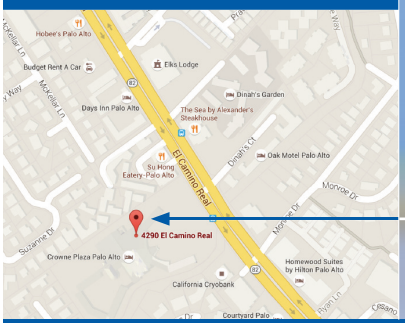
NLP FACULTY SUMMIT 2015

FACEBOOK HEADQUARTERS

1 HACKER WAY, MENLO PARK, CA 94025

OCTOBER 9, 2015

FROM:	TO:		SPEAKER:	TEAM:	ABSTRACT:
900	930	BREAKFAST AND REGISTRATION			
930	1000	INTRODUCTION AND KEYNOTE	Xiao Li		
1000	1100	ARABIC DIALECT CLASSIFICATION FOR TRANSLATION	Fei Huang	Translation	Arabic dialect classification has been an important and challenging problem for Arabic language processing, especially for social media text analysis and machine translation. We propose an approach to improving Arabic dialect classification with semi-supervised learning: multiple classifiers are trained with weakly supervised, strongly supervised, and unsupervised data. Their combination yields significant and consistent improvement on two different test sets. The dialect classification accuracy is improved by 5% over the strongly supervised classifier and 20\% over the weakly supervised classifier. Furthermore, when applying the improved dialect classifier to build a Modern Standard Arabic (MSA) language model (LM), the new model size is reduced by 70% while the English-Arabic translation quality is improved by 0.6 BLEU point.
		LARGE VOCABULARY NEURAL LANGUAGE MODELS	David Grangier	FAIR	Continuous space language models are helpful to various NLP applications, such as speech recognition and machine translation. Their training over large vocabularies is still expensive compared to classical count-based models. This presentation compares various training strategies such as softmax, hiearchical softmax and sub vocabulary sampling over different datasets and analyze their tradeoff in terms of perplexity versus training speed.
1100	1115	MINI BREAK			
1115	1215	POST CLASSIFICATION	Jeff Pasternack	Feed Ranking	With over a billion posts per day, people create and consume an incredible amount of content on Facebook. Knowing what these posts are about helps inform feed ranking, search, "related" posts, feed curation, trending, and numerous other applications. Our methods of classifying posts are diverse, ranging from industry stalwarts like LDA and logistic regression to the more recently popular embedding models; distilling the collective wisdom of these models is then done with an ensemble metaclassifier. Inference at Facebook scale can be challenging, but having virtually unlimited data with diverse supervision represents a unique opportunity for content understanding and the societal impact of applying the results to a product used by over a billion people.
		SENTIMENT CLASSIFIER	Satya Satyavarta	QuantEng	Sentiment analysis on Facebook posts and comments can be challenging because of the linguistic diversity on the site, in terms of both the variety of writing styles, as well as the number of languages. We frame the problem as a supervised learning task by deriving proxy labels from user interactions with the site. This enables building high quality models with high volumes of training data, for multiple languages, trained continuously as fresh data becomes available. This talk details the training techniques used in our sentiment classification (including proxy label design, multi-class classification, arbitration, and calibration), and the constraints that inform the choices made therein.
1215	1315	LUNCH			
1315	1545	QUERY UNDERSTANDING	Ves Stoyanov	Search NLP	Producing quality search results requires understanding the semantic of the query. In our query understanding efforts, we tackle several tasks such as segmenting the query, predicting its intent, recognizing entities mentioned in the query and their relations and even producing a semantic parse for structured queries. While none of these problems is completely new, the characteristics of Search on Facebook add some original flavors: our searches are highly personalized and our query distribution is tail-heavy. I will describe some of our recent work on query understanding that improves accuracy by performing joint segmentation and intent prediction.
		UNDERSTANDING MEDIA CONTENT FOR INSTAGRAM	Maxime Boucher	Instagram	Instagram recently launched a new version of its search and explore product, which helps users find interesting content on Instagram. As part of this launch, Instagram added new functionality to find interesting content via trending tags. There are multiple challenges involved in detecting trends in real time ranging from detecting trends that are popular within the community, filtering out spurious or generally non-trending hashtags and selecting a diverse set of trends to show to the users. During this talk, I will describe the architecture behind realtime detection of trending tags, along with how trend selection is performed before being shown to users.
		FINDING HIGHLIGHTS IN FREEFORM TEXT	Christy Sauper	Data Science	Facebook is a great melting pot where people can express a variety of opinions about any given subject; for example, giving details about a visit to a restaurant, sharing their viewpoint about a news article, or submitting a survey about their experience using Facebook. In each of these cases, we want to understand and highlight key points mentioned by multiple people. This talk will give an overview of where these highlights are useful both internally and externally and the methods that we use to explore topic highlights from freeform text.
1545	1600	AFTERNOON BREAK			
1600	1700	PANEL DISCUSSION			
1700	1900	HAPPY HOUR			



CROWNE PLAZA PALO ALTO

4290 EL CAMINO REAL
PALO ALTO, CA 94306

FACEBOOK HQ

1 HACKER WAY,
MENLO PARK, CA 94025

NOTE: *A van has been arranged to transport guests from the hotel to the Facebook Campus. Please meet the van outside the lobby at 8:30am.*