# Weakly Supervised Models of Aspect-Sentiment for Online Course Discussion Forums

[1]**Arti Ramesh,** [2]**Shachi H. Kumar,** [2]**James Foulds,** [2]**Lise Getoor**
[1]University of Maryland, College Park   [2]University of California, Santa Cruz
artir@cs.umd.edu, {shulluma, jfoulds, getoor}@ucsc.edu

## Abstract

Massive open online courses (MOOCs) are redefining the education system and transcending boundaries posed by traditional courses. With the increase in popularity of online courses, there is a corresponding increase in the need to understand and interpret the communications of the course participants. Identifying topics or *aspects* of conversation and inferring sentiment in online course forum posts can enable instructor interventions to meet the needs of the students, rapidly address course-related issues, and increase student retention. Labeled aspect-sentiment data for MOOCs are expensive to obtain and may not be transferable between courses, suggesting the need for approaches that do not require labeled data. We develop a weakly supervised joint model for aspect-sentiment in online courses, modeling the dependencies between various aspects and sentiment using a recently developed scalable class of statistical relational models called hinge-loss Markov random fields. We validate our models on posts sampled from twelve online courses, each containing an average of around 10,000 posts, and demonstrate that jointly modeling aspect together with sentiment improves the prediction accuracy for both aspect and sentiment.

## 1 Introduction

Massive Open Online Courses (MOOCs) have emerged as a powerful medium for imparting education to a wide geographical population. Discussion forums are the primary means of communication between MOOC participants (students, TAs, and instructors). Due to the open nature of these courses, they attract people from all over the world leading to large numbers of participants and hence, large numbers of posts in the discussion forums. In the courses we worked with, we found that over the course of the class there were typically over 10,000 posts.

Within this slew of posts, there are valuable *problem-reporting* posts that identify issues such as broken links, audio-visual glitches, and inaccuracies in the course materials. Automatically identifying these reported problems is important for several reasons: *i*) it is time-consuming for instructors to manually screen through all of the posts due to the highly skewed instructor-to-student ratio in MOOCs, *ii*) promptly addressing issues could help improve student retention, and *iii*) future iterations of the course could benefit from identifying technical and logistical issues currently faced by students. In this paper, we investigate the problem of determining the fine-grained topics of posts (which we refer to as "MOOC aspects") and the sentiment toward them, which can potentially be used to improve the course.

While aspect-sentiment has been widely studied, the MOOC discussion forum scenario presents a unique set of challenges. Labeled data are expensive to obtain, and posts containing fine-grained aspects occur infrequently in courses and differ across courses, thereby making it expensive to get sufficient coverage of all labels. Few distinct aspects occur per course, and only 5-10% of posts in a course are relevant. Hence, getting labels for fine-grained labels involves mining and annotating posts from a large number of courses. Further, cre-

ating and sharing labeled data is difficult as data from online courses is governed by IRB regulations. Privacy restrictions are another reason why unsupervised/weakly-supervised methods can be helpful. Lastly, to design a system capable of identifying all possible MOOC aspects across courses, we need to develop a system that is not fine-tuned to any particular course, but can adapt seamlessly across courses.

To this end, we develop a weakly supervised system for detecting aspect and sentiment in MOOC forum posts and validate its effectiveness on posts sampled from twelve MOOC courses. Our system can be applied to any MOOC discussion forum with no or minimal modifications.

Our contributions in this paper are as follows:

- We show how to encode weak supervision in the form of seed words to extract extract course-specific features in MOOCs using SeededLDA, a seeded variation of topic modeling (Jagarlamudi et al., 2012).

- Building upon our SeededLDA approach, we develop a joint model for aspects and sentiment using the *hinge-loss Markov random field (HL-MRF)* probabilistic modeling framework. This framework is especially well-suited for this problem because of its ability to combine information from multiple features and jointly reason about aspect and sentiment.

- To validate the effectiveness of our system, we construct a labeled evaluation dataset by sampling posts from twelve MOOC courses, and annotating these posts with fine-grained MOOC aspects and sentiment via *crowdsourcing*. The annotation captures fine-grained aspects of the course such as content, grading, deadlines, audio and video of lectures and sentiment (i.e., *positive, negative, and neutral*) toward the aspect in the post.

- We demonstrate that the proposed HL-MRF model can predict fine-grained aspects and sentiment and outperforms the model based only on SeededLDA.

## 2   Related Work

To the best of our knowledge, the problem of predicting aspect and sentiment in MOOC forums has not yet been addressed in the literature. We review prior work in related areas here.

**Aspect-Sentiment in Online Reviews**   It is valuable to identify the sentiment of online reviews towards aspects such as hotel cleanliness and cellphone screen brightness, and sentiment analysis at the aspect-level has been studied extensively in this context (Liu and Zhang, 2012). Several of these methods use latent Dirichlet allocation topic models (Blei et al., 2003) and variants of it for detecting aspect and sentiment (Lu et al., 2011; Lin and He, 2009). Liu and Zhang (2012) provide a comprehensive survey of techniques for aspect and sentiment analysis. Here, we discuss works that are closely related to ours.

Titov and McDonald (2008) emphasize the importance of an unsupervised approach for aspect detection. However, the authors also indicate that standard LDA (Blei et al., 2003) methods capture global topics and not necessarily pertinent aspects — a challenge that we address in this work. Brody and Elhadad (2010), Titov and McDonald (2008), and Jo and Oh (2011) apply variations of LDA at the sentence level for online reviews. We find that around 90% of MOOC posts have only one aspect, which makes sentence-level aspect modeling inappropriate for our domain.

Most previous approaches for sentiment rely on manually constructed lexicons of strongly positive and negative words (Fahrni and Klenner, 2008; Brody and Elhadad, 2010). These methods are effective in an online review context, however sentiment in MOOC forum posts is often implicit, and not necessarily indicated by standard lexicons. For example, the post "Where is my certificate? Waiting for it for over a month." expresses negative sentiment toward the certificate aspect, but does not include any typical negative sentiment words. In our work, we use a data-driven model-based approach to discover domain-specific lexicon information guided by small sets of seed words.

There has also been substantial work on joint models for aspect and sentiment (Kim et al., 2013; Diao et al., 2014; Zhao et al., 2010; Lin et al., 2012), and we adopt such an approach in this paper. Kim et al. (2013) use a hierarchical aspect-sentiment model and evaluate it for online reviews. Mukherjee and Liu (2012) use seed words for discovering aspect-based sentiment topics. Drawing on the ideas of Mukherjee and Liu (2012) and Kim et al. (2013), we propose a statistical rela-

| Post 1: I have not received the **midterm**. |
| Post 2: No **lecture subtitles** week, will they be uploaded? |
| Post 3: I am ... and I am looking forward to learn more ... |

Table 1: Example posts from MOOC forums. Aspect words are highlighted in **bold**.

tional learning approach that combines the advantages of seed words, aspect hierarchy, and flat aspect-sentiment relationships. It is important to note that a broad majority of the previous work on aspect sentiment focuses on the specific challenges of online review data. As discussed in detail above, MOOC forum data have substantially different properties, and our approach is the first to be designed particularly for this domain.

**Learning Analytics**  In another line of research, there is growing body of work on the analysis of online courses. Regarding MOOC forum data, Stump et al. (2013) propose a framework for taxonomically categorizing forum posts, leveraging manual annotations. We differ from their approach in that we develop an automatic system to predict MOOC forum categories without using labeled training data. Ramesh et al. (2014b) categorize forum posts into three broad categories in order to predict student engagement. Unlike this method, our system is capable of fine-grained categorization and of identifying aspects in MOOCS. Chaturvedi et al. (2014) focus on predicting instructor intervention using lexicon features and thread features. In contrast, our system is capable of predicting fine MOOC aspects and sentiment of discussion forum posts and thus provides a more informed analysis of MOOC posts.

## 3 Problem Setting and Data

MOOC participants primarily communicate through discussion forums, consisting of posts, which are short pieces of text. Table 1 provides examples of posts in MOOC forums. Posts *1* and *2* report issues and feedback for the course, while post *3* is a social interaction message. Our goal is to distinguish *problem-reporting* posts such as *1* and *2* from *social* posts such as *3*, and to *identify the issues* that are being discussed.

We formalize this task as an *aspect-sentiment* prediction problem (Liu and Zhang, 2012). The issues reported in MOOC forums can be related to the different elements of the course such as *lectures* and *quizzes*, which are referred to as *aspects*. The aspects were selected based on MOOC domain expertise and inspiration from Stump et al.

| COARSE-ASPECT | FINE-ASPECT | Description | # of posts |
|---|---|---|---|
| LECTURE | LECTURE-CONTENT | Content of lectures. | 559 |
| | LECTURE-VIDEO | Video of lectures. | 215 |
| | LECTURE-SUBTITLES | Subtitles of lecture. | 149 |
| | LECTURE-AUDIO | Audio of lecture. | 136 |
| | LECTURE-LECTURER | Delivery of instructor. | 69 |
| QUIZ | QUIZ-CONTENT | Content in quizzes. | 439 |
| | QUIZ-GRADING | Grading of quizzes. | 360 |
| | QUIZ-SUBMISSION | Quiz submission. | 329 |
| | QUIZ-DEADLINE | Deadline of quizzes. | 142 |
| CERTIFICATE | | Course certificates. | 194 |
| SOCIAL | | Social interaction posts. | 1187 |

Table 2: Descriptions of *coarse* and *fine* aspects.

(2013), aiming to cover common concerns that could benefit from intervention. The task is to predict these aspects for each post, along with the *sentiment* polarity toward the aspect, which we code as *positive*, *negative*, or *neutral*. The negative-sentiment posts, along with their aspects, allow us to identify potentially correctable issues in the course. As labels are expensive in this scenario, we formulate the task as *weakly supervised* prediction problem. In our work, we assume that a post has at most one fine-grained aspect, as we found that this was true for 90% of the posts in our data. This property is due in part to the brevity of forum posts, which are much shorter documents than those considered in other aspect-sentiment scenarios such as product reviews.

### 3.1 Aspect Hierarchy

While we do not require labeled data, our approaches allow the analyst to instead relatively easily encode a small amount of domain knowledge by seeding the models with a few words relating to each aspect of interest. Hence, we refer to our approach as *weakly supervised*. Our models can further make use of hierarchical structure between the aspects. The proposed approach is flexible, allowing the aspect seeds and hierarchy to be selected for a given MOOC domain.

For the purposes of this study, we represent the MOOC aspects with a two-level hierarchy. We identify a list of *nine* fine-grained aspects, which are grouped into *four* coarse topics. The *coarse* aspects consist of LECTURE, QUIZ, CERTIFICATE, and SOCIAL topics. Table 2 provides a description of each of the aspects and also gives the number of posts in each aspect category after annotation.

As both LECTURE and QUIZ are key coarse-level aspects in online courses, and more nuanced aspect information for these is impor-

tant to facilitate instructor interventions, we identify fine-grained aspects for these topics. For LECTURE we identify LECTURE-CONTENT, LECTURE-VIDEO, LECTURE-AUDIO, LECTURE-SUBTITLES, and LECTURE-LECTURER as fine aspects. For QUIZ, we identify the fine aspects QUIZ-CONTENT, QUIZ-GRADING, QUIZ-DEADLINES, and QUIZ-SUBMISSION. We use the label SOCIAL to refer to social interaction posts that do not mention a problem-related aspect.

## 3.2 Dataset

We construct a dataset by sampling posts from MOOC courses to capture the variety of aspects discussed in online courses. We include courses from different disciplines (business, technology, history, and the sciences) to ensure broad coverage of aspects. Although we adopt an approach that does not require labeled data for training, which is important for most practical MOOC scenarios, in order to validate our methods we obtain labels for the sampled posts using *Crowdflower*,[1] an online crowd-sourcing annotation platform. Each post was annotated by at least 3 annotators. Crowdflower calculates *confidence* in labels by computing trust scores for annotators using test questions. Kolhatkar et al. (2013) provide a detailed analysis of Crowdflower trust calculations and the relationship to inter-annotator agreement. We follow their recommendations and retain only labels with *confidence* $> 0.5$.

## 4 Aspect-Sentiment Prediction Models

In this section, we develop models and feature-extraction techniques to address the challenges of aspect-sentiment prediction for MOOC forums. We present two weakly-supervised methods—first, using a seeded topic modeling approach (Jagarlamudi et al., 2012) to identify aspects and sentiment. Second, building upon this method, we then introduce a more powerful statistical relational model which reasons over the seeded LDA predictions as well as sentiment side-information to encode hierarchy information and correlations between sentiment and aspect.

## 4.1 Seeded LDA Model

Topic models (Blei et al., 2003), which identify latent semantic themes from text corpora, have

previously been successfully used to discover aspects for sentiment analysis (Diao et al., 2014). By equating the topics, i.e. discrete distributions over words, with aspects and/or sentiment polarities, topic models can recover aspect-sentiment predictions. In the MOOC context we are specifically interested in problems with the courses, rather than general topics which may be identified by a topic model, such as the topics of the course material. To guide the topic model to identify aspects of interest, we use *SeededLDA* (Jagarlamudi et al., 2012), a variant of LDA which allows an analyst to "seed" topics by providing key words that should belong to the topics.

We construct SeededLDA models by providing a set of seed words for each of the coarse and fine aspects in the aspect hierarchy of Table 2. We also seed topics for *positive*, *negative* and *neutral* sentiment polarities. The seed words for coarse topics are provided in Table 3, and fine aspects in Table 4. For the sentiment topics (Table 5), the seed words for the topic *positive* are positive words often found in online courses such as *thank*, *congratulations*, *learn*, and *interest*. Similarly, the seed words for the *negative* topic are negative in the context of online courses, such as *difficult*, *error*, *issue*, *problem*, and *misunderstand*.

Additionally, we also use SeededLDA for isolating some common problems in online courses that are associated with sentiment, such as *difficulty*, *availability*, *correctness*, and course-specific seed words from the syllabus as described in Table 6. Finally, having inferred the SeededLDA model from the data set, for each post $p$ we predict the most likely aspect and the most likely sentiment polarity according to the post's inferred distribution over topics $\theta^{(p)}$.

In our experiments, we tokenize and stem the posts using NLTK toolkit (Loper and Bird, 2002), and use a stop word list tuned to online course discussion forums. The topic model Dirichlet hyper-parameters are set to $\alpha = 0.01$, $\beta = 0.01$ in our experiments. For SeededLDA models corresponding to the seed sets in Tables 3, 4, and 5, the number of topics is equal to the number of seeded topics. For SeededLDA models corresponding to the seed words in Tables 6 and 3, we use 10 topics, allowing for some *unseeded* topics that are not captured by the seed words.

---

[1] http://www.crowdflower.com/

LECTURE: lectur, video, download, volum, low, headphon, sound, audio, transcript, subtitl, slide, note
QUIZ: quiz, assignment, question, midterm,exam, submiss, answer, grade, score, grad, midterm, due, deadlin
CERTIFICATE: certif, score, signatur, statement, final, course, pass, receiv, coursera, accomplish, fail
SOCIAL: name, course, introduction, stud, group, everyon, student

Table 3: Seed words for *coarse* aspects

LECTURE-VIDEO: video, problem, download, play, player, watch, speed, length, long, fast, slow, render, qualiti
LECTURE-AUDIO: volum, low, headphon, sound, audio, hear, maximum, troubl, qualiti, high, loud, heard
LECTURE-LECTURER: professor, fast, speak, pace, follow, speed, slow, accent, absorb, quick, slowli
LECTURE-SUBTITLES: transcript, subtitl, slide, note, lectur, difficult, pdf
LECTURE-CONTENT: typo, error, mistak, wrong, right, incorrect, mistaken
QUIZ-CONTENT: question, challeng, difficult, understand, typo, error, mistak, quiz, assignment
QUIZ-SUBMISSION: submiss, submit, quiz, error, unabl, resubmit
QUIZ-GRADING: answer, question, answer, grade, assignment, quiz, respons ,mark, wrong, score
QUIZ-DEADLINE: due, deadlin, miss, extend, late

Table 4: Seed words for *fine* aspects

POSITIVE: interest, excit, thank, great, happi, glad, enjoy, forward, insight, opportun, clear, fantast, fascin, learn, hope, congratul
NEGATIVE: problem, difficult, error, issu, unabl, misunderstand, terribl, bother, hate, bad, wrong, mistak, fear, troubl
NEUTRAL: coursera, class, hello, everyon, greet, nam, meet, group, studi, request, join, introduct, question, thank

Table 5: Seed words for *sentiment*

DIFFICULTY: difficult, understand, ambigu, disappoint, hard, follow, mislead, difficulti, challeng, clear
CONTENT: typo, error, mistak, wrong, right, incorrect, mistaken, score
AVAILABILITY: avail, nowher, find, access, miss, view, download, broken, link, bad, access, deni, miss, permiss
COURSE-1: develop, eclips, sdk, softwar, hardware, accuser, html, platform, environ, lab, ide, java,
COURSE-2: protein, food, gene, vitamin, evolut, sequenc, chromosom, genet, speci, peopl, popul, evolv, mutat, ancestri
COURSE-3: compani, product, industri, strategi, decision, disrupt, technolog, market

Table 6: Seed words for sentiment specific to online courses

## 4.2 Hinge-loss Markov Random Fields

The approach described in the previous section automatically identifies user-seeded aspects and sentiment, but it does not make further use of structure or dependencies between these values, or any additional side-information. To address this, we propose a more powerful approach using hinge-loss Markov random fields (HL-MRFs), a scalable class of continuous, conditional graphical models (Bach et al., 2015). HL-MRFs have achieved state-of-the-art performance in many domains including knowledge graph identification (Pujara et al., 2013), understanding engagements in MOOCs (Ramesh et al., 2014a), biomedicine and multi-relational link prediction (Fakhraei et al., 2014), and modelling social trust (Huang et al., 2013). These models can be specified using *Probabilistic Soft Logic (PSL)* (Bach et al., 2015), a weighted first order logical templating language. An example of a PSL rule is

$$\lambda : P(a) \wedge Q(a, b) \rightarrow R(b),$$

where *P*, *Q*, and *R* are predicates, *a* and *b* are *variables*, and $\lambda$ is the weight associated with the rule. The weight of the rule indicates its importance in

the HL-MRF probabilistic model, which defines a probability density function of the form

$$P(\mathbf{Y}|\mathbf{X}) \propto \exp\Big( -\sum_{r=1}^{M} \lambda_r \phi_r(\mathbf{Y}, \mathbf{X}) \Big)$$

$$\phi_r(\mathbf{Y}, \mathbf{X}) = (\max\{l_r(\mathbf{Y}, \mathbf{X}), 0\})^{\rho_r} , \quad (1)$$

where $\phi_r(\mathbf{Y}, \mathbf{X})$ is a *hinge-loss potential* corresponding to an instantiation of a rule, and is specified by a linear function $l_r$ and optional exponent $\rho_r \in \{1, 2\}$. For example, in our MOOC aspect-sentiment model, if *P* and *F* denote *post P* and *fine aspect F*, then we have predicates SEEDLDA-FINE(P, F) to denote the value corresponding to topic F in SeededLDA, and FINE-ASPECT(P, F) is the target variable denoting the fine aspect of the post *P*. A PSL rule to encode that the SeededLDA topic *F* suggests that aspect *F* is present is

$$\lambda : \text{SEEDLDA-FINE}(P, F) \rightarrow \text{FINE-ASPECT}(P, F).$$

We can generate more complex rules connecting the different features and target variables, e.g.

$$\lambda : \text{SEEDLDA-FINE}(P, F) \wedge \text{SENTIMENT}(P, S)$$
$$\rightarrow \text{FINE-ASPECT}(P, F).$$

This rule encodes a dependency between SENTI-MENT and FINE-ASPECT, namely that the Seed-edLDA topic and a strong sentiment score increase the probability of the fine aspect. The HL-MRF model uses these rules to encode domain knowledge about dependencies among the predicates. The continuous value representation further helps in understanding the confidence of predictions.

## 4.3 Joint Aspect-Sentiment Prediction using Probabilistic Soft Logic (PSL-Joint)

In this section, we describe our joint approach to predicting aspect and sentiment in online discussion forums, leveraging the strong dependence between aspect and sentiment. We present a system designed using HL-MRFs which combines different features, accounting for their respective uncertainty, and encodes the dependencies between aspect and sentiment in the MOOC context.

Table 7 provides some representative rules from our model.[2] The rules can be classified into two broad categories—1) rules that combine multiple features, and 2) rules that encode the dependencies between aspect and sentiment.

### 4.3.1 Combining Features

The first set of rules in Table 7 combine different features extracted from the post. SEEDLDA-FINE, SEEDLDA-COARSE and SEEDLDA-SENTIMENT-COURSE predicates in rules refer to SeededLDA posterior distributions using *coarse*, *fine*, and course-specific *sentiment* seed words respectively. The strength of our model comes from its ability to encode different combinations of features and weight them according to their importance. The first rule in Table 7 combines the SeededLDA features from both SEEDLDA-FINE and SEEDLDA-COARSE to predict the fine aspect. Interpreting the rule, the fine aspect of the post is more likely to be LECTURE-LECTURER if the coarse Seed-edLDA score for the post is LECTURE, *and* the fine SeededLDA score for the post is LECTURE-LECTURER. Similarly, the second rule provides combinations of some of the other features used by the model—two different SeededLDA scores for sentiment, as indicated by seed words in Tables 5 and 6. The third rule states that certain fine aspects occur together with certain values of sentiment more than others. In online courses, posts that discuss grading usually talk about grievances

and issues. The rule captures that QUIZ-GRADING occurs with negative sentiment in most cases.

### 4.3.2 Encoding Dependencies Between Aspect and Sentiment

In addition to combining features, we also encode rules to capture the taxonomic dependence between coarse and fine aspects, and the dependence between aspect and sentiment (Table 7, bottom). Rules 4 and 5 encode pair-wise dependency between FINE-ASPECT and SENTIMENT, and COARSE-ASPECT and FINE-ASPECT respectively. Rule 4 uses the SeededLDA value for QUIZ-DEADLINES to predict both SENTIMENT, and FINE-ASPECT jointly. This together with other rules for predicting SENTIMENT and FINE-ASPECT individually creates a constrained satisfaction problem, forcing aspect and sentiment to agree with each other. Rule 5 is similar to rule 4, capturing the taxonomic relationship between target variables COARSE-ASPECT and FINE-ASPECT.

Thus, by using conjunctions to combine features and appropriately weighting these rules, we account for the uncertainties in the underlying features and make them more robust. The combination of these two different types of weighted rules, referred to below as *PSL-Joint*, is able to reason collectively about aspect and sentiment.

## 5 Empirical Evaluation

In this section, we present the quantitative and qualitative results of our models on the annotated MOOC dataset. Our models do not require labeled data for training; we use the label annotations only for evaluation. Tables 9 – 12 show the results for the SeededLDA and PSL-Joint models. Statistically significant differences, evaluated using a paired t-test with a rejection threshold of *0.01*, are typed in bold.

### 5.1 SeededLDA for Aspect-Sentiment

For SeededLDA, we use the seed words for *coarse*, *fine*, and *sentiment* given in Tables 3 – 5. After training the model, we use the SeededLDA multinomial posterior distribution to predict the target variables. We use the maximum value in the posterior for the distribution over topics for each post to obtain predictions for coarse aspect, fine aspect, and sentiment. We then calculate precision, recall and F1 values comparing with our ground truth labels.

| PSL-JOINT RULES |
|---|

**Rules combining features**
SEEDLDA-FINE(POST, LECTURE-LECTURER) ∧ SEEDLDA-COARSE(POST, LECTURE) → FINE-ASPECT(POST, LECTURE-LECTURER)
SEEDLDA-SENTIMENT-COURSE(POST, NEGATIVE) ∧ SEEDLDA-SENTIMENT(POST, NEGATIVE) → SENTIMENT(POST, NEGATIVE)
SEEDLDA-SENTIMENT-COURSE(POST, NEGATIVE) ∧ SEEDLDA-FINE(POST, QUIZ-GRADING) → FINE-ASPECT(POST, QUIZ-GRADING)
**Encoding dependencies between aspect and sentiment**
SEEDLDA-FINE(POST, QUIZ-DEADLINES) ∧ SENTIMENT(POST, NEGATIVE) → FINE-ASPECT(POST, QUIZ-DEADLINES)
SEEDLDA-FINE(POST, QUIZ-SUBMISSION) ∧ FINE-ASPECT(POST, QUIZ-SUBMISSION) → COARSE-ASPECT(POST, QUIZ)

Table 7: Representative rules from PSL-Joint model

| Model | LECTURE-CONTENT | | | LECTURE-VIDEO | | | LECTURE-AUDIO | | | LECTURE-LECTURER | | | LECTURE-SUBTITLES | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Prec. | Rec. | F1 | Prec. | Rec. | F1 | Prec. | Rec. | F1 | Prec. | Rec. | F1 | Prec. | Rec. | F1 |
| SEEDEDLDA | 0.137 | 0.057 | 0.08 | 0.156 | 0.256 | 0.240 | 0.684 | 0.684 | 0.684 | 0.037 | 0.159 | 0.06 | 0.289 | 0.631 | 0.397 |
| PSL-JOINT | **0.407** | **0.413** | **0.410** | **0.411** | **0.591** | **0.485** | 0.635 | 0.537 | 0.582 | **0.218** | **0.623** | **0.323** | **0.407** | 0.53 | **0.461** |

Table 8: Precision, recall and F1 scores for some fine aspects

| Model | LECTURE-CONTENT | | | LECTURE-VIDEO | | | LECTURE-AUDIO | | | LECTURE-LECTURER | | | LECTURE-SUBTITLES | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Prec. | Rec. | F1 | Prec. | Rec. | F1 | Prec. | Rec. | F1 | Prec. | Rec. | F1 | Prec. | Rec. | F1 |
| SEEDEDLDA | 0.137 | 0.057 | 0.08 | 0.156 | 0.256 | 0.240 | 0.684 | 0.684 | 0.684 | 0.037 | 0.159 | 0.06 | 0.289 | 0.631 | 0.397 |
| PSL-JOINT | **0.407** | **0.413** | **0.410** | **0.411** | **0.591** | **0.485** | 0.635 | 0.537 | 0.582 | **0.218** | **0.623** | **0.323** | **0.407** | 0.53 | **0.461** |

Table 9: Precision, recall and F1 scores for LECTURE fine aspects

| Model | QUIZ-CONTENT | | | QUIZ-SUBMISSION | | | QUIZ-DEADLINES | | | QUIZ-GRADING | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Prec | Rec. | F1 | Prec | Rec. | F1 | Prec. | Rec. | F1 | Prec. | Rec. | F1 |
| SEEDEDLDA | 0.042 | 0.006 | 0.011 | 0.485 | 0.398 | 0.437 | 0.444 | 0.141 | 0.214 | 0.524 | 0.508 | 0.514 |
| PSL-JOINT | **0.324** | **0.405** | **0.36** | **0.521** | 0.347 | 0.416 | **0.667** | **0.563** | **0.611** | **0.572** | **0.531** | **0.550** |

Table 10: Precision, recall and F1 scores for QUIZ fine aspects

| Model | LECTURE | | | QUIZ | | | CERTIFICATE | | | SOCIAL | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Prec | Rec. | F1 | Prec. | Rec. | F1 | Prec. | Rec. | F1 | Prec. | Rec. | F1 |
| SEEDEDLDA | 0.597 | 0.673 | 0.632 | 0.752 | 0.583 | 0.657 | 0.315 | 0.845 | 0.459 | 0.902 | 0.513 | 0.654 |
| PSL-JOINT | 0.563 | **0.715** | 0.630 | 0.724 | **0.688** | **0.706** | **0.552** | 0.711 | **0.621** | 0.871 | **0.530** | **0.659** |

Table 11: Precision, recall and F1 scores for coarse aspects

| Model | POSITIVE | | | NEGATIVE | | | NEUTRAL | | |
|---|---|---|---|---|---|---|---|---|---|
| | Prec | Rec. | F1 | Prec. | Rec. | F1 | Prec. | Rec. | F1 |
| SEEDEDLDA | 0.104 | 0.721 | 0.182 | 0.650 | 0.429 | 0.517 | 0.483 | 0.282 | 0.356 |
| PSL-JOINT | **0.114** | 0.544 | **0.189** | **0.571** | **0.666** | **0.615** | **0.664** | **0.322** | **0.434** |

Table 12: Precision, recall and F1 scores for sentiment

## 5.2 PSL for Joint Aspect-Sentiment (PSL-Joint)

Tables 9 and 10 give the results for the fine aspects under LECTURE and QUIZ. PSL-JOINT performs better than SEEDEDLDA in most cases, without suffering any statistically significant losses. Notable cases include the increase in scores for LECTURE-LECTURER, LECTURE-SUBTITLES, LECTURE-CONTENT, QUIZ-CONTENT, QUIZ-GRADING, and QUIZ-DEADLINES, for which the scores increase by a large margin over SeededLDA. We observe that for LECTURE-CONTENT and QUIZ-CONTENT, the increase in scores is more significant than others with SeededLDA performing very poorly. Since both lecture and quiz content have the same kind of words related to the course material, SeededLDA is not able to distinguish between these two aspects. We found that in 63% of these missed predictions, SeededLDA predicts LECTURE-CONTENT, instead of QUIZ-CONTENT, and vice versa. In contrast, PSL-Joint uses both coarse and fine SeededLDA scores and captures the dependency between a coarse aspect and its corresponding fine aspect. Therefore, PSL-Joint is able to distinguish between LECTURE-CONTENT and QUIZ-CONTENT. In the

next section, we present some examples of posts that SEEDEDLDA misclassified but were predicted correctly by PSL-Joint.

Table 11 presents results for the *coarse-aspects*. We observe that PSL-Joint performs better than SeededLDA for all classes. In particular for CERTIFICATE and QUIZ, PSL-Joint exhibits a marked increase in scores when compared to SeededLDA. This is also true for sentiment, for which the scores for NEUTRAL and NEGATIVE sentiment show significant improvement (Table 12).

## 5.3 Interpreting PSL-Joint Predictions

Table 13 presents some examples of posts that PSL-Joint predicted correctly, and which SeededLDA misclassified. The first two examples illustrate that PSL can predict the subtle difference between LECTURE-CONTENT and QUIZ-CONTENT. Particularly notable is the third example, which contains mention of both *subtitles* and *audio*, but the negative sentiment is associated with *audio* rather than *subtitles*. PSL-Joint predicts the fine aspect as LECTURE-AUDIO, even though the underlying SeededLDA feature has a high score for LECTURE-SUBTITLES. This example illustrates the strength of the joint reasoning approach in PSL-Joint. Finally, in the last example, the post mentions starting a *group* to discuss videos. This is an ambiguous post containing the keyword *video*, while it is in reality a social post about starting a group. PSL-Joint is able to predict this because it uses both the sentiment scores associated with the post and the SeededLDA scores for fine aspect, and infers that social posts are generally positive. So, combining the feature values for social aspect and positive sentiment, it is able to predict the fine aspect as SOCIAL correctly.

The continuous valued output predictions produced by PSL-Joint allow us to rank the predicted variables by output prediction value. Analyzing the predictions for posts that PSL-Joint misclassified, we observe that for *four* out of *nine* fine aspects, more than 70% of the time the correct label is in the top three predictions. And, for all fine aspects, the correct label is found in the top 3 predictions around 40% of the time. Thus, using the top three predictions made by PSL-Joint, we can understand the fine aspect of the post to a great extent. Table 14 gives some examples of posts for which the second best prediction by PSL-Joint is the correct label. For these examples, we found

that PSL-Joint misses the correct prediction by a small margin($< 0.2$). Since our evaluation scheme only considers the maximum value to determine the scores, these examples were treated as misclassified.

## 5.4 Understanding Instructor Intervention using PSL-Joint Predictions

In our 3275 annotated posts, the instructor replied to 787 posts. Of these, 699 posts contain a mention of some MOOC aspect. PSL-Joint predicts 97.8% from those as having an aspect and 46.9% as the correct aspect. This indicates that PSL-Joint is capable of identifying the most important posts, i.e. those that the instructor replied to, with high accuracy. PSL-Joint's MOOC aspect predictions can potentially be used by the instructor to select a subset of posts to address in order to cover the main reported issues. We found in our data that some fine aspects, such as CERTIFICATE, have a higher percentage of instructor replies than others, such as QUIZ-GRADING. Using our system, instructors can sample from multiple aspect categories, thereby making sure that all categories of problems receive attention.

## 6 Conclusion

In this paper, we developed a weakly supervised joint probabilistic model (PSL-Joint) for predicting aspect-sentiment in online courses. Our model provides the ability to conveniently encode domain information in the form of seed words, and weighted logical rules capturing the dependencies between aspects and sentiment. We validated our approach on an annotated dataset of MOOC posts sampled from twelve courses. We compared our PSL-Joint probabilistic model to a simpler SeededLDA approach, and demonstrated that PSL-Joint produced statistically significantly better results, exhibiting a 3–5 times improvement in F1 score in most cases over a system using only SeededLDA. As further shown by our qualitative results and instructor reply information, our system can potentially be used for understanding student requirements and issues, identifying posts for instructor intervention, increasing student retention, and improving future iterations of the course.

| Correct Label | PSL | SeededLDA | Post |
|---|---|---|---|
| QUIZ-CONTENT | QUIZ-CONTENT | LECTURE-CONTENT | There is a typo or other mistake in the assignment instructions (e.g. essential information omitted) Type ID: programming-content Problem ID: programming-mistake Browser: Chrome 32 OS: Windows 7 |
| QUIZ-CONTENT | QUIZ-CONTENT | LECTURE-CONTENT | There is a typo or other mistake on the page (e.g. factual error information omitted) Week 4 Quiz Question 6: Question 6 When a user clicks on a View that has registered to show a Context Menu which one of the following methods will be called? |
| LECTURE-AUDIO | LECTURE-AUDIO | LECTURE-SUBTITLES | Thanks for the suggestion about downloading the video and referring to the subtitles. I will give that a try but I would also like to point out that what the others are saying is true for me too: The audio is just barely audible even when the volume on my computer is set to 100%. |
| SOCIAL | SOCIAL | LECTURE-VIDEO | Let's start a group for discussing the lecture videos. |

Table 13: Example posts that PSL-Joint predicted correctly, but were misclassified by SeededLDA

| Correct Label | Predicted Label | Second Prediction | Post |
|---|---|---|---|
| LECTURE-CONTENT | QUIZ-CONTENT | LECTURE-CONTENT | I have a difference of opinion to the answer for Question 6 too. It differs from what is presented in lecture 1. |
| SOCIAL | LECTURE-SUBTITLES | SOCIAL | Hello guys!!! I am ... The course materials are extraordinary. The subtitles are really helpful! Thanks to instructors for giving us all a wonderful opportunity. |
| LECTURE-CONTENT | QUIZ-CONTENT | LECTURE-CONTENT | As the second lecture video told me I started windows telnet and connected to the virtual device. Then I typed the same command for sending an sms that the lecture video told me to. The phone received a message all right and I was able to open it but the message itself seems to be written with some strange characters. |

Table 14: Example posts whose second-best prediction is correct

# References

S. H. Bach, M. Broecheler, B. Huang, and L. Getoor. 2015. Hinge-loss Markov random fields and probabilistic soft logic. arXiv:1505.04406 [cs.LG].

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, March.

Samuel Brody and Noemie Elhadad. 2010. An unsupervised aspect-sentiment model for online reviews. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 804–812. Association for Computational Linguistics.

Snigdha Chaturvedi, Dan Goldwasser, and Hal Daumé III. 2014. Predicting instructor's intervention in mooc forums. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1501–1511. Association for Computational Linguistics.

Qiming Diao, Minghui Qiu, Chao-Yuan Wu, Alexander J. Smola, Jing Jiang, and Chong Wang. 2014. Jointly modeling aspects, ratings and sentiments for movie recommendation (jmars). In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, pages 193–202, New York, NY, USA. ACM.

Angela Fahrni and Manfred Klenner. 2008. Old wine or warm beer: Target-specific sentiment analysis of adjectives. In *Proc. of the Symposium on Affective Language in Human and Machine, AISB*, pages 60–63.

Shobeir Fakhraei, Bert Huang, Louiqa Raschid, and Lise Getoor. 2014. Network-based drug-target interaction prediction with probabilistic soft logic. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 11(5):775–787, Sept.

Bert Huang, Angelika Kimmig, Lise Getoor, and Jennifer Golbeck. 2013. A flexible framework for probabilistic models of social trust. In *The 2013 International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction (SBP 2013)*.

Jagadeesh Jagarlamudi, Hal Daumé, III, and Raghavendra Udupa. 2012. Incorporating lexical priors into topic models. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '12.

Y. Jo and A.H. Oh. 2011. Aspect and sentiment unification model for online review analysis. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 815–824. ACM.

Suin Kim, Jianwen Zhang, Zheng Chen, Alice Oh, and Shixia Liu. 2013. A hierarchical aspect-sentiment model for online reviews. In *Proceedings of The Twenty-Seventh AAAI Conference on Artificial Intelligence (AAAI-13)*. AAAI, July.

Varada Kolhatkar, Heike Zinsmeister, and Graeme Hirst. 2013. Annotating anaphoric shell nouns with their antecedents. In *7th Linguistic Annotation Workshop and Interoperability with Discourse*. ACL.

Chenghua Lin and Yulan He. 2009. Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09, pages 375–384, New York, NY, USA. ACM.

Chenghua Lin, Yulan He, R. Everson, and S. Ruger. 2012. Weakly supervised joint sentiment-topic detection from text. *Knowledge and Data Engineering, IEEE Transactions on*, 24(6):1134–1145, June.

Bing Liu and Lei Zhang. 2012. A survey of opinion mining and sentiment analysis. In Charu C. Aggarwal and ChengXiang Zhai, editors, *Mining Text Data*, pages 415–463. Springer.

Edward Loper and Steven Bird. 2002. Nltk: The natural language toolkit. In *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics - Volume 1*, ETMTNLP '02, pages 63–70, Stroudsburg, PA, USA. Association for Computational Linguistics.

Bin Lu, Myle Ott, Claire Cardie, and Benjamin K. Tsou. 2011. Multi-aspect sentiment analysis with topic models. In *Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops*, ICDMW '11, pages 81–88, Washington, DC, USA. IEEE Computer Society.

Arjun Mukherjee and Bing Liu. 2012. Aspect extraction through semi-supervised modeling. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1*, ACL '12, pages 339–348, Stroudsburg, PA, USA. Association for Computational Linguistics.

Jay Pujara, Hui Miao, Lise Getoor, and William Cohen. 2013. Knowledge graph identification. In *International Semantic Web Conference (ISWC)*. Winner of Best Student Paper award.

Arti Ramesh, Dan Goldwasser, Bert Huang, Hal Daume III, and Lise Getoor. 2014a. Learning latent engagement patterns of students in online courses. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*.

Arti Ramesh, Dan Goldwasser, Bert Huang, Hal Daumé 2014b. Understanding mooc discussion forums using seeded lda. In *9th ACL Workshop on Innovative Use of NLP for Building Educational Applications*. ACL.

Glenda S. Stump, Jennifer DeBoer, Jonathan Whittinghill, and Lori Breslow. 2013. Development of a framework to classify mooc discussion forum posts: Methodology and challenges. In *NIPS Workshop on Data Driven Education*.

Ivan Titov and Ryan McDonald. 2008. Modeling online reviews with multi-grain topic models. In *Proceedings of the 17th International Conference on World Wide Web*, WWW '08, pages 111–120, New York, NY, USA. ACM.

Wayne Xin Zhao, Jing Jiang, Hongfei Yan, and Xiaoming Li. 2010. Jointly modeling aspects and opinions with a maxent-lda hybrid. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10, pages 56–65, Stroudsburg, PA, USA. Association for Computational Linguistics.