Feature ID	Description
D_TFIDF_N (ngram $N = \{1, 2, 3\}$)	Cosine, Manhattan, Euclidean and Jaccard between q^* and the thread t_i.
D_WEMB	Cosine, Manhattan, Euclidean distances between q^* and the thread using word2vec.
SOCIAL	
SOCIAL_N_REPLIES	Number of replies.
SOCIAL_NDIF_REPLIERS	Number of different repliers.
SOCIAL_RATE_FAVORITES	Average of favorites.
SOCIAL_RATE_RETWEETS	Average of retweets.
SOCIAL_MENTIONS	Number of mentions.
SOCIAL_NDIF_MENTIONS	Number of different @mentions.
SOCIAL_NDIF_HASHTAGS	Number of different #hashtags.
USERS	
USERS_NDIF_FOLLOWERS	Number of followers of different users of the thread.
USERS_NDIF_FOLLOWINGS	Number of followings of different users of the thread.
USERS_AVE_AGE	Age average between replies and user date of Twitter.
USERS_RATE_VERIFIED_ACCOUNT	Average of users with verified account.
CONTENT	
CONTENT_NDIF_URLS	Number of different URLS.
CONTENT_N_WORDS	Number of words.
CONTENT_DENSITY	Number of words / number of tweets.
CONTENT_RATE_UPPER	Average of uppercase.
CONTENT_RATE_LOWER	Average of lowercase.
CONTENT_EMOTICONS_POS	Number of positive emoticons.
CONTENT_EMOTICONS_NEG	Number of negative emoticons.
CONTENT_EMOTICONS_NEU	Number of neutral emoticons.
CONTENT_RATE_VOCAB	Rate of well written words.
TIME	
TIME_LIFESPAN	Time difference between the first tweet and the last.
TIME_AVERAGE	time average between each tweet of the thread.
POS	A set of parts of speech tags.
REPW	Rate of the words that are in the 50% more representative words.
WEMB_Q	Explicit vector representation of the question q^* using word2vec (300 dim.).
WEMB_THR	Explicit vector representation of the thread using word2vec (300 dim.).

Table 1
Feature sets used for the ranking task. In some cases, they are a set of features and others single. For word embeddings (WEMB), we use a pre-trained model of 400 million tweets by (Godin et al., 2015).