dados de saúde. Tais usos ocorrem em setores públicos e privados para fins de propriedade, pesquisa e monitoramento com uma regulamentação inferior à abrangente. Os participantes concordaram que, na maioria dos casos, os provedores, médicos e seus pacientes geralmente desconhecem esse desenvolvimento, apesar do crescimento e sucesso de uma indústria de milhões de dólares com base na venda de dados relacionados à saúde e à saúde. Além disso, enquanto a HIPAA exige que muitos prestadores de cuidados de saúde e seguradoras de saúde obtenham documentação adicional antes de divulgarem informações de saúde específicas de cada pessoa e analisarem minuciosamente os pedidos de acesso a informações de saúde para fins secundários, como, por exemplo, para pesquisas, as regras HIPAA apenas abordam o uso e a divulgação de informações de saúde por "entidades cobertas" (por exemplo, prestadores de cuidados de saúde, planos de saúde e câmaras de compensação).

Recomendação 1: aumentar a transparência do uso de dados e promover a conscientização pública. As discussões sobre políticas públicas em curso devem abordar explicitamente e diretamente o uso secundário de dados de saúde. Conduzir e gerenciar essas atividades deve recrutar diversas partes interessadas e divulgar completamente usos e salvaguardas através de processos abertos e facilmente acessíveis.

Encontrar 2: o foco precisa ser acesso e controle de dados, e não propriedade de dados. O consenso do grupo foi que o foco na "propriedade" desviou a atenção do desenvolvimento necessário de políticas e práticas sólidas. Os participantes reconheceram que a responsabilidade de garantir a privacidade e salvaguardar os dados do paciente se aplica ao diverso continuo de usuários de dados. Os avanços técnicos permitem a criação de muitos bancos de dados que agora são mantidos, atualizados, usados ​​e reutilizados para múltiplos propósitos, inclusive aqueles que estão fora do atendimento direto ao paciente. A tecnologia também permitiu a transmissão fácil de tais dados. Apesar dos requisitos da HIPAA em relação à des-identificação de dados e à adesão aos acordos de uso de dados, existe o potencial para a re-identificação de pacientes e provedores através da ligação de bancos de dados diferentes. Existe uma necessidade de explorar e abordar explicitamente questões de acesso e controle de dados de saúde ao longo dos ciclos de vida dos dados. Extensas discussões abordaram a necessidade de desenvolver políticas para usos secundários de dados de saúde, reconhecendo que essas políticas serão complexas.

Recomendação 2: Concentre as discussões em curso sobre acesso, uso e controle de dados - não sobre a propriedade. As reuniões de Consensusbuilding que abrangem um grupo eleitoral abrangente devem se concentrar nas políticas e práticas de acesso e controle de dados para o uso secundário de dados. O foco deve enfatizar o acesso e o controle, e não a propriedade. Os participantes devem considerar as melhores abordagens para gerenciamento de riscos e mitigação.

Encontrando 3: Problemas críticos incluem privacidade do paciente e confiança pública. O uso de dados de pacientes específicos de pessoas para fins diferentes do atendimento direto ao paciente e saúde pública não é bem compreendido e é mal monitorado. Isso suscita inúmeras questões éticas, técnicas, econômicas e processuais. O sentido dos participantes da reunião foi que existem poucas salvaguardas que abordam adequadamente as utilizações secundárias dos dados de saúde. Outras discussões sobre o consentimento informado devem esclarecer como os dados usam para fins específicos podem permanecer em conformidade com as leis federais, estaduais e locais. Os dados de dados de saúde não são abrangidos por regulamentos de privacidade, incluindo os usos dos dados obtidos por meio de consentimento forçado ou compelido, podem prejudicar a confiança pública e potencialmente prejudicar o bem público. Alguns membros do painel afirmaram que o desenvolvimento e a execução de opções de escolha de pacientes que envolvem autorização explícita para o uso de seus próprios dados (optando / desativando) fornece os únicos meios adequados para mitigar os problemas de privacidade do paciente. Os participantes reconheceram que não existe "uma única perspectiva de paciente unificado (consumidor)". Os consumidores verão o problema de muitas maneiras possíveis - assumindo que estão informados sobre isso. Assim, uma variação substancial nos pontos de vista do consumidor fará questões relacionadas ao consentimento do paciente (consumidor) e ao complexo de escolha.

Recomendação 3a: Discutir políticas de privacidade e segurança para uso secundário de dados de saúde. Para desenvolver o consenso sobre questões fundamentais, as organizações do setor público e privado que promovam o uso da informação de saúde devem promover discussões que incluam uma ampla gama de partes interessadas do que participaram dessa conferência. As discussões em curso devem abordar questões complexas relacionadas ao uso secundário privado e seguro de dados de saúde.

Recomendação 3b: aumentar a conscientização pública sobre benefícios e desafios associados ao uso secundário de dados de saúde. Uma ampla gama de partes interessadas, especialmente os grupos de pacientes e cuidadores orientados para o consumidor, devem promover a educação pública sobre os benefícios dos EHRs e sobre o uso secundário de dados de saúde. Um primeiro passo é identificar as organizações e agências apropriadas que têm um papel a desempenhar nesse esforço. O objetivo da educação é desenvolver a conscientização pública e a confiança no uso secundário de dados de saúde.

c88692cfb6fe378b306fc4f11724ff3ce8fb.pdf

FRR: recuperação remota justa de registros médicos privados terceirizados em redes eletrônicas de saúde

EHN gera e grava numerosos documentos médicos, incluindo imagens médicas, relatórios de verificação médica, relatórios clínicos, prescrições médicas, relatórios de monitoramento de pacientes e assim por diante [1,2]. Esses documentos são sensíveis e precisam ser protegidos, de modo que apenas as partes autorizadas possam acessá-los.

A computação em nuvem apresenta alguns problemas de segurança: verificação de integridade de dados remota, recuperação remota de dados e privacidade de dados.

Para alcançar a privacidade dos dados, uma abordagem geral é permitir que os clientes criptografem seus dados antes de enviá-los para o CS. A criptografia pode ser projetada de forma elegante para permitir pesquisar dados criptografados ou recuperar dados com conteúdo específico. Para melhorar a robustez da recuperação de dados, os códigos de correção de erros podem ser usados antes que os dados sejam armazenados no CS.

Nesse cenário, numerosos registros médicos são criptografados e armazenados em um CS. Ao contrário de outras aplicações de armazenamento na nuvem, aqui o CS, o hospital e os pacientes podem estar interessados em modificar desonestamente os dados médicos terceirizados. Por isso, um terceiro independente é empregado para garantir a integridade e a recuperação dos dados terceirizados. Além disso, quando a investigação médica é realizada no caso de acidentes médicos, o terceiro deve poder recuperar os registros médicos originais de forma justa. Nós argumentamos que é necessário um terceiro independente, uma vez que o hospital e os pacientes podem manipular os registros médicos originais ou se recusam a decifrar os registros criptografados se os registros originais pudessem ser usados como argumentos contra eles em uma disputa.

Para alcançar a privacidade dos dados, os registros médicos privados precisam ser criptografados antes de serem terceirizados para servidores em nuvem para que o CS curioso não consiga entender o conteúdo dos dados armazenados.

Ideia:

1. Paciente: quando um paciente vai ao hospital para tratamento de saúde, registros médicos privados são gerados e criptografados pelo paciente.

2. Hospital: o hospital ajuda o paciente a criar alguns registros médicos privados. Depois de receber os registros médicos criptografados do paciente, o hospital carrega-os para o CS.

3. CS: CS tem espaço de armazenamento significativo e recursos computacionais para manter os dados dos hospitais. No entanto, a CS também pode ocasionalmente se deportar e modificar algumas partes dos dados armazenados.

d43f81c6e8c3e46f5bdcb686d60d40915de8.pdf

Unicidade da mineração de dados médicos

colocamos várias questões que devem ser respondidas pela comunidade, de modo que ambos os pacientes em quem os dados são coletados, bem como os mineradores de dados, podem se beneficiar.

existem restrições éticas, legais e sociais na coleta e distribuição de dados, que não se aplicam a espécies não humanas, e que limitam as conclusões científicas que podem ser extraídas.

# Questões éticas, legais e sociais

Como os dados médicos são coletados em assuntos humanos, existe uma enorme tradição ética e legal, destinada a prevenir o abuso de pacientes e o uso indevido de seus dados. Os principais pontos das questões éticas, legais e sociais em medicina podem ser organizados sob cinco rubricas:

• Propriedade de dados

• Medo de ações judiciais

• Privacidade e segurança dos dados humanos

• Benefícios esperados

• Problemas administrativos

## Propriedade de dados

Na teoria jurídica, a propriedade é determinada por quem tem o direito de vender um determinado item de propriedade (Moore e Berman, 2000).

Uma vez que é considerado indecoroso vender dados humanos ou tecido, a questão da propriedade de dados em medicina é igualmente confusa.

Os dados médicos são enterrados em bancos de dados heterogêneos e espalhados por todo o estabelecimento de cuidados médicos, sem nenhum formato comum ou princípios de organização.

A questão da propriedade da informação do paciente não é resolvida, e o objeto de processos recorrentes e altamente divulgados e consultas do Congresso. Os pacientes individuais possuem dados coletados em si mesmos? Os seus médicos possuem os dados? Os seus provedores de seguros possuem os dados? Se os provedores de seguros não possuem os dados dos seus segurados, eles podem se recusar a pagar pela coleta e armazenamento dos dados? Se a capacidade de processar e vender dados médicos humanos é indecente, então, como devem ser compensados ​​os gerentes de dados, que organizam e minam os dados? Ou esse recurso incrivelmente rico para o melhoramento potencial da humanidade deve ser desminado?

## Medo de ações judiciais

o medo de ações judiciais dirigidas contra médicos e outros prestadores de cuidados de saúde.

a medicina de U. S. também tem o clima de negligência mais litigioso do mundo. Alguns argumentaram que esta sobretaxa de 30% sobre os cuidados médicos dos EUA, cerca de mil dólares per capita anualmente, é principalmente medico-legal: quer custos legais diretos, ou a sobrecarga de "medicina defensiva", ou seja, testes desnecessários ordenados pelos médicos para cobrir-se em futuros processos judiciais em potencial.

os médicos e outros produtores de dados médicos são compreensivelmente relutantes em entregar seus dados aos mineradores de dados. Mineradores de dados podem navegar esses dados para eventos adversos.

As anomalias aparentes na história médica de um paciente individual podem desencadear uma investigação. Em muitos casos, a aparência de negligência pode ser um erro de transcrição de dados ou dados; e nem todos os malos resultados em medicina são necessariamente o resultado do comportamento negligente do provedor.

Por expor-se a esse risco, qual recompensa os prestadores recebem em troca?

## Privacidade e segurança dos dados humanos

as regras federais da U. S. estabelecem diretrizes para ocultação de identificadores de pacientes individuais. A parte em jogo não é apenas uma violação potencial da confidencialidade do paciente, com a possibilidade de uma ação judicial subseqüente; mas também a erosão da relação médico-paciente, em que o paciente é extraordinariamente sincero com o médico na expectativa de que tais informações privadas nunca serão divulgadas.

Outra questão é a segurança de dados no tratamento de dados, e particularmente na transferência de dados. Antes que os identificadores estejam ocultos, somente as pessoas autorizadas devem ter acesso aos dados. Uma vez que a transferência dos dados eletronicamente através da Internet é insegura, os identificadores devem ser cuidadosamente ocultos, mesmo para transferências dentro de uma única instituição médica de uma unidade para outra.

existem pelo menos duas necessidades legítimas de pesquisa para a re-identificação de dados médicos des-identificados: primeiro, há uma necessidade de evitar registros duplicados acidentais sobre o mesmo paciente de esticar as conclusões da pesquisa; Em segundo lugar, pode haver uma necessidade convincente de se referir a registros médicos originais (re-identificados) para verificar a correção ou obter informações adicionais sobre pacientes específicos. Esses requisitos especiais poderiam ser gerenciados por agências reguladoras apropriadas, mas não podiam ser cumpridas se os dados fossem completamente anônimos.

Existem quatro formas de identificação de dados do paciente:

* Os dados anônimos são dados que foram coletados para que a identificação do paciente fosse removida no momento em que a informação foi coletada. Os identificadores do paciente não são registrados no momento da coleta da amostra e, portanto, nunca podem ser recuperados.
* Os dados anônimos são dados que são coletados inicialmente com os identificadores do paciente, que posteriormente são removidos irrevogavelmente. Ou seja, nunca pode haver possibilidade de retornar ao registro do paciente e obter informações adicionais.

dados anônimos, como descrito acima, podem ser duplicados acidentalmente e não podem ser verificados para correções ou dados adicionais.

* Dados des-identificados são dados coletados inicialmente com os identificadores do paciente, que posteriormente são codificados ou criptografados. O paciente pode ser re-identificado nas condições estipuladas por uma agência apropriada, tipicamente um Conselho de Revisão Institucional (IRB).
* Os dados identificados só podem ser coletados sob revisão significativa pela instituição, diretrizes federais, etc., com o paciente dando consentimento informado por escrito.

## Benefícios esperados

Qualquer uso de dados de pacientes, mesmo des-identificados, deve ser justificado para o IRB como tendo alguns benefícios esperados.

a Internet é a maneira mais barata e conveniente de distribuir dados e o mais acessível ao público, que pode ter motivos legítimos de acesso. Por exemplo, pode haver grupos de interesses de doenças raras, grupos de vigilância médica ou mesmo pesquisadores com perspectivas científicas não convencionais, que têm reivindicações razoáveis para extrair os dados, mas que não conseguiram montar os recursos financeiros e administrativos para ministrar bancos de dados privados.

## Problemas administrativos

As diretrizes federais emergentes da U. S. para a privacidade do paciente especificam uma série de políticas e procedimentos administrativos.

Deve haver políticas para avaliar e certificar que as medidas de segurança adequadas estão em vigor na instituição de pesquisa. Deve haver contratos legais entre a organização e quaisquer partes externas que tenham acesso a informações de saúde identificáveis ​​individualmente, exigindo que as partes externas protejam os dados. Deve haver planos de contingência para responder a emergências, incluindo um plano de backup de dados e um plano de recuperação de desastres. Deve haver um sistema de controle de acesso à informação que inclua políticas para a autorização, estabelecimento e modificação de privilégios de acesso a dados. Deve haver uma revisão interna contínua dos registros de acesso a dados, a fim de identificar possíveis violações de segurança.

A organização deve assegurar a supervisão do pessoal que realiza atividades de manutenção de sistemas técnicos para manter registros de autorização de acesso, para garantir que o pessoal operacional e de manutenção tenha acesso adequado, empregar procedimentos de segurança do pessoal e assegurar que os usuários do sistema sejam treinados na segurança do sistema.

Deve haver treinamento de segurança para todos os funcionários, incluindo treinamento de conscientização para todo o pessoal, lembretes de segurança periódicos, educação do usuário sobre proteção de vírus, educação do usuário na importância de monitorar falhas de login, gerenciamento de senhas e como registrar discrepâncias.

## Conclusões

A informação médica sobre o paciente individual é considerada altamente privada, e o público em geral é extremamente receoso em relação à divulgação (U.S., 1999). Todos nós apreciamos os benefícios da pesquisa médica realizada em outros pacientes, mas muitas vezes somos relutantes em contribuir ou divulgar nossa própria informação para tais fins. Quando os dados médicos são publicados, espera-se que os pesquisadores mantenham a dignidade do paciente individual e que os resultados serão usados ​​para fins socialmente benéficos (Saul, 2000). Sugeriu-se que as verdades científicas são fundamentalmente amorais; eles podem ser usados ​​para o bem ou o mal (Changeux e Connes, 1995).