

Une machine à raconter des histoires ?

Analyse et modélisation des processus de ré-éditorialisation de vidéos

Encadrement :

Georges Linarès (PR), Serigne Gueye (MCF)

Damien Malinas (MCF)

Unités de recherche impliquées :

LIA (EA 4128)

Centre Norbert Elias, Equipe *Culture et Communication* (UMR 8562)

Contexte et objectifs

Les recherches récentes en traitement automatique de documents multimédias se sont majoritairement concentrées sur les problèmes d'accès aux contenus numériques, de structuration de grandes collections, d'indexation automatique. L'idée que le numérique pourrait permettre de ré-utiliser des contenus et de produire une *œuvre* à partir de matériaux eux-mêmes multimédias est plus récente : il s'agit d'une pratique artistique courante, qui s'est répandue dans les milieux *vidéophiles*, dans des contextes assez variés : production de bandes annonces, de résumés, de portails d'accès à des collections, etc.

Ce principe d'une composition par assemblage de fragments a été par ailleurs assez largement utilisé en traitement automatique du langage, pour la génération de résumés automatiques ou, moins directement, en recherche d'information.

En résumé texte par extraction, il s'agit de produire un texte concis résumant un ou plusieurs documents sources par extraction puis assemblage de segments de textes. L'extension de ce principe à la vidéo a été introduite plus récemment [Crozat & al 2012, Ercolessi 2012, Rouvier, 2011]. Bien que cette approche ait suscité de très nombreux travaux en traitement du langage naturel, certains des problèmes de fond restent ouverts, relatifs à la définition des fonctions d'intérêt, aux méthodes d'optimisation de la couverture des sources par le résumé ou encore, et c'est une question qui a été peu abordée dans la littérature, à la structure narrative du texte qui devrait être généré.

En recherche d'information, le document produit par le système est un ensemble de pointeurs sur des sources supposées répondre à la requête utilisateur. Ainsi, un moteur de recherche doit sélectionner les documents les plus pertinents pour cette requête, produire une vue suffisamment informative des résultats pour permettre la décision de consultation, offrir à l'utilisateur une image synthétique des contenus dans un espace visuel limité. L'approche généralement suivie consiste à présenter l'ensemble des documents les plus pertinents comme la suite ordonnée des extraits de textes les plus proches de la requête.

La plupart des travaux réalisés sur ces sujets reposent sur l'analyse de contenus sémantiques, des cibles applicatives précises et un cadre expérimental relativement bien balisé. Les contenus culturels multimédias posent des problèmes spécifiques *a priori* plus difficiles à appréhender, notamment parce qu'ils intègrent des dimensions subjectives, liées à la perception ou à la sensibilité des auteurs et des spectateurs. Ces dimensions sortent des cadres bien définis dans lesquels les systèmes d'analyse des contenus sont habituellement conçus [Torres-Moreno, 2011].

La ré-éditorialisation ajoute à ces difficultés relatives à la nature des objets à traiter celle du processus de construction d'un objet composite destiné à donner une image fidèle de contenus sources, à raconter une histoire, à susciter l'envie de voir le document dans son intégralité [Bouchardon & al, 2011].

A quel point ce processus peut-il être reproduit, quelles sont les fonctions d'intérêt sous-jacentes, les mécanismes mis en œuvre dans la sélection des fragments et leur assemblage ?

L'étude proposée abordera ces questions par le biais de la modélisation ou de la simulation du processus de ré-éditorialisation tel qu'il peut être mis en œuvre par des acteurs humains. Ces modèles seront considérés à la fois comme les bases sur lesquelles des systèmes de ré-éditorialisation pourraient être construits et comme des outils d'analyse des pratiques de production et de réception des œuvres.

Approches

Les modèles proposés pourront s'appuyer sur l'extraction de descripteurs de différents niveaux (linguistiques et extra linguistiques) des transcriptions des contenus parlés, des fonctions d'intérêts multiples et la représentation en graphes de segments vidéo [Garg et al., 2009].

Ces trois aspects sont étroitement liés. Les fonctions d'intérêt doivent évaluer l'intérêt d'un segment de vidéos à partir des descripteurs. Il s'agit à la fois de l'intérêt individuel du fragment considéré et de l'intérêt de son inclusion dans l'ensemble qui formera le document cible. Dans [Rouvier & al, 2013, Rouvier 2011], nous avons identifié trois types des descripteurs représentant la saillance d'un fragment, son expressivité (dynamique et intensité de l'expression, charge émotive) et le fond sémantique des contenus, notamment des contenus parlés. Ces descripteurs relatifs au fond, à la forme et à la singularité d'un fragment pourront être étendus, contextualisés, ou ré-évalués en fonction des intentions ou du contexte de la ré-édition.

Le processus de composition, tel qu'il est généralement formulé dans les méthodes de résumé texte par extraction, repose sur la sélection de segments qui, regroupés, maximisent l'information contenue en minimisant la redondance. C'est un problème de couverture d'ensembles, assez classique en optimisation combinatoire et qu'on sait résoudre dans le cas particulier où la valeur informative d'un groupement est la somme des valeurs individuelles de chaque segment. Ceci n'est malheureusement pas le cas ici et les techniques standards d'optimisation évoquées plus haut ne s'appliquent plus directement. Ce problème a été abordé de différentes manières, par l'usage d'algorithmes gloutons ou d'algorithmes de programmation linéaire en nombres entiers [Garg et al., 2009].

Ces approches devront être revisitées ou reconsidérées dans le cas de documents vidéo, dont la complexité est d'un autre ordre, les fonctions d'intérêt plus complexes et la structure narrative ciblée issue de modèles plus explicites que ceux implantés d'ordinaire dans les systèmes de résumé texte.

Enfin, la simulation est un outil qui peut permettre d'évaluer des hypothèses d'analyse sociologique et, éventuellement, d'en susciter de nouvelles : les méthodes développées, tout en s'appuyant sur la connaissance des publics, des pratiques de ré-appropriation, du processus de création et de la réception des œuvres, permettront en retour de l'enrichir.

Cadre expérimental

Le travail expérimental sera mené sur les données collectées sur des sites dans lesquels des communautés de spectateurs partagent des documents qu'ils produisent, des avis, des commentaires. Ce travail de mise en place expérimental, d'identification des communautés et de leurs pratiques a été amorcé lors du stage de Master de Paul Lecomte, étudiant du Master *Publics de la culture et communication* de l'UAPV, qui est co-encadré par D. Malinas et Georges Linarès. La thèse proposée pourra s'appuyer sur cette première collaboration, notamment en s'intéressant à trois types de documents pour lesquels du matériel expérimental est dorénavant et déjà disponible : les bandes annonces, les résumés de séries télévisées et les *Mashups*, qui sont des objets d'intérêt grandissant dans des communautés très actives sur la toile.

Bibliographie

- [Berger, 2006] Peter L. Berger, Invitation à la sociologie, Paris, La Découverte, 2006.
- [Bouchardon & al, 2011] Bouchardon S., Cailleau I., Crozat S., Bachimont B., Hulin T. (2011). Explorer les possibles de l'écriture multimédia. Les Enjeux de l'information et de la communication, dossier ""Le(s) multi-média(s)".
- [Boudin & al, 2008] Florian Boudin and Torres Juan-Manuel and El Bèze Marc A Scalable MMR Approach to Sentence Scoring for Multi-Document Update Summarization COLING, 2008
- [Crozat & al, 2012] Crozat S. (2012). Structured and fragmented content in collaborative XML publishing chains. Proceedings of DocEng 2012 "12th edition of the ACM Symposium on Document Engineering ", Paris, France.
- [Ercolessi] [Philippe Ercolessi](#), [Hervé Bredin](#), Christine Sénac: StoViz: story visualization of TV series. [ACM Multimedia 2012](#): 1329-1330, New York, NY, USA ©2012
- [Ethis, 2006] Les Spectateurs du temps. Introduction à une sociologie de la réception des œuvres filmiques, L'Harmattan, Collection Logiques sociales, Paris, 2006.
- [Ethis al, 2008] Avignon, le public participant en collaboration avec Jean-Louis Fabiani et Damien Malinas (Éditions L'entretemps-La Documentation française), Paris.
- [Ethis 2009] Sociologie du cinéma et de ses publics (Armand Colin- Sociologie 128 - nouvelle édition revue et refondue), Paris, 2009.
- [Favre, 2007] B. Favre, 2007. Résumé automatique de parole pour un accès efficace aux bases de données audio. Ph. D. thesis, LIA.
- [Garg et al., 2009] N. Garg, B. Favre, K. Reidhammer, & D. Hakkani-Tür, 2009. ClusterRank : A Graph Based Method for Meeting Summarization. In Proc. Of Conference of the International Speech Communication Association.
- [Goldstein & al, 2000] J. Goldstein, V. Mittal, J. Carbonell, and J. Callan. 2000. Creating and evaluating multi-document sentence extract summaries. CIKM 2000.
- [Goffman, 1975] Erving Goffman, Stigmate, les usages sociaux des handicaps, Paris, Les Éditions de Minuit, 1975.
- [Jauss & al, 1978] Hans Robert Jauss, Pour une esthétique de la réception, Gallimard, Paris, 1978.
- [Jullier & al, 2008] Laurent Jullier, Jean-Marc Leveratto, La leçon de vie dans le cinéma hollywoodien, Paris, Vrin, 2008.
- [Lecomte, 2012] Paul Lecomte, « Le mashup film possède-t-il une identité générique ? » Rapport de Master 1, université d'Avignon et des Pays de Vaucluse, 2012
- [Metz, 1984] Christian Metz, Le Signifiant imaginaire, Paris, Christian Bourgois Éditeur, 1984.
- [Rouvier & al, 2013] Mickael Rouvier, Georges Linarès, Benoit Favre, Bernard Merialdo. Searching segments of interest in single story web-videos In International Workshop on Image and Audio Analysis for Multimedia Interactive Services, WIAMIS 2013, Paris, 2013.
- [Rouvier & al, 2009] Mickael Rouvier, Georges Linarès, Driss Matrouf Robust Audio-based Classification of Video Genre In Conference of the International conference of the Speech Communication Association, InterSpeech, ISCA (ed.), Brighton, UK, 2009.
- [Rouvier 2012] Mickael Rouvier, « Structuration de contenus audio-visuels pour le résumé automatique », thèse de l'Université d'Avignon et des Pays de Vaucluse, 2012.
- [Torres Moreno & al, 2011] Juan Manuel Torres-Moreno. Résumé automatique de documents - une approche statistique. Hermès-Lavoisier, 2011, ISBN 978-2-7462-3212-9
- [Xie et al., 2009] S. Xie, B. Favre, D. Hakkani-Tür, & Y. Liu, 2009a. Leveraging Sentence Weights in Concept-based Optimization Framework for Extractive Meeting Summarization. In Proc. of Conference of the International Speech Communication Association, InterSpeech'09, Brighton, UK
- [Yingbo & al, 2011] J. Yingbo, B. Merialdo, M. Rouvier, and G. Linarès. Static and dynamic video summaries. In ACM Multimedia Conference, Scottsdale, Arizona, USA , 2011.