Création automatique de résumés vidéo par programmation par contraintes

Haykel Boukadida

Membres du Jury :

Georges QUENOT	Directeur de recherche CNRS - rapporteur
Bernard MERIALDO	Professeur Eurécom - rapporteur
Alexandre TERMIER	Professeur Université de Rennes 1 - examinateur
Julien PINQUIER	Maitre de Conférences Université Paul Sabatier - examinateur
Nicolas BELDICEANU	Professeur École des Mines de Nantes - examinateur
Pascale SEBILLOT	Professeur INSA de Rennes - examinateur
Patrick GROS	Directeur de recherche Inria - directeur de thèse
Sid-Ahmed BERRANI	Resp. équipe de recherche Orange Labs - Co-directeur

04 Décembre 2015, INRIA Rennes, Bretagne Atlantique

Plan

Partie I : Introduction

- Contexte et motivations
- Principaux travaux existants
- Problématiques et contributions

Partie II : Programmation par contraintes

Partie III : Contributions de la thèse

- Modèle #1 : Sélection de plans
- Modèle #2 : Sélection d'extraits basée sur la segmentation en plans
- Modèle #3 : Sélection d'extraits sans aucune segmentation
- Évaluation de la qualité des résumés

Partie IV : Conclusions et perspectives

Contexte et motivations

- Avènement du numérique

Terminaux d'enregistrement (PVR)

Catalogues de VoD

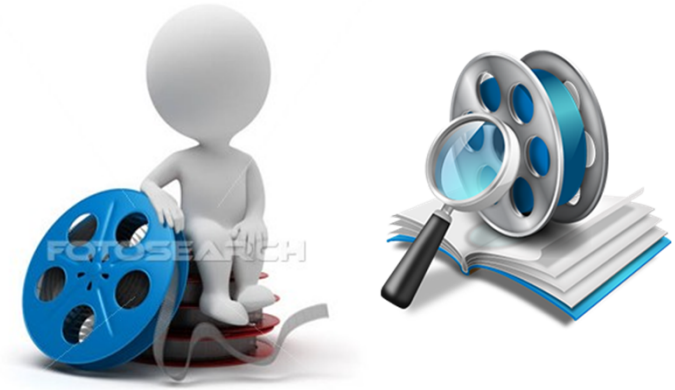
Services de rattrapage TV

Plateformes de partage de vidéos

Avalanche de contenus multimédia

Contexte et motivations

- Nécessité de pouvoir naviguer simplement au sein de grandes collections
  - Moteurs de recherche
  - Moteurs de recommandation



Une courte liste de vidéos

- Nécessité d'avoir un aperçu des vidéos permettant d'effectuer le choix final et pouvoir sélectionner le contenu à regarder

- Un résumé est une nouvelle vidéo :
  - ✓ Durée plus courte
  - ✓ Événements intéressants, moments forts
  - ✓ Vue d'ensemble sur le contenue audio-visuel

Principaux travaux existants

Trois mécanismes utilisés:

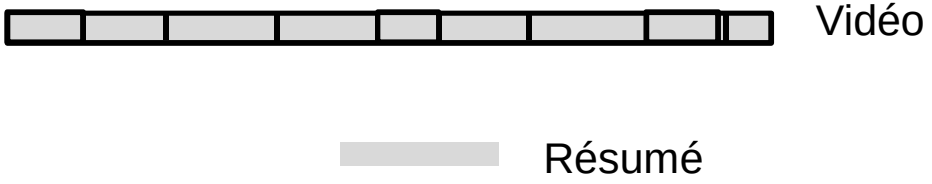
- Élimination des redondances
- Détection des moments forts
- Construction d'une courbe modélisant l'attention humaine

Trois types de résumés dynamiques:

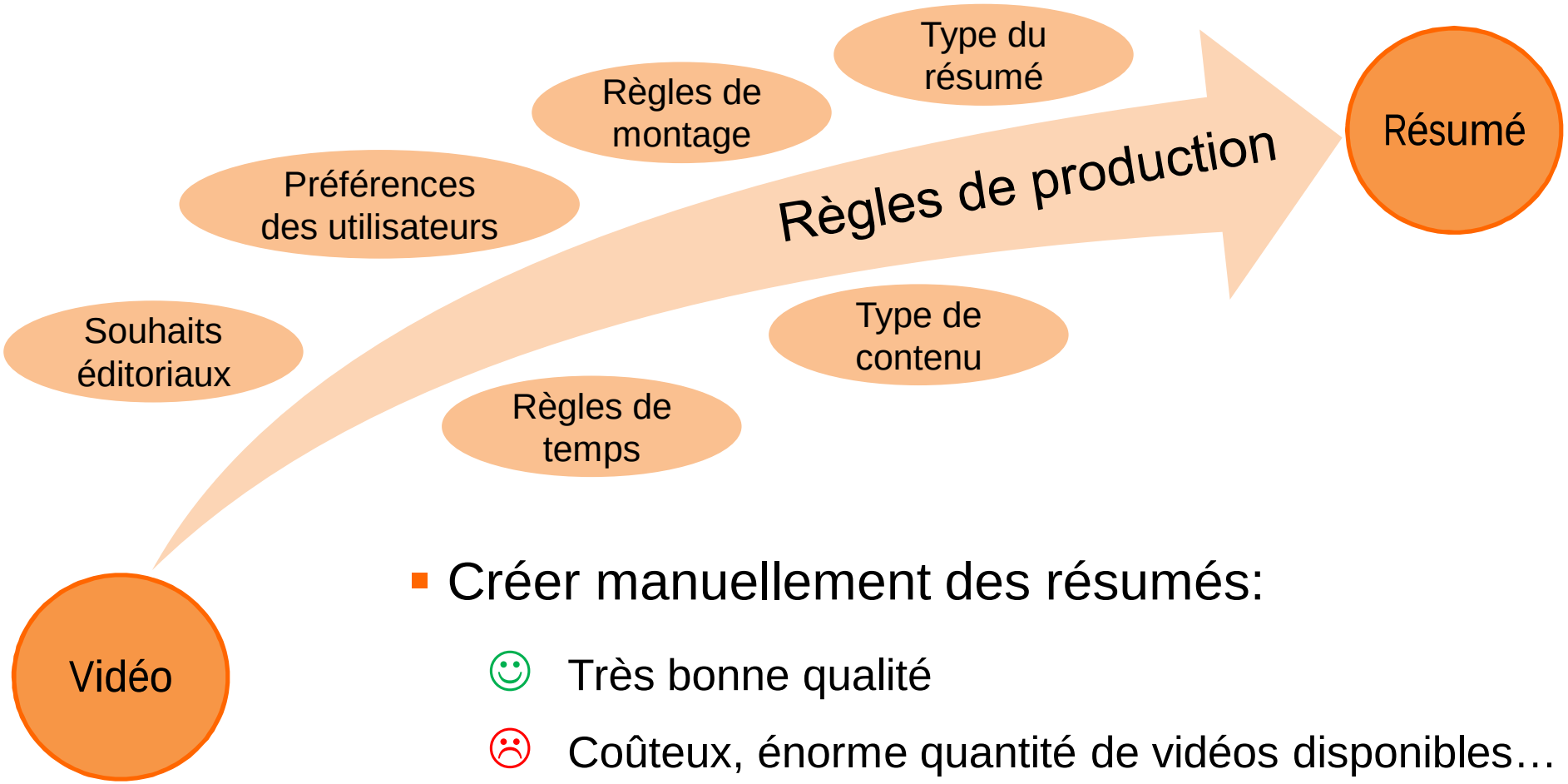
- Reportage informatif
- Événements importants
- Attention de l'utilisateur

Processus de la création d'un résumé:

1. Segmentation de la vidéo
2. Sélection des extraits
3. Concaténation des extraits



Problématiques et contributions



- Créer manuellement des résumés:
  - ☺ Très bonne qualité
  - ☹ Coûteux, énorme quantité de vidéos disponibles…
  - ☹ Non modifiables, une fois généré…

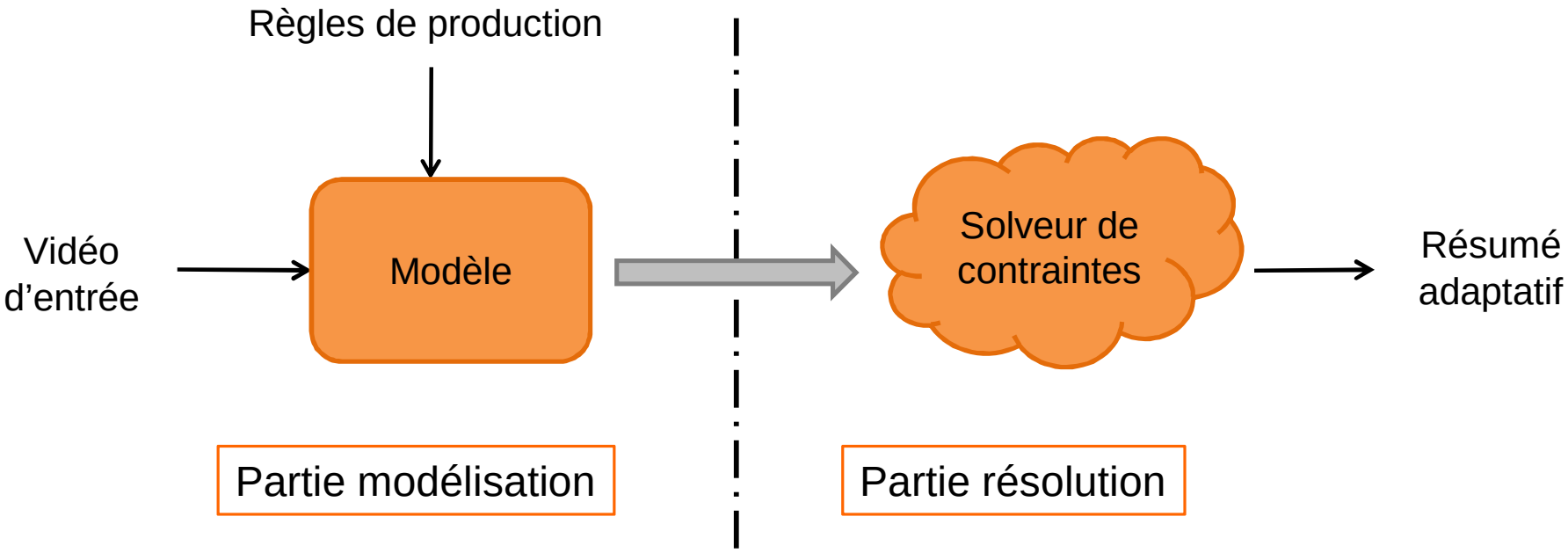
Problématiques et contributions

Problème de création automatique de résumés → Problème de satisfaction de contraintes (PSC)

- Explicitation des règles de production sous forme de contraintes à satisfaire
- Adaptabilité des résumés générés automatiquement
- Diversité des solutions (résumés) qui s'adaptent aux besoins de l'utilisateurs et qui satisfont les contraintes exprimées

Problématiques et contributions

- Séparation entre les règles de production de résumés et l'algorithme de génération de résumés



- Ajouter de nouvelles règles de production (ou modifier des règles) sans revoir tout le processus de création de résumés

Programmation par contraintes (1/4)

« Constraints Programming represents one of the closest approaches computer science has yet made to the Holy Grail of programming: the user states the problem, the computer solves it »

Eugene Freuder

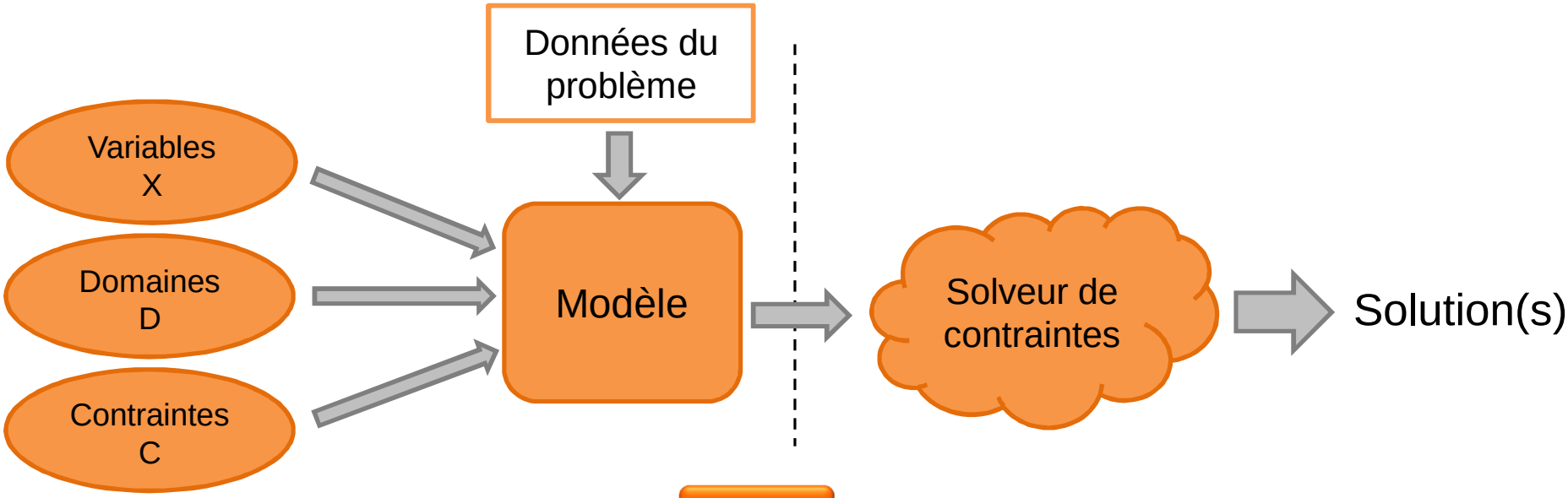
- Une technique de résolution de problèmes qui vient de l'intelligence artificielle et de la recherche opérationnelle
- Résoudre des problèmes d'optimisation combinatoires et de complexité élevée (problèmes NP-complets)
- Offrir une autre façon de formuler et résoudre des problèmes ayant un grand nombre de contraintes
- Exemple : ordonnancement, planification, emplois du temps, affectation de ressources, séquençage de l'ADN, logistique...

Programmation par contraintes (2/4)

- Problème défini par :
  - Ensemble de variables $X = \{x_1, x_2, \ldots, x_n\}$
  - Ensemble de domaines $D = \{D(x_1), D(x_2), \ldots, D(x_n)\}$

    $D(x)$ ensemble fini des valeurs possibles pour la variable x
  - Ensemble de contraintes $C = \{C_1, C_2, \ldots, C_m\}$

    C_i contrainte exprimant une propriété qui doit être satisfaite par un ensemble de variables



Programmation par contraintes (3/4)

Problèmes de satisfaction de contraintes	Problèmes de satisfaction de contraintes avec optimisation
• Trouver une solution réalisable • Énumérer toutes les solutions possibles.	• Trouver la solution optimale

- Méthodes de résolution des problèmes :

Filtrage des variables | Propagation de contraintes | *Backtracking*

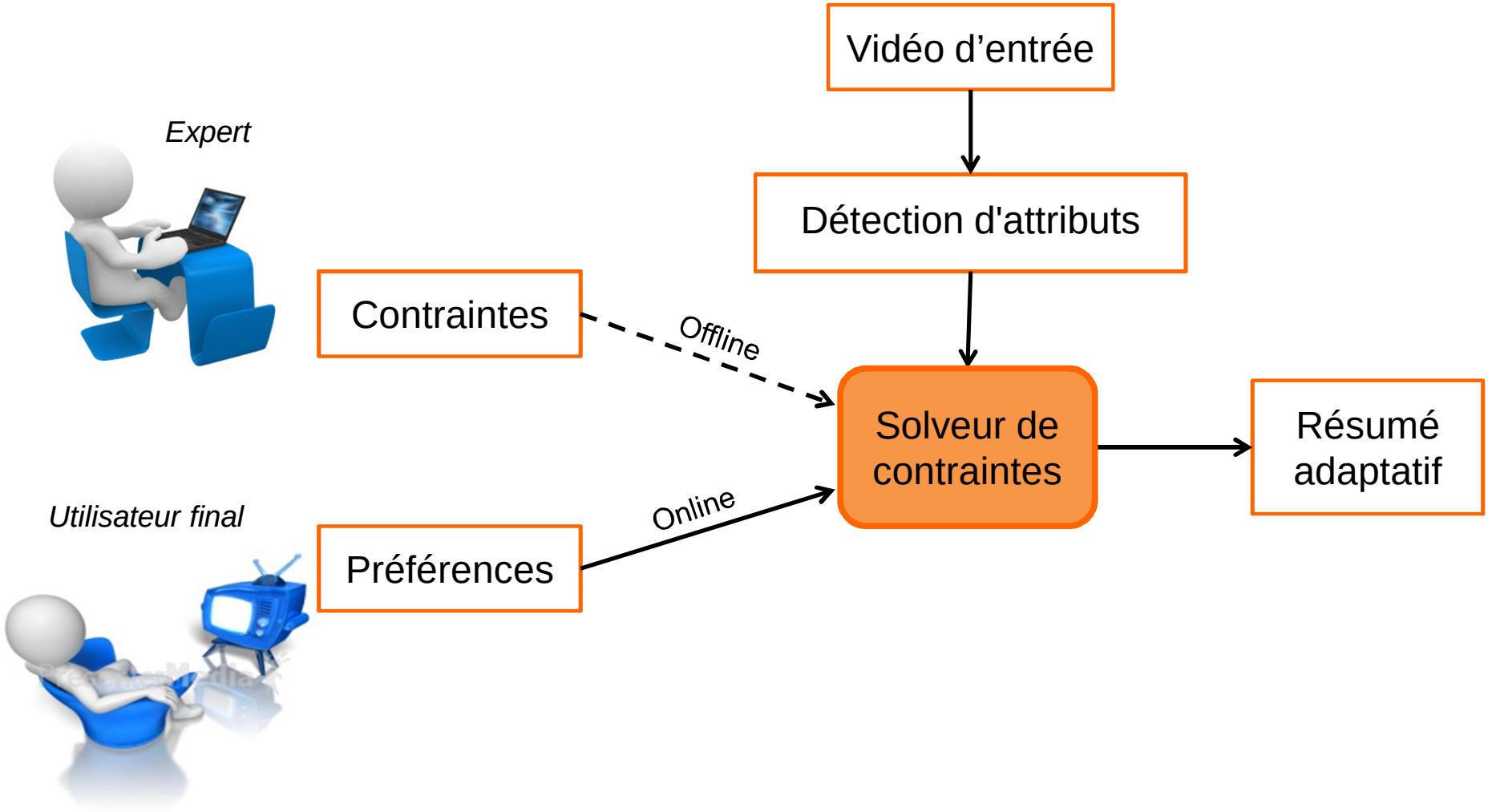
- Stratégies de résolutions :

Choix de variables | Choix de valeurs

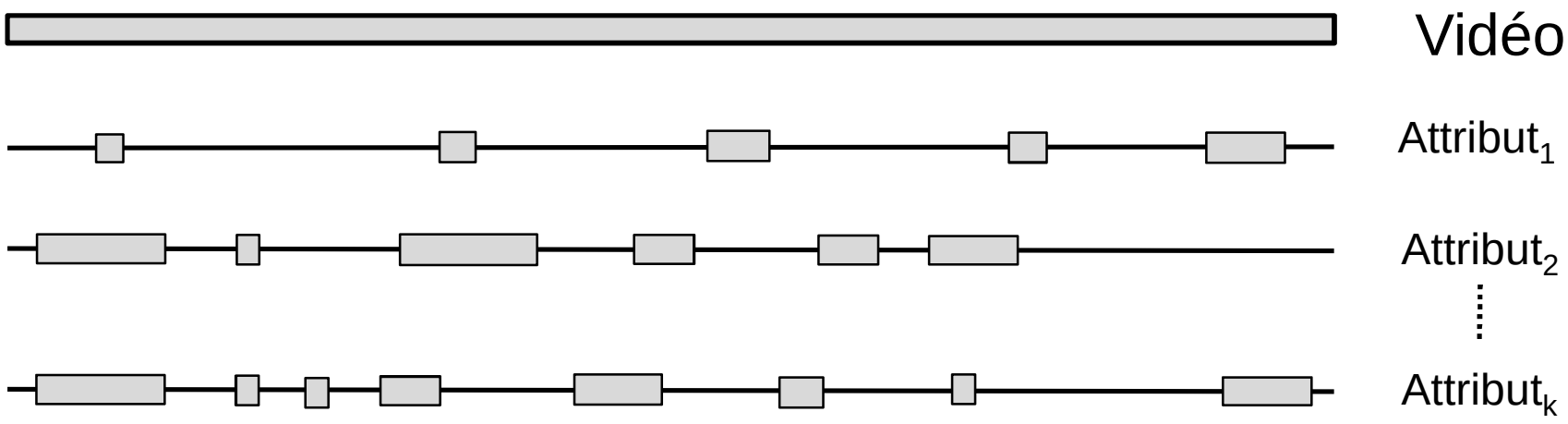
Programmation par contraintes (4/4)

- Exemples de solveurs de contraintes :
  - CHOCO (École des Mines de Nantes)
  - OR-tools (Google)
  - CPLEX (ILOG-IBM)
- Librairie Java libre pour les problèmes de satisfaction de contraintes
- CHOCO propose environ 80 contraintes basiques :
  - ✓ Contraintes arithmétiques classiques : equal, not equal, less or equal, greater or equal, …
  - ✓ Contraintes arithmétiques complexes : plus, minus, mult, sum, scalar, max, min, abs, …
  - ✓ Contraintes logiques: and, or, not, implies, ifOnlyIf, ...
  - ✓ Contrainte globales utiles: AllDifferent and BoundAllDifferent, GlobalCardinality and BoundGCC, AtMostNvalue, Cumulative, Occurence, Element, regular, ...
- Flexible: possibilité d'ajouter sa propre contrainte

Démarche


Vidéo d'entrée
Détection d'attributs
Expert
Contraintes
Offline
Solveur de contraintes
Résumé adaptatif
Utilisateur final
Online
Préférences

Démarche



Vidéo	Audio	Texte
• Luminosité et couleur (nombre, taille et position) • Suivi des visages • Mouvement des objets • Mouvement de la caméra • Détection d'objets • …	• Détection des applaudissements • Détection de silence • Détection de musique • Détection de parole • Clustering des locuteurs • …	• Transcription de parole • Détection de sous-titres • …

Démarche

- Exprimer les règles de production comme des contraintes sur les différents segments d'attributs sous forme de :

  - **Contraintes à satisfaire :** Décrivent ce que le résumé souhaité doit contenir et ce qu'il ne doit pas inclure: l'***obligatoire*** et l'***interdit***

    Les extraits sélectionnés contiennent des visages, précédent des applaudissements et ne contiennent pas de parole

  - **Fonction de coût à optimiser :** Maximise ou minimise certains critères : le ***souhaitable*** et le ***déconseillé***

    Les extraits maximisent la musique dans le résumé final

Démarche

Est-ce que la programmation par contraintes est utilisable pour créer des résumés ?

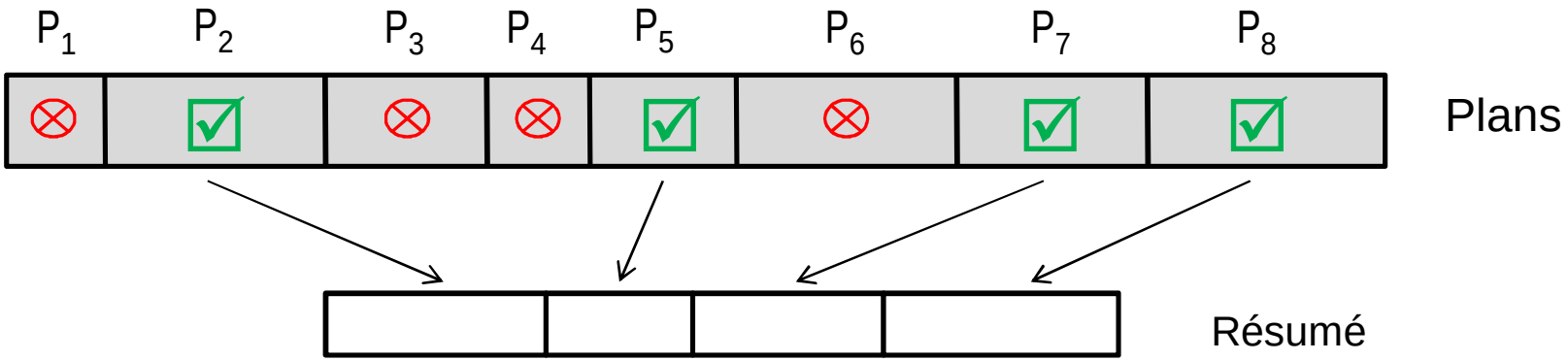
Trois modèles

Expressivité

Qualité des résumés

Modèle #1 : Principe et modélisation

- Une modélisation basée sur la sélection de plans.



Variables

$$Extraits = \{ extrait_i , i = 1..p \}$$

Domaines

$$extrait_i \in \{0,1\}$$

Modèle #1 : Formulation des contraintes

- Contrainte de durée du résumé

$$dmin \leq \sum_{i=1}^{p} duree\_plan_i * extrait_i \leq dmax$$

- Contrainte sur la durée minimale d'un extrait

$$(extrait_i * duree\_plan_i) \geq (extrait_i * duree\_plan\_min)$$

- Attribut non souhaité

$$\sum_{i=1}^{p} (plan\_Attribut_k[i] * extrait_i) = 0$$

- Présence d'un attribut

$$extrait_i \Rightarrow plan\_Attribut_k[i]$$

Modèle #1 : Discussion

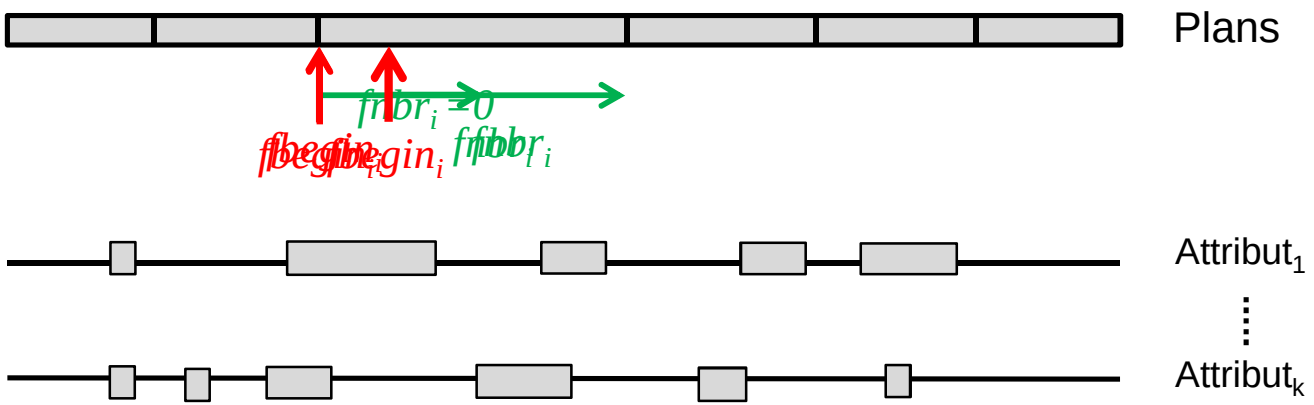
- Ce modèle est très simple à mettre en place
- Facilité d'exprimer les contraintes, opérateurs logiques
- Espace de recherche très réduit
- Exploration complète des domaines de recherche
- Temps de réponse très rapide

- Impossible d'exprimer toutes les contraintes (exp: plans juxtaposés)
- Extrêmement dépendant aux bordures des plans
- Ne profite pas pleinement des points forts de la PPC

Simple mais pas assez expressif

Modèle #2 : Principe

- Une modélisation basée sur la segmentation en plans
- Un résumé est un ensemble d'extraits (un extrait par plan)
  - Un plan entier
  - Une partie d'un plan
  - Rien n'est sélectionné d'un plan



Modèle #2 : Modélisation

- Une modélisation basée sur la sélection d'un extrait au sein du plan
- Cibler uniquement les parties intéressantes de la vidéo

Variables

$$Extraits \begin{cases} FDebut = \{ fdebut_i , & i = 1..p\} \\ FNbr = \{ fnbr_i , & i = 1..p\} \end{cases}$$

Domaines

$$fdebut_i \in \{debut\_plan_i .. fin\_plan_i\}$$
$$fnbr_i \in \{0 .. duree\_plan_i\}$$

Modèle #2 : Formulation des contraintes

- Contraintes de modélisation

$$fdebut_i + fnbr_i - 1 \leq fin\_plan_i$$

$$fnbr_i = 0 \Rightarrow fdebut_i = debut\_plan_i$$

- Contraintes globales

Durée du résumé : $dmin \leq \sum_{i=1}^{p} fnbr_i \leq dmax$

- Contraintes d'élagage

Élimination d'un attribut : $(fnbr_i * plan\_Attribut_k[i]) = 0$

Élimination des extraits courts : $(fnbr_i \geq duree\_extrait\_min) \vee (fnbr_i = 0)$

- Contraintes de voisinage

$$(fnbr_i \neq 0) \Rightarrow (plan\_Attribut_k[i + 1] \neq 0)$$

Modèle #2 : Formulation des contraintes

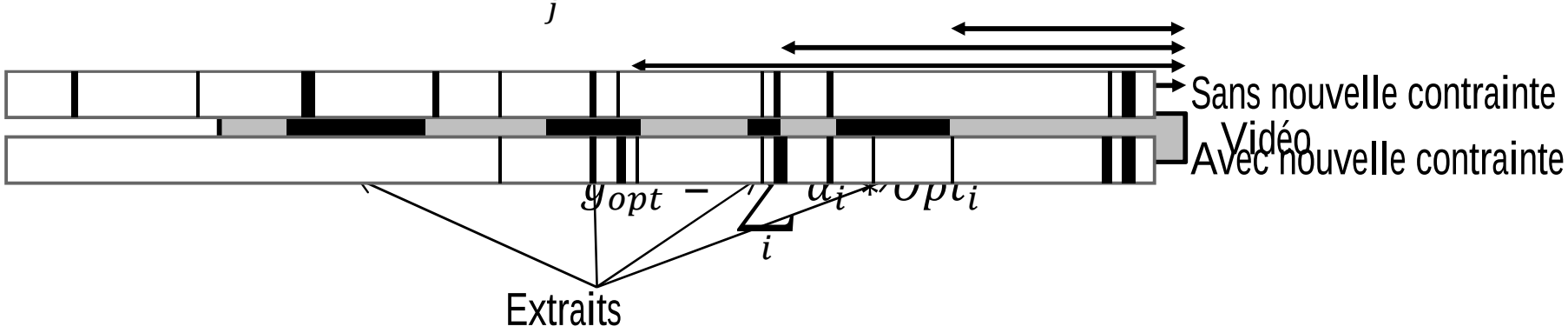
- Optimisation d'une fonction de coût

  - Maximiser ou minimiser la présence d'un attribut dans le résumé

$$opt_i = \sum Qte\_Attribut$$

  - Favoriser la sélection d'extraits à partir de la fin de la vidéo

$$opt_i = \sum_j (fin\_video - fin\_extrait_j)$$



Modèle #2 : Discussion

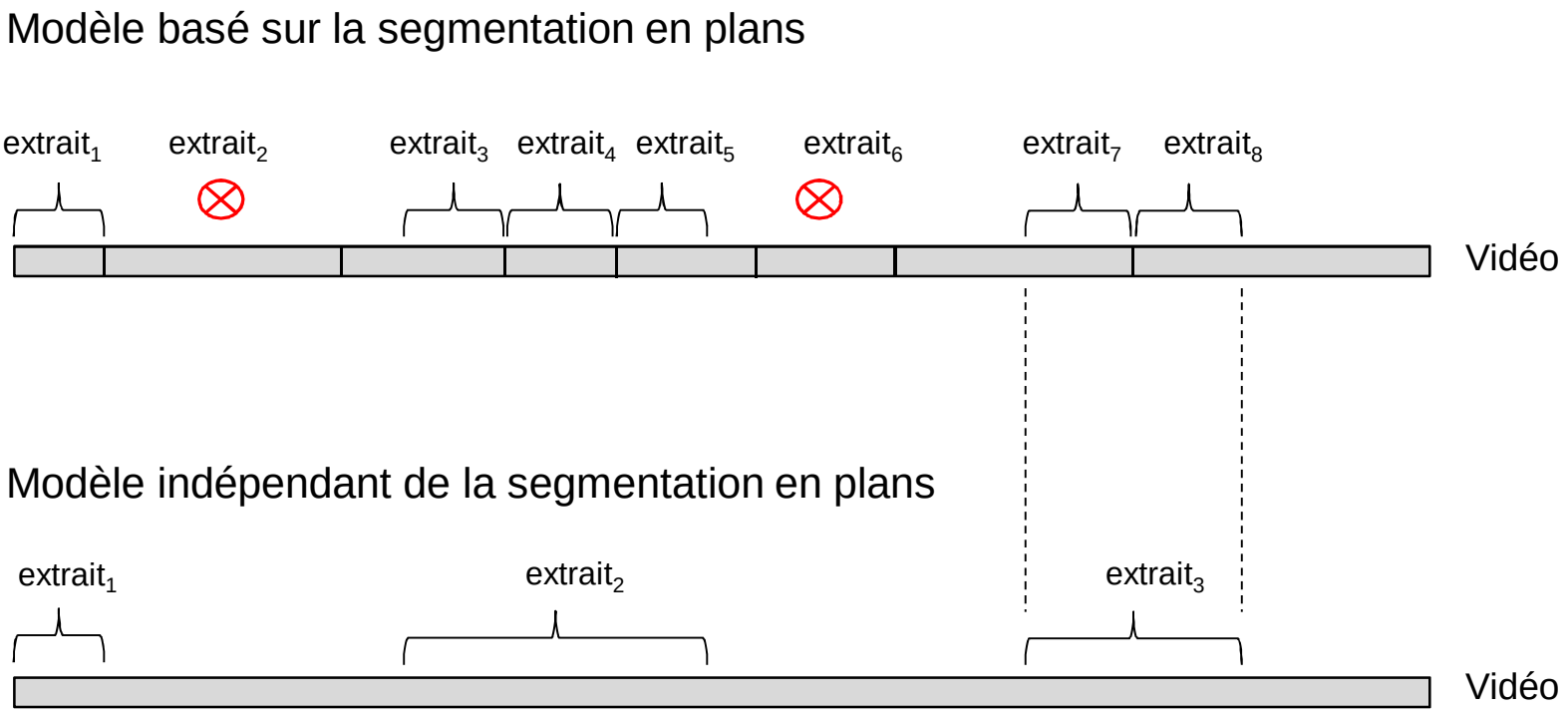
- Modèle flexible et expressif
  - ✓ Possibilité d'exprimer une large variété de contraintes

- Modèle très dépendant des frontières des plans
- Ne permettant pas d'exprimer quelques contraintes
  - ✓ Contraintes sur la durée minimale d'un extraits s'étalant sur plusieurs plans voisins

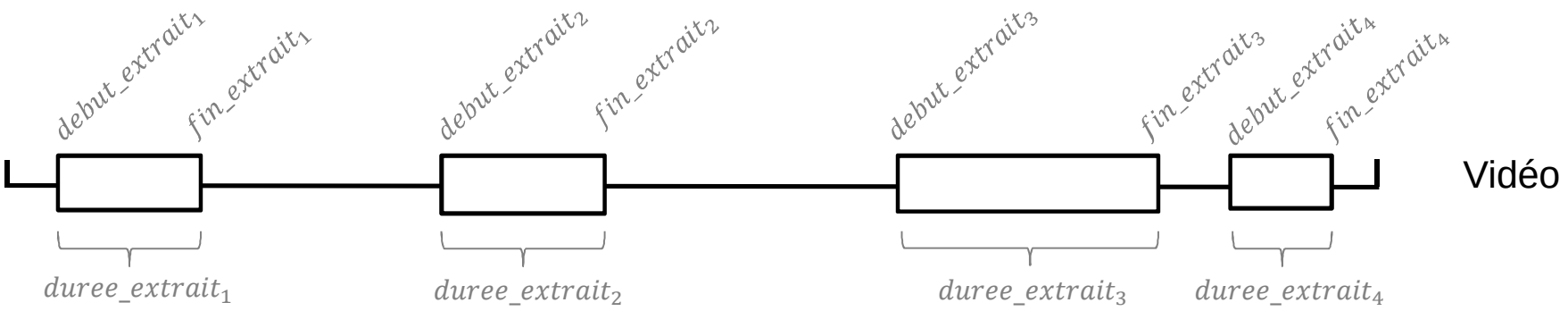
Plus flexible, plus complexe mais pas encore assez expressif

Modèle #3 : Principe

- Une modélisation qui s'affranchi complètement des frontières des plans
- La sélection d'extraits ne dépend d'aucune segmentation



Modèle #3 : Modélisation



Variables

$$Extraits \begin{cases} Debut\_extrait = \{ debut\_extrait_i , & i = 1..n\} \\ Fin\_extrait = \{ fin\_extrait_i , & i = 1..n\} \\ Duree\_extrait = \{ duree\_extrait_i , & i = 1..n\} \end{cases}$$

Domaines

$$debut\_extrait_i \in [debut\_video\ ..\ fin\_video\ - duree\_extrait\_min]$$

$$fin\_extrait_i \in [debut\_video + duree\_extrait\_\min\ ..\ fin\_video]$$

$$duree\_extrait_i \in [duree\_extrait\_\min\ ..\ duree\_extrait\_max]$$

Modèle #3 : Formulation des contraintes

- Contraintes de modélisation

➢ Dépendance intra-extrait :

$$fin\_extrait_i = debut\_extrait_i + duree\_extrait_i$$

➢ Dépendance inter-extraits :

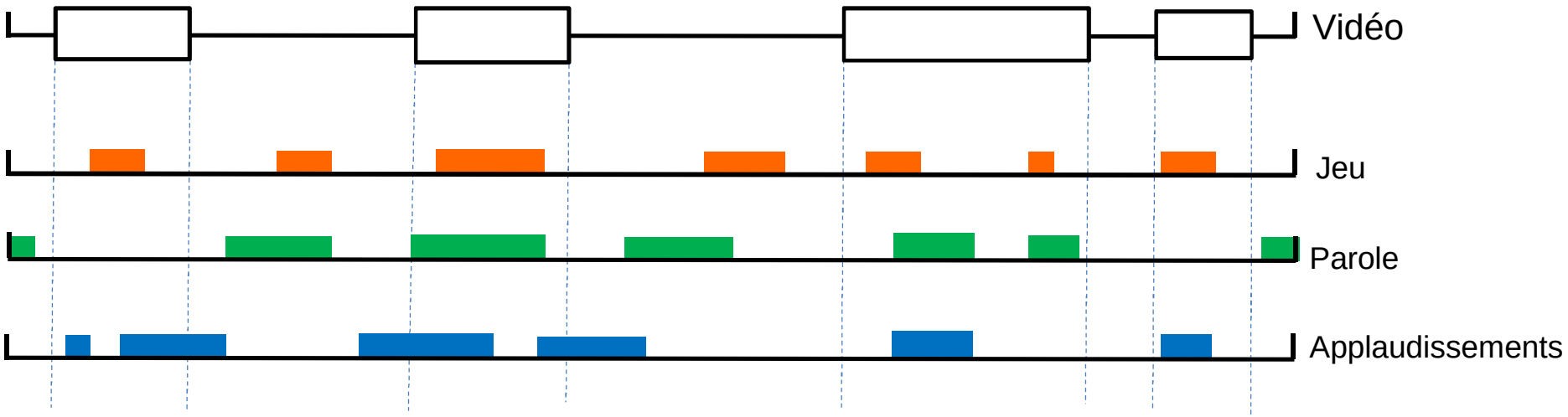
- Ordonnancement des extraits :

$$fin\_extrait_i < debut\_extrait_{i+1}$$

- Non chevauchement des extraits :

$$\left(debut\_extrait_i > fin\_extrait_j\right) \vee \left(fin\_extrait_i < debut\_extrait_j\right)$$

Modèle #3 : Formulation des contraintes



- **Exemple contrainte 1** : Chaque extrait sélectionné doit impérativement contenir un jeu (Contient jeu)
- **Exemple contrainte 2** : Le début et la fin de chaque extrait ne doivent pas correspondre à un segment de parole (Ne coupe pas parole)
- Calculer la durée totale de l'apparition des applaudissements dans les extraits (maximiser applaudissements dans le résumé)

Modèle #3 : Formulation des contraintes

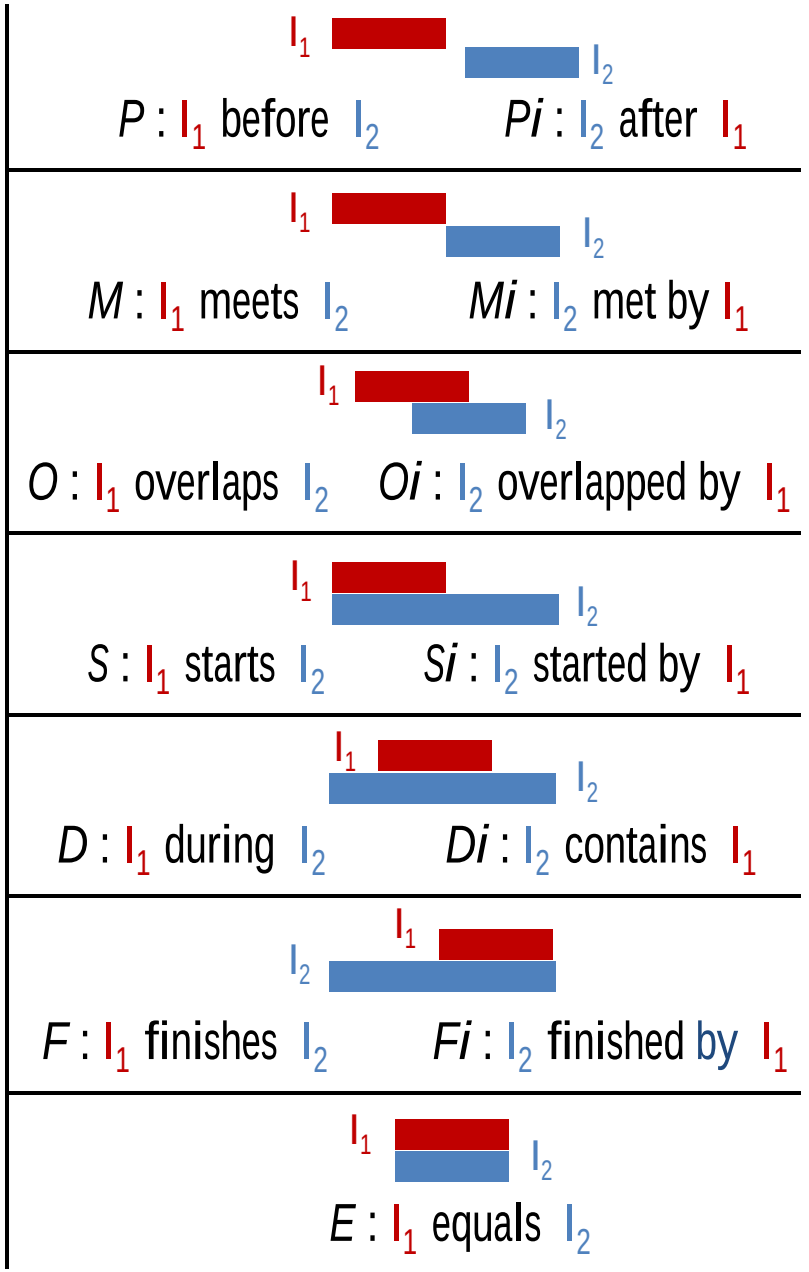
- L'algèbre d'Allen est une logique temporelle qui définie toutes les relations possibles entre deux intervalles
- Il décrit 13 relations atomiques disjonctives et conjonctives
- Problème :

Relations entre deux segments

⬇

Relations entre deux ensembles de segments

- Problème de combinatoire



Modèle #3 : Formulation des contraintes

- Contraintes élémentaires:

  Un extrait doit contenir un segment d'un attribut donné

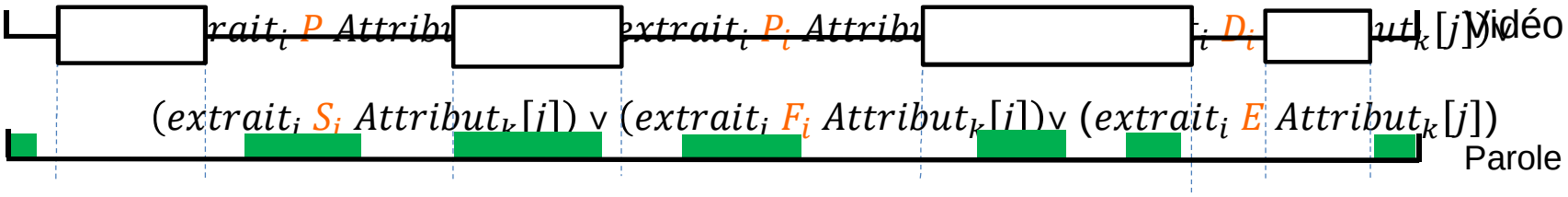
  $extrait_i \quad D_i \quad Attribut_k[j]$

  Un extrait doit commencer par un segment d'un attribut donné

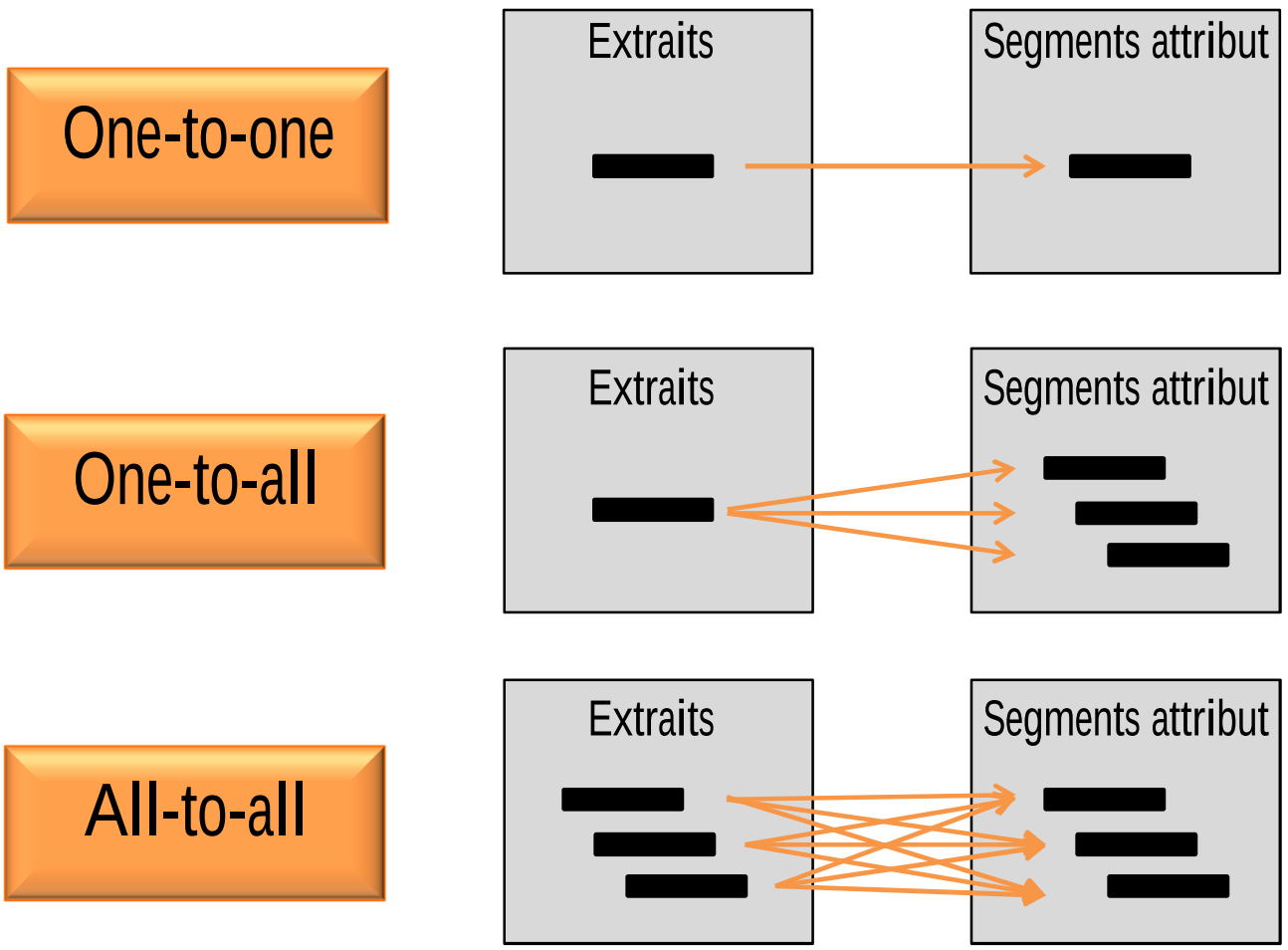
  $extrait_i \quad S_i \quad Attribut_k[j]$

- Contraintes complexes:

  Les extraits ne coupent aucun segment d'un attribut donné:

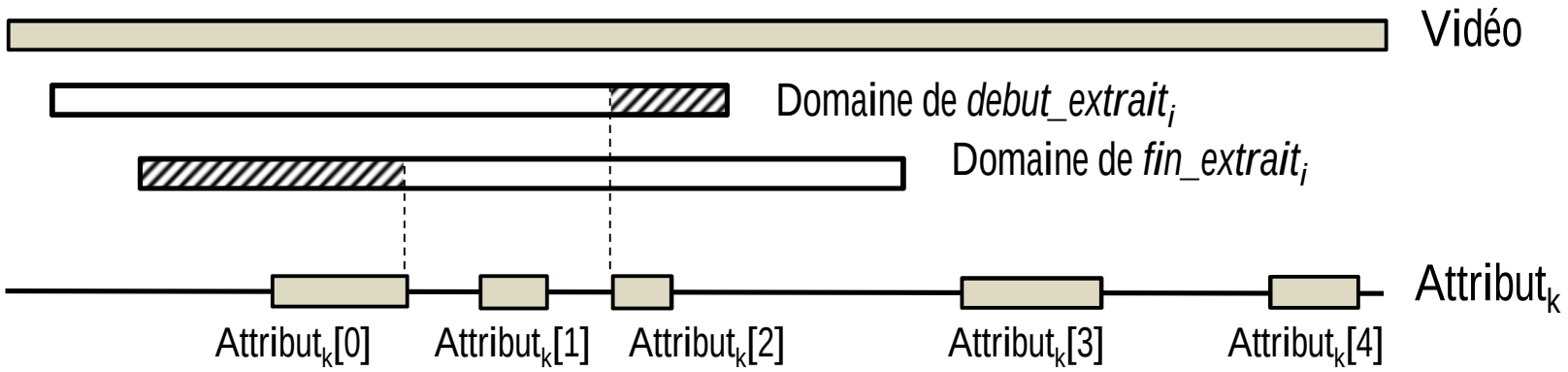


Modèle #3 : Formulation des contraintes

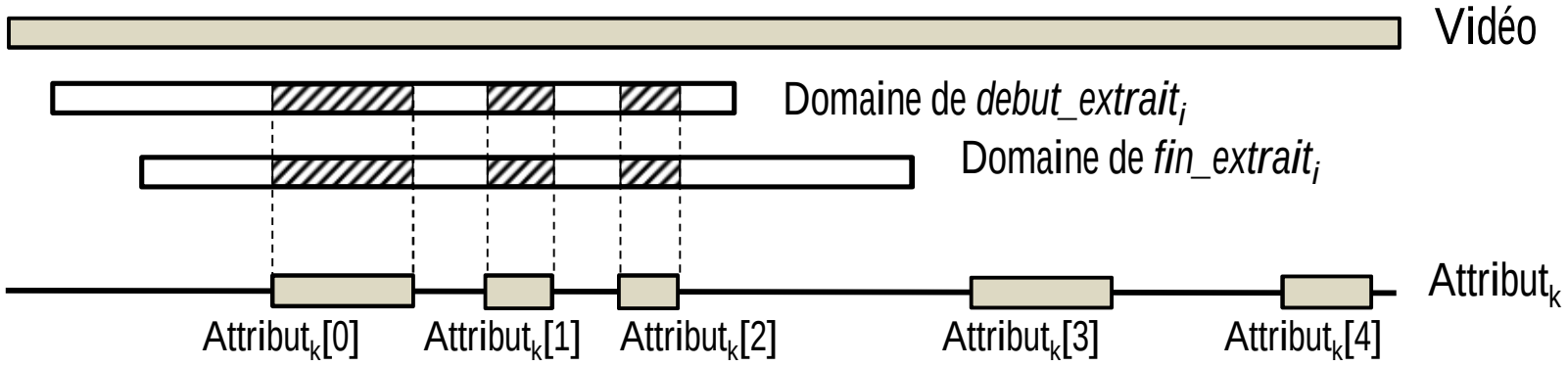


Modèle #3 : Formulation des contraintes

- La contrainte « contient »



- La contrainte « ne pas couper »



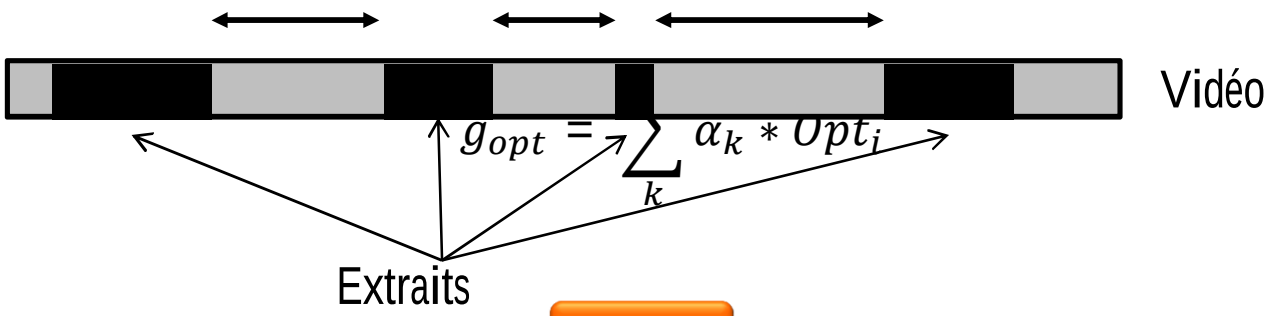
Modèle #3 : Formulation des contraintes

- Contraintes globales
  - Durée du résumé : $dmin \leq \sum_{i=1}^{n} duree\_extrait_i \leq dmax$
  - Présence d'un attribut : $Qte\_attribut_k \geq seuil$
- Optimisation d'une fonction de coût
  - Maximiser ou minimiser la présence d'un attribut dans le résumé

$$Opt_i = \sum Qte\_attribut_k$$

  - Favoriser la représentativité de la totalité de la vidéo dans le résumé

$$opt_i = \sum_j \sqrt{(debut\_extrait_{j+1} - fin\_extrait_j)}$$



Modèle #3 : Discussion

- Modèle plus flexible et expressif
- Possibilité d'exprimer toutes les relations temporelles entre les segments et les ensembles de segments

- Pousser CHOCO dans ses retranchements
- Introduire et implémenter de nouvelles contraintes

Mise en œuvre un peu plus complexe mais flexibilité et expressivité maximales

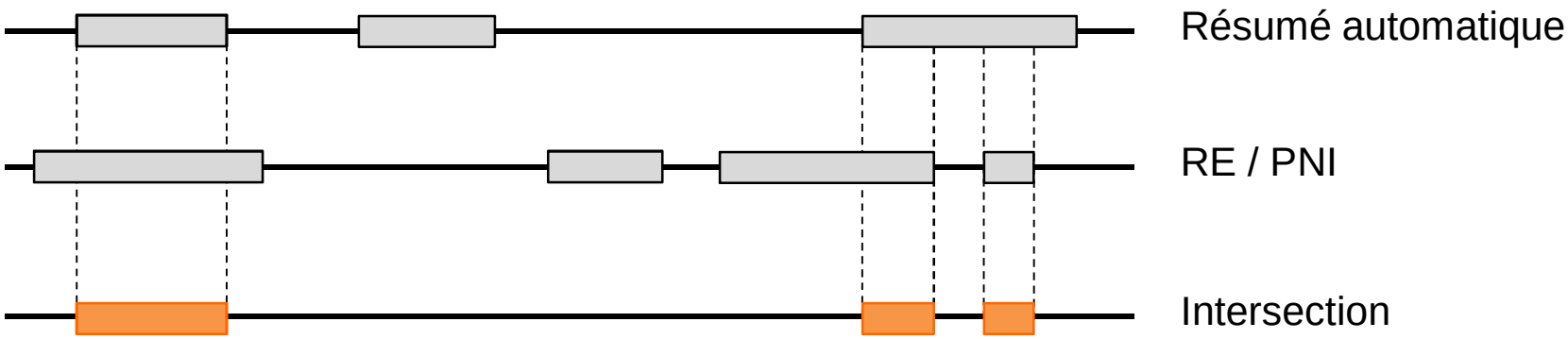
Évaluation

- Évaluation de la qualité des résumés générés automatiquement :
  - Une tâche délicate
  - Pas de définition objective → subjectivité
  - Dépend du type de la vidéo, de l'application cible et du spectateur

- Trois méthodes d'évaluation :
  - ✓ Métriques objectives
  - ✓ Description des résultats
  - ✓ Tests utilisateurs

Évaluation

- Évaluation sur des résumés de Matchs de Tennis
- Deux métriques objectives d'évaluation :
  - Capacité de sélectionner les parties intéressantes d'une vidéo :
    Intersection avec un résumé éditorial (RE) créé par des experts
  - Capacité d'éliminer les parties non-intéressantes (PNI) d'une vidéo :
    Intersection avec les PNI annotées par des volontaires



Évaluation

- Pourcentage de l'intersection entre le résumé éditorial (20 min) et le résumé automatique par rapport au résumé automatique :

	$Match_1$				$Match_2$			
Durée	4-5 min	9-10 min	14-15 min	19-20 min	4-5 min	9-10 min	14-15 min	19-20 min
RS *	1%	6%	9%	9%	9%	14%	16%	14%
Modèle #2	43%	32%	28%	21%	30%	28%	32%	33%
Modèle #3	51%	43%	37%	41%	56%	41%	**60%**	59%

- Pourcentage de l'intersection entre les parties non intéressantes (40 min) et le résumé automatique par rapport au résumé automatique :

	$Match_3$				$Match_4$			
Durée	4-5 min	9-10 min	14-15 min	19-20 min	4-5 min	9-10 min	14-15 min	19-20 min
RS	41%	50%	44%	40%	32%	46%	32%	33%
Modèle #2	15%	11%	13%	9%	10%	8%	15%	22%
Modèle #3	15%	10%	11%	**7%**	8%	8%	11%	11%

* RS : Sélection aléatoire de plans

Évaluation

- Pourcentage de l'intersection entre le résumé éditorial (3 min) et le résumé automatique (3 min) par rapport au résumé automatique :

	$Match_5$	$Match_6$	$Match_7$	$Match_8$	$Match_9$	$Match_{10}$	$Match_{11}$	$Match_{12}$
∩ Avec RE *	24%	**45%**	21%	17%	33%	21%	41%	18%

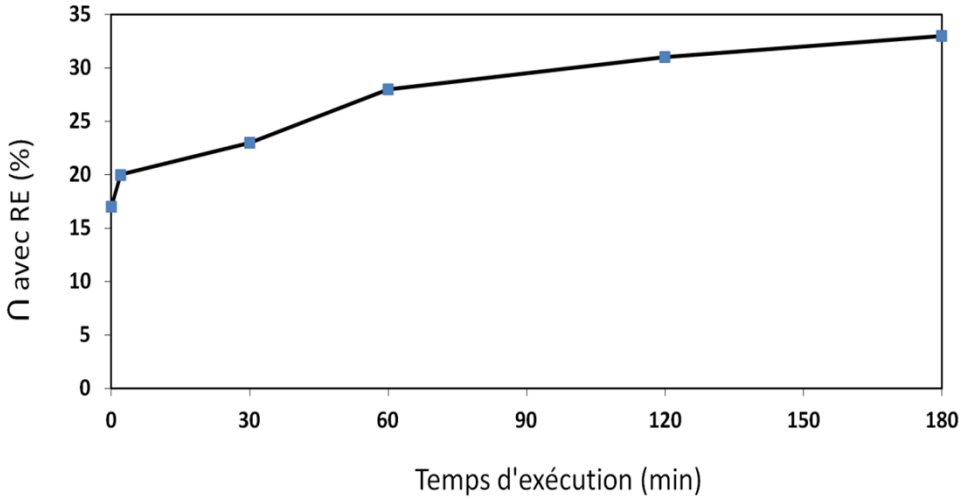
- Impact de la correction de la détection des attributs sur la qualité des résumés générés :

Durée	5 min		10 min		15 min		20 min	
Modèle	avant	après	avant	après	avant	après	avant	après
∩ Avec RE *	56%	66%	**41%** ↗	**68%**	60%	62%	59%	67%
∩ Avec PNI *	**15%** ↘	**4%**	10%	6%	11%	5%	7%	6%

* RE : Résumé éditorial
* PNI : Parties non intéressantes

Évaluation

- Impact du temps d'exécution et de l'utilisation de seuils sur la qualité du résumé :



	Sans la nouvelle contrainte	Avec la nouvelle contraintes
temps	180 min	2 min
∩ Avec RE *	33%	33%

- Comparaison des temps pris par le solveur pour retourner une première solution :

	1re solution retournée avec Modèle #2	1re solution retournée avec Modèle #3
Match	1 minute et 27 secondes	1 seconde

* RE : Résumé éditorial

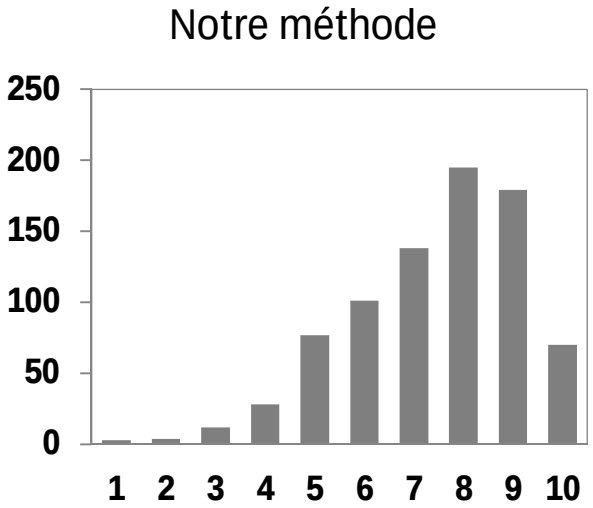
Évaluation

- Évaluation en ligne par des volontaires
- Remplir un formulaire et attribuer une note
- Différents résumés présentés aléatoirement aux évaluateurs :
  - Résumés éditoriaux
  - Résumés générés automatiquement (modèle #3)
  - Résumés générés par des méthodes basiques
- Compagne d'évaluation sur 5 semaines :
  - ✓ 61 évaluateurs, 1096 évaluations
  - ✓ Chaque résumé : 6,85 fois
  - ✓ Chaque évaluateur : 18 résumés

Évaluation

	Moyenne (μ)	Écart-type (σ)
Résumés éditoriaux	8,12	1,56
Notre méthode	7,42	1,75
Résumés basiques	1,65	2,46

- Les résumés éditoriaux sont légèrement meilleurs que ceux générés par notre méthode



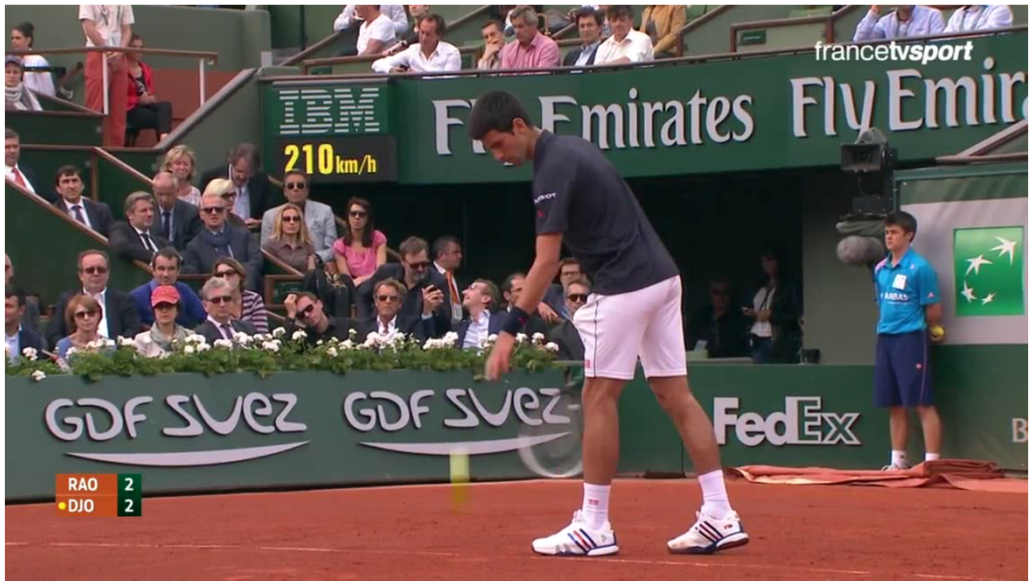
Conclusion et Perspectives

- Une approche nouvelle de création automatique de résumés vidéo
  - ✓ Programmation par contraintes
  - ✓ Évolution du solveur CHOCO : ajout de nouvelles contraintes
- Avantage principal: séparation entre la modélisation et la résolution du problème
- Trois modèles différents
  - Performance
  - Efficacité et expressivité
  - Qualité des résumés résultants
- Une évaluation de la qualité des résumés générés automatiquement
  - ➢ Métriques objectives
  - ➢ Description des résultats
  - ➢ Tests utilisateurs

Conclusion et Perspectives

- Expérimentations supplémentaires sur d'autres types de contenus vidéo
- Prise en compte la structure des vidéos dans le processus de création des résumés
- Prise en compte de la transcription de la parole
- Résumé d'un ensemble de vidéos
- Proposer un langage de haut niveau permettant aux utilisateurs de spécifier leurs contraintes sans avoir des connaissances en programmation
- Effets de montage: simple extraction à un aspect plus éditorial

Merci pour votre attention



Match : 3h30 résumé : 3 minutes