### Medical Data Segmentation for Privacy

Ellick Chan, Peifung E. Lam, and John C. Mitchell

---

**Introduction**

Medical record segmentation is a technique to provide privacy and protect against discrimination for certain medical conditions such as STDs, substance abuse and mental health, by sequencing or redacting certain medical codes from a patient's record.

We present an initial study that describes an approach for segmenting sensitive medical codes to protect patient privacy and to comply with privacy laws.

Firstly, we describe segmentation strategies for sensitive codes and explore the link between medical concepts using sources of medical knowledge. Secondly, we mine medical knowledge sources for correlations between medical concepts. Thirdly, we describe an approach that a privacy attacker may use to infer redacted codes based off second order knowledge. More specifically, the attacker could use the presence of multiple related concepts to strengthen the attack. Finally, we evaluate possible defensive approaches against techniques that an adversary may use to infer the segmented condition.

**Algorithm**

- Hypothesis \( H \), input \( q \)
- \( q = \{ \text{query text} \} \)
- \( c \) is the set of concepts in the query
- \( H \) is a hypothesis
- \( c \) is a concept
- \( \alpha \) is the support
- \( \beta \) is the confidence

1. Select a concept from the EHR using an information extraction method
2. Search the EHR for documents that match the query
3. Identify hypotheses from medical concepts in the EHR
4. Select hypotheses according to plausibility criteria
5. Evaluate hypotheses according to support
6. Output hypothesis results

---

**References**