

Systemes de question/réponse : Évaluation des requêtes en amont et en aval du processus d'appariement

Karine LAVENUS, Jens GRIVOLLA
Laboratoire d'Informatique d'Avignon

1 Introduction

Nous présentons les travaux que nous effectuons actuellement au LIA. Nos travaux portent tous deux sur l'évaluation des systèmes de question/réponse, mais à des niveaux différents. Karine Lavenus se penche sur la phase de transformation qui a lieu lors du passage de la question posée par l'utilisateur à la requête traitée par le système. De son côté, Jens Grivolla s'intéresse à l'évaluation des réponses candidates proposées par un ou plusieurs systèmes.

2 Catégorisation et appariement question réponse

En observant un corpus de questions et de réponses valides tirées de la campagne d'évaluation de TREC-9, on devrait pouvoir définir des critères de pertinence autres que la simple présence de mots-clefs à proximité de la réponse, et ce afin d'améliorer l'appariement question/réponse.

2.1 Catégorisation des questions

Tout d'abord, dresser une catégorisation des questions axée sur la typologie sémantique des réponses et rattachée à des patrons lexico-syntaxiques devrait permettre une meilleure désambiguïsation des questions. Il est possible aussi, dans certains cas, d'ajouter à la catégorisation des traits déduits de l'information contenue dans la question afin de préciser le nombre attendu de réponses ou encore le type de lien sémantique liant le terme cible de la question à la réponse.

2.2 Mise en correspondance des patrons lexico-syntaxiques des questions avec ceux des réponses

Ensuite, à travers l'étude d'éléments récurrents, chaque catégorie peut être rattachée à des types de questions représentés par des patrons lexico-syntaxiques eux-même rattachés à des types de réponses représentés par des patrons lexico-syntaxiques. Cette opération a pour avantage de mettre en exergue des traits discrets permettant d'associer à un taux de confiance relativement élevé une catégorie de question à un type de réponse.

2.3 Indices linguistiques pour l'appariement

Enfin, l'analyse des réponses valides fournies par les concurrents de TREC-9 dans des bribes de 250 caractères permet de dégager des indices linguistiques utiles au sein de l'appariement question réponse. Ces indices prennent en compte les dérivés sémantiques et morphologiques des mots-clefs issus de la questions, ainsi que les relations syntaxiques qui sont en jeu, impliquant la notion de distance. Nous insisterons plus particulièrement sur cette partie lors de notre exposé.

Les résultats de cette étude de corpus devrait permettre aux concepteurs de systèmes de question/réponse d'améliorer les performances de leur système, notamment en proposant une meilleure délimitation de la brise correspondant à la réponse et en augmentant la précision du système.

3 Estimation automatique de pertinence

Dans la recherche documentaire et en particulier dans des applications de type “questions/réponses” il est souvent intéressant d’avoir une bonne mesure de confiance dans les réponses trouvées par le moteur de recherche.

Une bonne estimation de pertinence peut permettre de faire un choix entre plusieurs réponses (venant éventuellement de différents systèmes), d’appliquer des méthodes d’enrichissement additionnelles selon les besoins, ou encore de permettre à l’utilisateur de prendre des décisions (comme d’approfondir la recherche à travers un dialogue).

3.1 Approche générale

L’approche poursuivie consiste à déterminer un ensemble d’attributs calculables sur la base de la requête même et (selon les informations disponibles) des réponses ainsi que du processus ayant conduit à la réponse (scores internes du système de recherche documentaire, documents dont la réponse est extraite, etc.)

Chaque requête ou chaque couple requête-réponse sera ainsi représenté par un vecteur d’attributs d’une dimension variant largement selon l’ensemble d’attributs choisi. En utilisant un corpus de requêtes pour lesquelles la pertinence ou non-pertinence des réponses est connue, on peut ensuite appliquer des méthodes de classification automatique afin d’aboutir à des prédictions par rapport au critère de classification (par ex. la probabilité de pertinence de la réponse) pour des nouvelles requêtes non comprises dans ce corpus d’apprentissage.

Alors que les mesures de probabilité de pertinence utilisées par le système de recherche même sont généralement optimisées purement pour le classement comparatif des différentes réponses candidates, une estimation extérieure au système peut apporter des informations additionnelles et permettre la comparaison des résultats livrés par différents systèmes.

3.2 Expérimentations

Des expérimentations ont été conduites surtout sur la base de TREC, en particulier la tâche de recherche “*ad hoc*”, c.à.d. la recherche documentaire classique répondant à une requête par une liste de documents estimés pertinents par le système. Quelques essais ont également été faits dans le domaine du “*question answering*”, mais n’ont pas encore pu être approfondis.

Une liste d’attributs pouvant être corrélés avec la difficulté de la requête (en terme de précision obtenue par le moteur de recherche) a été développée et évolue constamment. Une fois la représentation vectorielle de la requête calculée, différentes méthodes de classification et décision automatique (arbres de décision, SVM, etc.) ont été appliquées pour obtenir des prédictions. Celles-ci ont ensuite été évaluées par validation croisée utilisant la méthode “*leave-one-out*”.

Actuellement sur Adhoc les prédictions obtenues sont généralement meilleures que des prédictions basées purement sur la distribution connues des classes, mais dépendent trop fortement du corpus traité. En particulier, un système développé sur TREC 6 ne donnera pas de résultats utilisables sur TREC 8.

Sur la tâche QA, la qualité des prédictions ne se détache pas clairement du résultat obtenu sur la distribution seule, ce qui montre que l’ensemble d’attributs ne semble pas approprié dans ce contexte. Ceci n’est cependant pas surprenant tenant compte du fait que les quelques essais faits pour l’instant n’utilisent comme base que les requêtes mêmes sans avoir accès aux réponses fournies et les documents correspondants.

3.3 Perspectives

Pour pouvoir espérer obtenir des classifications plus pertinentes, il sera indispensable de développer une intégration plus serrée avec le moteur question/réponse utilisé afin de pouvoir définir des attributs additionnels tenant compte non seulement de la requête mais également des réponses.