

Système de recherche de musique adaptable à la perception de chaque utilisateur

Najeh HAJLAOUI

Laboratoire LIG, BP 53
385, Rue de la Bibliothèque F-38041 Grenoble cedex 9
Najeh.Hajlaoui@imag.fr

1 Contexte et objectifs

Dans le cadre de nos travaux sur le portage linguistique des systèmes de gestion de contenu traitant des énoncés spontanés en langue naturelle, nous présentons ici une évaluation du portage d'IMRS (système de recherche de morceau de musique en langue naturelle) Kumamoto (2007) du japonais vers le français. Cette évaluation peut se faire au niveau des représentations internes en les comparant, ou au niveau de la tâche.

Ici, nous nous intéressons à une évaluation liée à la tâche en proposant un service Web qui permet de mesurer la performance globale de la nouvelle version obtenue. Nous avons par la suite cherché à améliorer et ajouter de nouvelles fonctionnalités en proposant un service de recherche de musique adaptable à la perception de chaque utilisateur. En effet, un même morceau de musique peut être jugé calme pour un premier auditeur, très calme pour un deuxième, et assez calme pour un troisième, etc. On se demande l'impression finale que porte ce dernier morceau de musique. C'est naturel que les utilisateurs évaluent différemment un même morceau de musique car ils ont des perceptions différentes. Devant cette situation, nous proposons un service de recherche de musique basé des méthodes simples et automatisées et qui sont adaptables à la perception de chaque utilisateur.

2 IMRS (Impression-based Music-Retrieval System)

IMRS est un service Web prototype, non opérationnel, développé au Japon à titre de recherche, et permettant d'obtenir des morceaux de musique au moyen de requêtes en japonais, ou bien en manipulant une interface graphique présentant 10 critères réglables par des cases à cocher (vitesse, niveau sonore, rythme...).

Dans IMRS, les énoncés traités par le système sont des phrases ou fragments de phrase spontanés qui sont transformés en vecteurs de requête. Chaque vecteur est composé de dix composantes. La valeur d'une composante est un nombre réel entre 1 et 7. Le symbole particulier « *nil* » veut dire « *don't care* ».

Les dix critères sont *Quiet-Noisy*, *Calm-Agitated*, *Refreshing-Depressing*, *Bright-Dark*, *Solemn-Flippant*, *Leisurely-Restricted*, *Pretty-Unattractive*, *Happy-Sad*, *Relax-Arouse*, *The mind is restored-The mind is vulnerable*. Par exemple, le composant « *Happy – Sad* » est

IMRSg

caractérisé par sept valeurs ou impressions intermédiaires : « *very happy* », « *happy* », « *a little happy* », « *medium* », « *a little sad* », « *sad* », et « *very sad* ». Ces termes dénotent des impressions correspondant respectivement aux valeurs 7, 6, 5, 4, 3, 2, et 1. Une requête pour trouver un morceau de musique donnant une impression de bonheur (*happy*) correspond alors au vecteur à dix dimensions (*nil nil nil nil nil nil 6,0 nil nil*) Kumamoto (2007).

Malheureusement, le code n'est pas accessible, et la documentation existante ne décrit que la représentation interne d'IMRS, à savoir ces vecteurs de dimension dix. Les articles publiés contiennent cependant des scénarios et des images d'écran, à partir desquels nous avons pu imaginer le reste de l'architecture du système, à savoir une base de données des morceaux de musiques (métadonnées, fichiers son et vecteurs), un composant de gestion des requêtes sous forme interne, et un autre pour l'extraction de contenu à partir des requêtes en langue naturelle (ici le japonais).

3 IMRSg

Nous avons construit en Php/MySQL un service Web dit IMRS-g réimplémentant aussi exactement que possible le système IMRS, le généralisant, et le personnalisant. N'ayant pas de compétences suffisantes en japonais, nous avons remplacé le japonais par le français. Nous avons aussi adapté la partie traitement de la langue à l'anglais et à l'arabe.

Nous avons ensuite généralisé la représentation interne en ajoutant d'autres critères de recherche, non exprimables par le même type de vecteurs et en utilisant les ensembles flous.

L'extracteur de contenu traite des énoncés simples et en extrait si possible les 10 premières composantes d'un vecteur de type IMRS, contenant donc un réel de l'intervalle [1,7], et les informations supplémentaires, sous forme d'éléments ou de listes d'éléments.

La demande d'un morceau de musique peut-être faite directement, à travers une interface graphique permettant de manipuler un vecteur à dix dimensions. L'interface permet aussi de taper une requête en langue naturelle, dont l'interprétation apparaît comme un vecteur. Comme réponse, l'utilisateur reçoit des liens vers des morceaux de musique répondant à ses critères de choix. Il lui suffit de cliquer sur un lien pour entendre le morceau correspondant.

La représentation initiale de IMRS semble trop rigide pour exprimer des énoncés comme "un peu doux", ou pour faire évoluer la requête courante par un énoncé comme "un peu moins rapide". Même si l'on convient que chaque terme correspond à un intervalle de longueur 1 centré sur sa valeur de référence, par exemple 5,5..6,5 pour *happy* (et 6,5..7 pour *very happy*). C'est pourquoi nous avons étudié la possibilité de mieux modéliser et traiter les requêtes en faisant appel à la logique floue Zadeh (1965). Afin de raisonner à partir de connaissances imparfaites difficiles à utiliser dans le cadre de la logique classique, la logique floue propose de remplacer les variables booléennes, utilisées dans la logique classique, par des variables floues, à valeurs dans l'intervalle réel [0,1], et les ensembles classiques par des ensembles flous.

En faisant appel à la logique floue, on pourra ainsi dire qu'un morceau de musique est « rapide » si son tempo moyen est de 100 à la noire, ou s'il est dans l'intervalle [84,112]. On pourra aussi associer un ensemble flou de tempi à cette valeur « rapide », et l'appeler « impression de rapidité ». Voici un exemple qui montre comment l'impression de rapidité peut être caractérisée par une courbe représentant la « fonction d'appartenance » des fréquences à cet ensemble flou. Cette courbe, ou de façon équivalente l'impression de rapidité, peut être différente d'une personne à une autre.

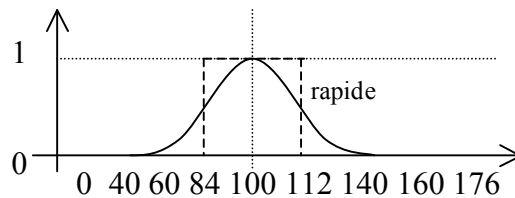


FIG. 1 – Courbe de l'impression de rapidité.

IMRS-g généralise IMRS par plusieurs aspects. Le plus intéressant nous semble être l'utilisation de la logique floue pour mieux traiter les "impressions" des utilisateurs, et les modificateurs tels que *plus, un peu plus, beaucoup plus, moins...*

Nous avons eu plus d'une trentaine d'utilisateurs participants à la campagne d'évaluation. En compensation, chaque utilisateur a pu écouter des morceaux de musique adaptés à sa façon de percevoir et à ses goûts.

Conclusion. Nous avons construit un système permettant d'effectuer une évaluation liée à la tâche (*end to end*) en effectuant la réingénierie du prototype IMRS et en l'étendant en IMRS-g. C'est un site Web qui permet de faire participer des personnes diverses à l'évaluation d'un ensemble de morceaux de musique en leur proposant de rechercher leur musique avec différents modes possibles. On peut essayer IMRS-g à l'adresse suivante : <http://www-clips.imag.fr/geta/User/najeh.hajlaoui/Musique/>.

Références

Kumamoto, T., (2007). *A Natural Language Dialogue System for Impression-based Music-Retrieval*. Proc. CICLING-07 (Computational Linguistics and Intelligent Text Processing). Mexico. February 12-24, 2007. 12 p.

Zadeh, L. A. (1965). *Fuzzy sets*. Information and Control 8: pp. 338-353.

Summary

The paper present a Web service which adapts the internal representation of a request to find a piece of music to the perception of each user. Indeed, the same piece of music can be judged as quiet by a first listener, and very quiet by another. For this, we transform the internal representation using "fuzzy"sets.