

EstiHMM: een efficiënt algoritme ter bepaling van de maximale sequenties in een imprecies hidden Markovmodel

Probleemstelling

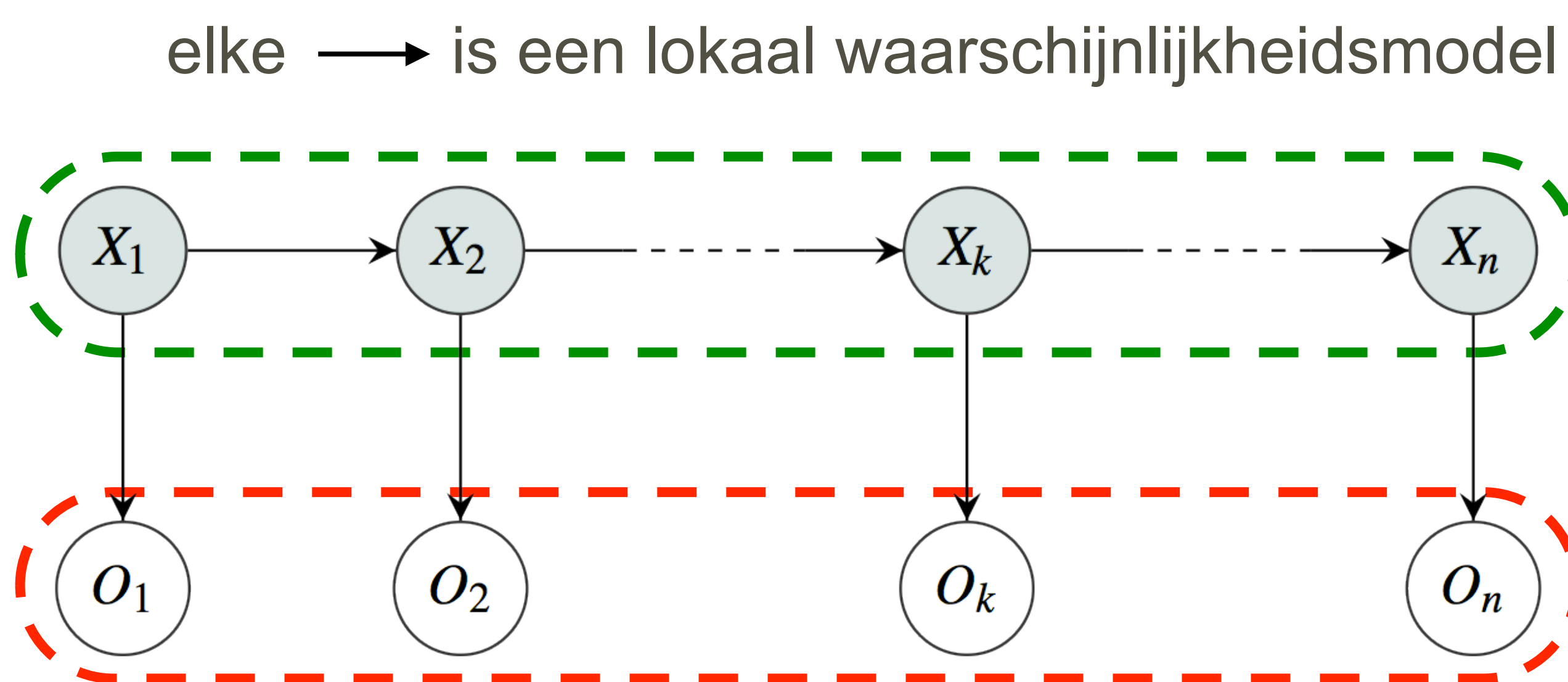
We zijn geïnteresseerd in de waarde (toestand) van een sequentie van **verborgen variabelen** en hebben enkel **observaties** van een andere reeks variabelen die met de vorige in verband staat door waarschijnlijkheidsmodellen. Hoe kunnen we uitgaande van deze **observaties** de **(maximale) toestandssequentie** vinden, zijnde de toestandswaarde van de verborgen variabelen die de hoogste kans heeft om voor te komen?

Modellering

Verborgen **TOESTANDEN**:

Hidden Markovmodel

Gekende **OBSERVATIES**:



Toepassingen



Bio-informatica

DNA-sequenties bepalen op basis van **metingen**

Spraakherkenning

De **geschreven boodschap** vinden die correspondeert met een opname van **uitgesproken klanken**

Optische tekenherkenning

Omzetten van een **ingescand document** in een **tekstbestand**

Crypto-analyse

De **oorspronkelijke boodschap** vinden die overeenstemt met een **gecodeerde tekst**

In deze masterproef wordt het **EstiHMM-algoritme** ontwikkeld. Dit algoritme veralgemeent het gekende Viterbi-algoritme en stelt het in staat om coherent om te gaan met gebrek aan data en **imprecieze gegevens**. Aangezien het Viterbi-algoritme reeds lange tijd met succes toegepast wordt in **diverse domeinen**, is dit nieuwe algoritme **direct toepasbaar**. De grote meerwaarde is dat het - in tegenstelling tot Viterbi - **robuste resultaten** geeft. Dit past in het bredere en maatschappelijk steeds relevanter wordende thema van **sensitiviteitsanalyse**.

Jasper De Bock, Faculteit Ingenieurswetenschappen en Architectuur, **Universiteit Gent**

Het belang van imprecisie

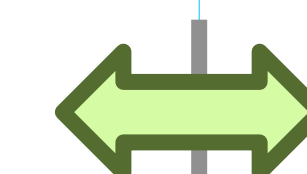
- ✓ In een eerste fase met de pc een **ruwe verzameling oplossingen** bekomen die manueel verfijnd kan worden
- ✓ **Beperkte** hoeveelheid **gegevens** op correcte wijze modelleren
- ✓ **Risico's vermijden**

VITERBI

De verschillen

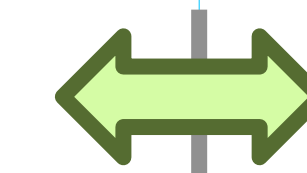
ESTIHMM

Enkel precieze modellen



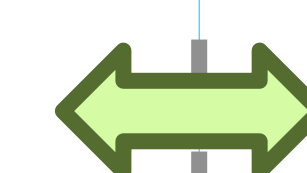
Ook imprecieze modellen

Altijd één oplossing
(soms zeer gevoelig aan afwijkingen van het model)



Meerdere oplossingen mogelijk (in die situaties waar Viterbi niet robuust is)

Zeer efficiënt



Nog steeds zeer efficiënt (lineair in # oplossingen)