

Un système automatique de sélection de réponse en domaine ouvert intégrable à un système de dialogue social

Franck Charras¹ Guillaume Dubuisson Duplessis¹ Vincent Letard² Anne-Laure Ligozat³ Sophie Rosset¹

(1) LIMSI, CNRS, Université Paris-Saclay, F-91405 Orsay

(2) LIMSI, CNRS, Univ. Paris-Sud, Université Paris-Saclay, F-91405 Orsay

(3) LIMSI, CNRS, ENSIIE, Université Paris-Saclay, F-91405 Orsay

{charras, gdubuisson, letard, annlor, rosset}@limsi.fr

RÉSUMÉ

Cette démonstration présente un système de dialogue en domaine ouvert qui utilise une base d'exemples de dialogue automatiquement constituée depuis un corpus de sous-titres afin de gérer un dialogue social de type « chatbot ».

ABSTRACT

An automatic open-domain response selection system integrable to a social dialogue system

This demonstration exhibits an example-based system that uses a database of indexed dialogue examples automatically built from a television drama subtitle corpus to manage social open-domain dialogue.

MOTS-CLÉS : Modèle de dialogue basé sur des exemples ; Système de dialogue en domaine ouvert.

KEYWORDS: Example-based dialogue modelling ; Open-domain dialogue system.

1 Description générale du système

Ces travaux s'inscrivent dans le cadre du projet Joker qui vise à développer une interface utilisateur intelligente générique fournissant un système de dialogue multimodal démontrant des compétences de communication sociale comme l'humour (Devillers *et al.*, 2015). Ce projet s'intéresse à des interactions divertissantes intervenant dans un environnement social comme une cafétéria.

Cette démonstration présente un système de dialogue permettant un échange entre deux interlocuteurs : un participant humain et le système conversationnel. Ce système de dialogue présente deux caractéristiques principales. La première est d'être « social », c'est-à-dire non contraint par une tâche clairement établie entre les interlocuteurs (e.g., réserver un billet de train). La seconde est d'être en domaine ouvert, c'est-à-dire non restreint à un domaine particulier (e.g., le domaine ferroviaire). Il appartient donc à la famille des systèmes de type « chatbot ».

Ce système utilise un module de sélection de réponse « système » qui a été automatiquement créé depuis un corpus de sous-titres de séries télévisées, décrit et évalué dans (Dubuisson Duplessis *et al.*, 2016). Il sélectionne des contributions surprenantes et amusantes en réponse à un énoncé humain afin de maintenir sa participation dans le dialogue. L'interaction se déroule par le biais d'une interface web de type « salon de conversation en ligne ». Le dialogue est mené par le participant humain. Il prend

la forme d'une succession de paires de type énoncé de l'humain/réponse du système. Le système fonctionne en langue anglaise ; une version en langue française est actuellement en développement.

2 Description technique du système

Le modèle de dialogue sous-jacent au système appartient à la famille des modèles de dialogue fondé sur des exemples (Lee *et al.*, 2009). Cette dernière se fonde sur l'exploitation d'une base d'exemples de dialogue indexés qui représentent un ensemble de stratégie de dialogue. Différentes approches, dans ce cadre, ont été proposées dans la littérature, s'appuyant sur des scénarios de films (Banchs & Li, 2012; Nio *et al.*, 2014) ou encore des sous-titres (Ameixa *et al.*, 2014). Une particularité de notre approche est une automatisation complète des deux processus constituant ce type de modèle de dialogue : (i) la création de la base d'exemples de dialogue, et (ii) la sélection de la contribution du système en réponse à un énoncé humain et depuis la base d'exemples.

La base d'exemples de dialogue est automatiquement construite depuis un corpus de sous-titres de séries télévisées comprenant différents genres tels que « comédie » et « science-fiction » (e.g., « The Big Bang Theory », « Malcolm »). Ce corpus subit un ensemble de pré-traitements à partir d'outils du traitement automatique des langues (e.g., normalisation lexicale, lemmatisation, reconnaissance d'entités nommées). La base contient approximativement 270000 exemples de dialogue extraits depuis 432 fichiers de sous-titres.

La gestion du dialogue consiste en la sélection d'une réponse appropriée depuis la base d'exemples et en son adaptation compte-tenu de l'énoncé du participant humain. Notre approche discerne trois grandes étapes : (1) la sélection d'un ensemble de réponses candidates depuis la base d'exemple, (2) la sélection de la réponse la plus appropriée, et (3) la transformation de la réponse sélectionnée en prenant en compte l'énoncé humain (e.g., substitution d'entités nommées).

Des développements futurs prévoient d'étendre cette approche au corpus OpenSubtitles 2016 (Lison & Tiedemann, 2016) pour extraire jusqu'à 50 millions d'exemples de dialogue. Pour exploiter cette large base d'exemples, la sélection des réponses candidates se fera par calculs de similarité dans un plongement vectoriel du type *doc2vec* (Le & Mikolov, 2014). De premières expériences sont d'ores et déjà en cours et le système résultant sera également montré à titre de comparaison.

3 Informations complémentaires

Ce système a été utilisé afin de collecter un corpus de 41 dialogues avec 27 participants, disponible librement à l'URL : <https://ucar.limsi.fr>. Le système est en évolution active. La version courante est disponible en ligne : <https://webjoker.limsi.fr/>¹.

1. L'accès est limité. Pour y accéder, utiliser le nom d'utilisateur `batman` et le mot de passe `You:Joker`).

Références

- AMEIXA D., COHEUR L., FIALHO P. & QUARESMA P. (2014). Luke, I am your father : dealing with out-of-domain requests by using movies subtitles. In *Intelligent Virtual Agents*, p. 13–21 : Springer.
- BANCHS R. E. & LI H. (2012). IRIS : a chat-oriented dialogue system based on the vector space model. In *Proceedings of the ACL 2012 System Demonstrations*, p. 37–42 : Association for Computational Linguistics.
- DEVILLERS L., ROSSET S., DUBUISSON DUPLESSIS G., SEHILI M., BÉCHADE L., DELABORDE A., GOSSART C., LETARD V., YANG F., YEMEZ Y., TÜRKER B., SEZGIN M., EL HADDAD K., DUPONT S., LUZZATI D., ESTÈVE Y., GILMARTIN E. & CAMPBELL N. (2015). Multimodal data collection of human-robot humorous interactions in the joker project. In *6th International Conference on Affective Computing and Intelligent Interaction (ACII)*.
- DUBUISSON DUPLESSIS G., LETARD V., LIGOZAT A.-L. & ROSSET S. (2016). Purely Corpus-based Automatic Conversation Authoring. In *10th edition of the Language Resources and Evaluation Conference (LREC)*, Portorož, Slovenia.
- LE Q. V. & MIKOLOV T. (2014). Distributed representations of sentences and documents. *arXiv preprint arXiv :1405.4053*.
- LEE C., JUNG S., KIM S. & LEE G. G. (2009). Example-based dialog modeling for practical multi-domain dialog system. *Speech Communication*, **51**(5), 466–484.
- LISON P. & TIEDEMANN J. (2016). Opensubtitles2016 : Extracting large parallel corpora from movie and tv subtitles. In *10th edition of the Language Resources and Evaluation Conference (LREC)*, Portorož, Slovenia.
- NIO L., SAKTI S., NEUBIG G., TODA T., ADRIANI M. & NAKAMURA S. (2014). Developing non-goal dialog system based on examples of drama television. In *Natural Interaction with Robots, Knowbots and Smartphones*, p. 355–361. Springer.

Tag Thunder : plateforme de démonstration et d'expérimentation

Jean-Marc Lecarpentier Elena Manishina Maxence Busson

Fabrice Maurel Stéphane Ferrari

Université de Normandie, UNICAEN, GREYC UMR CNRS 6072

Campus Côte de Nacre, Bd du Maréchal Juin, 14032 CAEN cedex 5, France

{firstname.lastname}@unicaen.fr

RÉSUMÉ

Dans cette démonstration, nous proposons un système qui permettrait aux utilisateurs non-voyants d'obtenir le *first glance* d'une page web. L'objectif est de réduire le temps d'accès à la structure logico-thématique de la page et de favoriser le développement de stratégies de lecture de haut niveau. Notre concept, appelé Tag Thunder, s'appuie sur une phase de segmentation de la page en zones, suivie d'une étape de représentation des zones par un mot ou groupe de mots, puis une vocalisation simultanée de ces représentants.

ABSTRACT

Tag Thunder : demonstration and experimentation platform

In this demonstration we present an architecture to allow visually impaired people to perceive the layout of a web page. We aim to speed up access to the page structure and the topics of its sections, as well as develop high level reading strategies. Our concept, named Tag Thunder, is based on 3 steps : first, the segmentation of a web page into areas, then keyword extraction for each area and finally a vocalization of the keywords.

MOTS-CLÉS : Synthèse de la Parole, Segmentation, Extraction des Mots Clés.

KEYWORDS: Speech Synthesis, Content Segmentation, Keyword Extraction.

1 Introduction et objectifs

La navigation visuelle sur internet permet d'accéder à une vue précoce et globale des pages Web ('skimming'). (Dias & Conde, 2007) montrent que dans ce cas la typographie et la disposition des éléments dans le document revêtent une importance capitale. Les usagers non-voyants n'ont que peu l'usage de cette possibilité (Ahmed *et al.*, 2012), bien qu'il existe des stratégies palliatives (Borodin *et al.*, 2010) telles que l'accélération du débit de la synthèse de la parole, le saut de titres en titres ou de liens hypertextes en liens hypertextes, la lecture des premières ou dernières phrases de tous les paragraphes, etc. Néanmoins, la différence d'efficacité, en comparaison avec la navigation visuelle, reste significative (Bigham *et al.*, 2007).

Dans la présente démonstration, nous proposons une approche qui permettra aux utilisateurs non-voyant d'obtenir une vue globale (*first glance*) d'une page web donnée. Notre représentation combine un filtrage du contenu informationnel d'une page avec la prise en compte de ses propriétés visuelles.